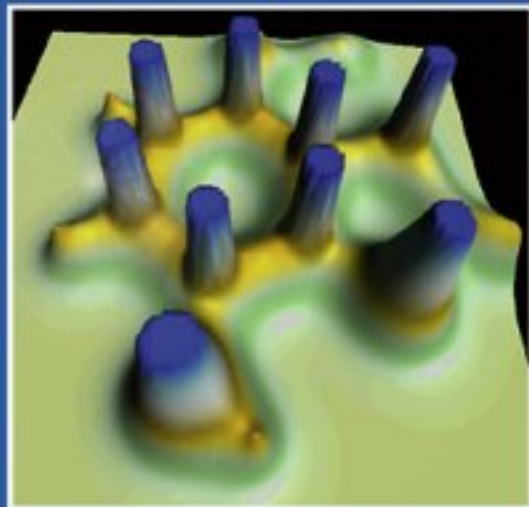


WILEY-VCH



Quantum Medicinal Chemistry

Edited by P. Carloni and F. Alber



**Methods
and Principles
in Medicinal
Chemistry**

Volume 17

Edited by
R. Mannhold,
H. Kubinyi,
G. Folkers

Quantum Medicinal Chemistry

Edited by
Paolo Carloni and Frank Alber

Methods and Principles in Medicinal Chemistry

Edited by

R. Mannhold, H. Kubinyi, G. Folkers

Editorial Board

H.-D. Höltje, H. Timmerman,

J. Vacca, H. van de Waterbeemd, T. Wieland

Quantum Medicinal Chemistry

Edited by

Paolo Carloni and Frank Alber



WILEY-
VCH

WILEY-VCH GmbH & Co. KGaA

Series Editors

Prof. Dr. Raimund Mannhold

Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
Germany
raimund.mannhold@uni-duesseldorf.de

Prof. Dr. Hugo Kubinyi

BASF AG Ludwigshafen
c/o Donnersbergstraße 9
67256 Weisenheim am Sand
Germany
kubinyi@t-online.de

Prof. Dr. Gerd Folkers

Department of Applied Biosciences
ETH Zürich
Winterthurer Straße 190
8057 Zürich
Switzerland
folkers@pharma.ethz.ch

Volume Editors

Prof. Dr. Paolo Carloni

International School for Advanced Studies
Via Beirut 4
34014 Trieste
Italy
carloni@sissa.it

Dr. Frank Alber

Laboratories of Molecular Biophysics
The Rockefeller University
1230 York Avenue, Box 270
New York, NY 10021-6399
USA
current address:
Dept. of Biopharmaceutical Sciences
University of California
San Francisco, CA 94143
USA
frank.alber@salilab.org

This book was carefully produced. Nevertheless, editors, authors and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Cover illustration

Electron density map of morphine showing the aromatic and part of the furanoid ring. Courtesy of C. Matta.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

ISBN 3-527-30456-8

© 2003 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Printed on acid free paper

All rights reserved (including those of translation in other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Composition K+V Fotosatz GmbH,
64743 Beerfelden

Printing Strauss Offsetdruck GmbH,
Mörtenbach

Bookbinding J. Schäffer GmbH & Co. KG.,
Grünstadt

Printed in the Federal Republic of Germany

Contents

Preface XI

Foreword XIII

List of Contributors XV

Outline of the Book 1

Density Functional Theory

- 1 **Advances in Density-functional-based Modeling Techniques – Recent Extensions of the Car-Parrinello Approach** 5
Daniel Sebastiani and Ursula Röthlisberger
- 1.1 Introduction 5
- 1.2 The Car-Parrinello Approach – Basic Ideas 6
- 1.2.1 How It Can be Done 8
- 1.2.2 *Ab Initio* Molecular Dynamics Programs 15
- 1.3 Mixed Quantum Mechanical/Molecular Mechanical (QM/MM) Car-Parrinello Simulations 15
- 1.3.1 Gly-Ala Dipeptide in Aqueous Solution – Do We Need a Polarizable Force Field? 19
- 1.4 Density-functional Perturbation Theory and the Calculation of Response Properties 21
- 1.4.1 Introduction to Density-functional Perturbation Theory 21
- 1.4.2 Basic Equations of Density-functional Perturbation Theory 22
- 1.4.3 NMR Chemical Shieldings within DFPT 26

1.4.3.1	Introduction to Nuclear Magnetic Resonance Chemical Shifts	26
1.4.3.2	NMR Chemical Shielding	27
1.4.3.3	Calculation of NMR Chemical Shifts in QM/MM Car-Parrinello Simulations	30
1.5	Introduction to Time-dependent Density-functional Theory (TD-DFT)	32
1.5.1	Basic Equations of TD-DFT	33
1.5.2	Applications of TD-DFT within the QM/MM Framework – Opsochromic Shift of Acetone in Water	35
1.6	Acknowledgments	36
1.7	References	36
2	Density-functional Theory Applications in Computational Medicinal Chemistry	41
	<i>Andrea Cavalli, Gerd Folkers, Maurizio Recanatini, and Leonardo Scapozza</i>	
2.1	Introduction	41
2.2	Density-functional Theory and Related Methods	42
2.2.1	Density-functional Theory	42
2.2.2	<i>Ab Initio</i> Molecular Dynamics	45
2.3	SAR Studies of Ligand-Target Interactions	48
2.3.1	The Case Study: Herpes Simplex Virus Type 1 Thymidine Kinase Substrates and Inhibitors	48
2.3.1.1	Rationalizing Substrate Diversity – SAR of HSV1 TK Ligands	51
2.3.1.2	What Can be Learned from this Case Study – From SAR to Drug Design	56
2.4	Theoretical Studies of Enzymatic Catalysis	57
2.4.1	The Phosphoryl Transfer Reaction	58
2.4.1.1	Cdc42-catalyzed GTP Hydrolysis	58
2.4.1.2	HIV-1 Integrase	63
2.5	Studies on Transition Metal Complexes	64
2.5.1	Radiopharmaceuticals	65
2.6	Conclusions and Perspectives	67
2.7	References	68

3	Applications of Car-Parrinello Molecular Dynamics in Biochemistry – Binding of Ligands in Myoglobin	73
	<i>Carme Rovira</i>	
3.1	Introduction	74
3.2	Computational Details	79
3.3	Myoglobin Active Center	81
3.3.1	Structure, Energy, and Electronic State	81
3.3.2	The <i>Picket-fence</i> -oxygen Biomimetic Complex	86
3.3.2.1	Interplay Structure/Electronic State	86
3.3.2.2	Optimized Structure and Energy of O ₂ Binding	90
3.3.3	Heme-Ligand Dynamics	93
3.4	Interaction of the Heme with the Protein	99
3.5	Conclusions	106
3.6	Acknowledgments	108
3.7	References	108
4	Density-functional Theory in Drug Design – the Chemistry of the Anti-tumor Drug Cisplatin and Photoactive Psoralen Compounds	113
	<i>Johan Raber, Jorge Liano, and Leif A. Eriksson</i>	
4.1	Introduction	113
4.2	Density-functional Theory	114
4.2.1	Basic Equations	115
4.2.2	Gradient Corrections and Hybrid Functionals	117
4.2.3	Time-dependent Density-functional Response Theory (TD-DFRT)	120
4.2.4	Applicability and Applications	122
4.3	Modes of Action of Anti-tumor Drug Cisplatin	124
4.3.1	Activation Reactions	127
4.3.2	Interactions Between DNA and Cisplatin	134
4.4	Photochemistry of Psoralen Compounds	141
4.4.1	Ionization Potentials	143
4.4.2	Excitation Spectra	146
4.5	Acknowledgments	150
4.6	References	150

QM/MM Approaches

- 5 *Ab Initio* Methods in the Study of Reaction Mechanisms – Their Role and Perspectives in Medicinal Chemistry 157**
Mikael Peräkylä
- 5.1 Introduction 157
- 5.2 Methods 161
- 5.2.1 Hybrid QM/MM Potential 161
- 5.2.2 QM/MM Boundary – The Link Atom Approach 161
- 5.2.3 QM/MM Boundary – The Hybrid Orbital Approach 165
- 5.3 Thermodynamically Coupled QM/MM 166
- 5.4 Selected Applications of QM/MM Methods 168
- 5.4.1 Uracil-DNA Glycosylase 168
- 5.4.2 QM/MM Simulations of Quantum Effects 169
- 5.4.3 Miscellaneous Applications 170
- 5.5 Conclusions 173
- 5.6 References 173
- 6 Quantum-mechanical/Molecular-mechanical Methods in Medicinal Chemistry 177**
Francesca Perruccio, Lars Ridder, and Adrian J. Mulholland
- 6.1 Introduction 177
- 6.2 Theory 178
- 6.2.1 Methodology 178
- 6.2.2 Basic Theory 179
- 6.2.3 QM/MM Partitioning Schemes 180
- 6.3 Practical Aspects of Modeling Enzyme Reactions 182
- 6.3.1 Choice and Preparation of the Starting Structure 182
- 6.3.2 Definition of the QM Region 183
- 6.3.3 Choice of the QM Method 184
- 6.4 Techniques for Reaction Modeling 185
- 6.4.1 Optimization of Transition Structures and Reaction Pathways 185
- 6.4.2 Dynamics and Free Energy Calculations 186
- 6.5 Some Recent Applications 189
- 6.5.1 Human Aldose Reductase 189
- 6.5.2 Glutathione S-Transferases 191
- 6.5.3 Influenza Neuraminidase 193

- 6.5.4 Human Thrombin 193
- 6.5.5 Human Immunodeficiency Virus Protease 194
- 6.6 Conclusions 195
- 6.7 References 195

Molecular Properties

- 7 Atoms in Medicinal Chemistry 201**
Richard F. W. Bader, Cherif F. Matta, and Fernando J. Martin
- 7.1 Why Define Atoms in Molecules? 201
- 7.2 Theory of Atoms in Molecules 202
- 7.2.1 Definition of Atoms and Molecular Structure 203
- 7.3 Definition of Atomic Properties 208
- 7.3.1 Atomic Charges, Multipole Moments and Volumes 209
- 7.4 QTAIM and Correlation of Physicochemical Properties 211
- 7.4.1 Use of Atomic Properties in QSAR 211
- 7.4.2 Use of Bond Critical Point Properties in QSAR 213
- 7.4.3 QTAIM and Molecular Similarity 215
- 7.5 Use of QTAIM in Theoretical Synthesis of Macromolecules 218
- 7.5.1 Assumed Perfect Transferability in the Synthesis of a Polypeptide 219
- 7.5.2 The Assembly of Buffered Open Systems in a Macrosynthesis 222
- 7.6 The Laplacian of the Density and the Lewis Model 224
- 7.6.1 The Laplacian and Acid-Base Reactivity 225
- 7.6.2 Molecular Complementarity 228
- 7.7 Conclusions 229
- 7.8 References 230
- 8 The Use of the Molecular Electrostatic Potential in Medicin Chemistry 233**
Jane S. Murray and Peter Politzer
- 8.1 Introduction 233
- 8.2 Methodology 235
- 8.3 An Example that Focuses on V_{\min} – the Carcinogenicity of Halogenated Olefins and their Epoxides 239

- 8.4 An Example Focusing on the General Patterns of Molecular Electrostatic Potentials – Toxicity of Dibenzo-*p*-dioxins and Analogs 244
- 8.5 Statistical Characterization of the Molecular Surface Electrostatic Potential – the General Interaction Properties Function (GIPF) 246
- 8.6 Summary 250
- 8.7 Acknowledgment 250
- 8.8 References 250

- 9 Applications of Quantum Chemical Methods in Drug Design 255**
Hans-Dieter Höltje and Monika Höltje
- 9.1 Introduction 255
- 9.2 Application Examples 256
 - 9.2.1 Force Field Parameters from *Ab Initio* Calculations 256
 - 9.2.1.1 Equilibrium Geometry for a Dopamine-D₃-Receptor Agonist 260
 - 9.2.1.2 Searching for a Bioactive Conformation 262
 - 9.2.2 Atomic Point Charges 264
 - 9.2.3 Molecular Electrostatic Potentials 266
 - 9.2.4 Molecular Orbital Calculations 268
- 9.3 Outlook 273
- 9.4 Acknowledgment 274
- 9.5 References 274

Subject Index 275

Preface

Everyone relies on the power of computers, including chemical and pharmaceutical laboratories. Increasingly faster and more exact simulation algorithms have made quantum chemistry a valuable tool in the search for bioactive substances. The much larger computational cost is more than compensated by a deeper understanding of the physicochemical events taking place at the interaction of ligands and proteins. Special interest in biomolecular simulation is now given to catalytic centers in proteins which contain metals. Many of the DNA binding proteins involved in the control of the transcription processes contain metallic centers. Standard empirical methods, which have undeniable merits in the field of structure-based design, nevertheless fail to describe subtle chemical phenomena as partially covalent bonds or non-rigid aromatic moieties. Another field of high interest in medicinal chemistry are ligands that interfere with ion channels. Also here the presence of large electric fields demands a more sophisticated approach. *Ab initio* molecular dynamics which typically make use of density functional theory add another piece to the mosaic pattern of understanding ligand-protein interactions. Experience that has accumulated in recent years in the fields of material sciences and medicinal chemistry shows a unique role of *ab initio* molecular dynamics in studying complex interaction phenomena with a close coupling to experimental, mostly spectroscopical data.

These few remarks highlight that quantum-chemical methods have adapted an important role in medicinal chemistry. It is the intention of the present volume to document this role in adequate detail. Accordingly, the book is divided into three main sections. The first section is dedicated to density functional theory.

A description of advances in density functional based modelling techniques is followed by application examples in computational medicinal chemistry, biochemistry and drug design. The following section is focussed on QM/MM approaches and describes e.g. the use of *ab initio* methods in the study of reaction mechanisms. The last section presents a survey of pharmaceutically relevant properties derived by quantum-chemical calculations such as molecular electrostatic potentials. In a finalizing chapter applications of quantum-chemical methods to systems of biological and pharmacological relevance are described.

The series editors would like to thank the authors and in particular the volume editors, Paolo Carloni and Frank Alber, that they devoted their precious time to compiling and structuring the comprehensive information on medicinal quantum chemistry. Last, but not least we want to express our gratitude to Frank Weinreich and Gudrun Walter from Wiley-VCH publishers for the fruitful collaboration.

December 2002

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Ludwigshafen
Gerd Folkers, Zürich

Foreword

Quantum-chemical (QC) calculations are a key element in biological research. When constantly tested for their range of validity QC methods provide a description of how molecules interact and form their three-dimensional shape, which in turn determines molecular function. They can aid the formulation of hypotheses that provide the connecting link between experimentally determined structures and biological function. QC calculations can be used to understand enzyme mechanisms, hydrogen bonding, polarization effects, spectra, ligand binding and other fundamental processes both in normal and aberrant biological contexts. The power of parallel computing and progress in computer algorithms are enlarging the domain of QC applications to ever more realistic models of biological macromolecules. This book is meant to serve as a reference for chemists, biochemists, and pharmacologists interested in learning about and using state-of-the-art QC techniques to investigate systems and processes of pharmaceutical relevance. We are confident that the contributions presented here will provide further support for the development and applications of quantum-mechanical methods in the applied biosciences.

New York and Trieste
July 2002

Frank Alber and Paolo Carloni

List of Contributors

Editors

Dr. FRANK ALBER
Laboratories of Molecular
Biophysics
The Rockefeller University
1230 York Avenue, Box 270
New York, NY 10021-6399
USA
frank.alber@salilab.org

Prof. Dr. PAOLO CARLONI
International School for Advanced
Studies (SISSA/ISAS)
Via Beirut 4
34014 Trieste
Italy
carloni@sissa.it

Authors

Prof. Dr. RICHARD F. W. BADER
Department of Chemistry
McMaster University
1280 Main Street West
Hamilton
Ontario, L8S 4M1
Canada
bader@mcmil.cis.mcmaster.ca

Dr. ANDREA CAVALLI
Department of Pharmaceutical
Sciences
University of Bologna
Via Belmeloro 6
40126 Bologna
Italy
cavan@alma.unibo.it

Prof. Dr. LEIF A. ERIKSSON
Division of Structural and
Computational Biophysics
Department of Biochemistry
Uppsala University
Box 576
75123 Uppsala
Sweden
leif.eriksson@xray.bmc.uu.se

Prof. Dr. GERD FOLKERS
Department of Applied Biosciences
ETH Zürich
Winterthurerstraße 190
8057 Zürich
Switzerland
folkers@pharma.anbi.aethz.ch

XVI | *List of Contributors*

Prof. Dr. HANS-DIETER HÖLTJE
Institute of Pharmaceutical
Chemistry
Heinrich-Heine Universität
Universitätsstraße 1
40225 Düsseldorf
Germany
hoeltje@pharm.uni-duesseldorf.de

Dr. MONIKA HÖLTJE
Institute of Pharmaceutical
Chemistry
Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
Germany
mhoeltje@pharm.
uni-duesseldorf.de

Dr. JORGE LLANO
Division of Structural and
Computational Biophysics
Department of Biochemistry and
Department of Quantum Chemistry
Uppsala University
Box 576
75123 Uppsala
Sweden
Jorge.Llano@xray.bmc.uu.se

Dr. FERNANDO J. MARTÍN
Department of Pharmaceutical
Chemistry
University of California
San Francisco
513 Parnassus Avenue
Box 0446
San Francisco, CA 94143
USA

Dr. CHERIF F. MATTA
Department of Chemistry
McMaster University
Hamilton
1280 Main Street West
Ontario, L8S 4M1
Canada
mattacf@mcmaster.ca
current adress:

Lash Miller Chemical Laboratories
Chemistry Department
University of Toronto
Toronto, Ontario
Canada, M5S 1A1

Prof. Dr. ADRIAN J. MULHOLLAND
School of Chemistry
University of Bristol
Bristol BS8 1TS
UK
Adrian.Mulholland@bris.ac.uk

Prof. Dr. JANE S. MURRAY
Department of Chemistry
University of New Orleans
Elysian Fields Ave.
New Orleans, LA 70148
USA
jsmurray@uno.edu

Dr. FRANCESCA PERRUCCIO
School of Chemistry
University of Bristol
Bristol BS8 1TS
UK
current adress:
Molecular Informatics
Structure and Design
Pfizer Global Research
and Development
Ramsgate Road
Sandwich, Kent CT13 9NJ
UK

Prof. Dr. MIKAEL PERÄKYLÄ
 Department of Chemistry
 University of Kuopio
 Savilahdentie 9 F
 P.O. Box 1627
 70211 Kuopio
 Finland
 mikael.perakyla@uku.fi

Prof. Dr. PETER POLITZER
 Department of Chemistry
 University of New Orleans
 Elysian Fields Ave.
 New Orleans, LA 70148
 USA
 ppolitzer@uno.edu

Dr. JOHAN RABER
 Division of Structural and Computational Biophysics
 Department of Biochemistry
 Uppsala University
 Savilahdentie 9 F
 Box 576
 75123 Uppsala
 Sweden
 Johan.Raber@xray.bmc.uu.se

Prof. Dr. MAURIZIO RECANATINI
 Department of Pharmaceutical Sciences
 University of Bologna
 Via Belmeloro 6
 40126 Bologna
 Italy
 mreca@alma.unibo.it

Dr. LARS RIDDER
 School of Chemistry
 University of Bristol
 Bristol BS8 1TS
 UK
current adress:
 Molecular Design & Informatics
 N.V. Organon, 5340 BH Oss
 The Netherlands

Prof. Dr. URSULA RÖTHLISBERGER
 Institute of Molecular and Biological Chemistry
 Swiss Federal Institute of Technology
 EPFL
 1015 Lausanne
 Switzerland
 and Institute of Inorganic Chemistry
 ETH Zürich
 Universitätsstraße 6
 8092 Zürich
 Switzerland
 uro@inorg.chem.ethz.ch

Dr. CARME ROVIRA
 Centre de Recerca en Química Teòrica
 Parc Científic de Barcelona (PCB)
 Annex A, pta. 1
 Josep Samitier 1–5
 08028 Barcelona
 Spain
 and Departament de Química Física
 Facultat de Química
 Universitat de Barcelona
 Martí i Franquès 1
 08028 Barcelona
 Spain
 crovira@pcb.ub.es

XVIII | *List of Contributors*

Prof. Dr. LEONARDO SCAPOZZA
Department of Applied Biosciences
ETH Zürich
Winterthurerstraße 190
8057 Zürich
Switzerland
leonardo.scapozza@pharma.
anbi.ethz.ch

Dr. DANIEL SEBASTIANI
Institute of Molecular
and Biological Chemistry
Swiss Federal Institute
of Technology
EPFL
1015 Lausanne
Switzerland
and Max-Planck-Institute
of Polymer Research
Ackermannweg 10
55128 Mainz
Germany
sebastia@mpip-mainz.mpg.de

Outline of the Book

The book is organized into three major parts which cover important and emerging methods and applications in biological and pharmacological research.

The first part focuses on density functional theory (DFT), one of the most successful first-principle approaches to investigation of the electronic structure of relatively large model systems; its suitability for tackling chemical problems was recognized by the Nobel prize awarded to W. Kohn in 1998. In the first chapter, U. Röthlisberger and D. Sebastiani outline the principles of DFT and describe selected major advances, from the DFT-based molecular dynamics (Car-Parrinello) method (and its extension to hybrid DFT/classical molecular dynamics), to time-dependent DFT methods which enable extension of DFT to investigation of excited states and DFT perturbation theory for calculation of NMR chemical shifts. This rather methodological chapter serves as a reference for Chapters 2–4, which describe a wide spectrum of applications. A. Cavalli et al. (Chapter 2) discuss DFT studies of ligand-target interactions and the mechanism of specific enzyme systems. Subsequently, C. Rovira presents a detailed survey of Car-Parrinello applications to iron porphyrin proteins (Chapter 3). Finally, J. Raber et al. describe applications of DFT to the mode of action of the antitumor drug cisplatin and to description of the excitation spectra of photochemotherapeutic compounds (Chapter 4).

The second part of the book provides a description of QM/MM approaches. Because of their large size, biological systems such as proteins and nucleic acids cannot be treated fully at the

quantum chemical (QC) level. QM/MM approaches combine a QC method (performed at semiempirical *ab initio* or density-functional levels) for the region of interest (for example, the active site of an enzyme) with molecular mechanics (MM) treatment of the environment. M. Peräkylä introduces the basic theory of structurally and thermodynamically coupled QM/MM approaches. He describes, in detail, treatment of the boundary between the QM and the MM regions (Chapter 5). Practical aspects and selected applications are further discussed by F. Peruccio et al. (Chapter 6).

The final part is devoted to a survey of molecular properties of special interest to the medicinal chemist. The Theory of Atoms in Molecules by R. F. W. Bader et al., presented in Chapter 7, enables the quantitative use of chemical concepts, for example those of the functional group in organic chemistry or molecular similarity in medicinal chemistry, for prediction and understanding of chemical processes. This contribution also discusses possible applications of the theory to QSAR. Another important property that can be derived by use of QC calculations is the molecular electrostatic potential. J. S. Murray and P. Politzer describe the use of this property for description of noncovalent interactions between ligand and receptor, and the design of new compounds with specific features (Chapter 8). In Chapter 9, H. D. and M. Höltje describe the use of QC methods to parameterize force-field parameters, and applications to a pharmacophore search of enzyme inhibitors. The authors also show the use of QC methods for investigation of charge-transfer complexes.

Density Functional Theory

1

Advances in Density-functional-based Modeling Techniques

– Recent Extensions of the Car-Parrinello Approach

DANIEL SEBASTIANI and URSULA RÖTHLISBERGER

1.1 Introduction

During the last decade, density-functional theory (DFT)-based approaches [1, 2] have advanced to prominent first-principles quantum chemical methods. As computationally affordable tools apt to treat fairly extended systems at the correlated level, they are also of special interest for applications in medicinal chemistry (as demonstrated in the chapters by Rovira, Raber et al. and Cavalli et al. in this book). Several excellent text books [3–5] and reviews [6] are available as introduction to the basic theory and to the various flavors of its practical realization (in terms of different approximations for the exchange-correlation functional). The actual performance of these different approximations for diverse chemical [7] and biological systems [8] has been evaluated in a number of contributions.

In this chapter we will focus on one particular, recently developed DFT-based approach, namely on first-principles (Car-Parrinello) molecular dynamics (CP-MD) [9] and its latest advancements into a mixed quantum mechanical/molecular mechanical (QM/MM) scheme [10–12] in combination with the calculation of various response properties [13–18] within DFT perturbation theory (DFTPT) and time-dependent DFT theory (TDDFT) [19].

First-principles (or ‘*ab initio*’) molecular dynamics, the direct combination of DFT with classical molecular dynamics (MD), was introduced in 1985 in a seminal paper by Car and Parrinello [9]. This novel scheme has first been applied to the study of metal clusters [20] and amorphous and liquid silicon [21] but has since moved rapidly into chemistry [22] and biology [23, 24]. CP-MD of

fers the unique possibility of performing parameter-free MD simulations in which all the interactions are calculated on-the-fly within the framework of DFT (or alternative electronic structure methods [70–73]). In this way, finite temperature and entropic effects are included in a straightforward manner in the context of a quantum chemical electronic structure calculation and simulations can be performed in realistic condensed-phase environments. Furthermore, this approach is also highly amenable to parallelization so that currently simulations of 100–1000 atoms can be performed at relative ease. CP-MD thus offers promising perspectives for applications in medicinal chemistry, and a growing number of such studies has started to emerge in recent years [25].

Lately, the CP-MD approach has been combined with a mixed QM/MM scheme [10–12] which enables the treatment of chemical reactions in biological systems comprising tens of thousands of atoms [11, 26]. Furthermore, CP-MD and mixed QM/MM CP-MD simulations have also been extended to the treatment of excited states within a restricted open-shell Kohn-Sham approach [16, 17, 27] or within a linear response formulation of TDDFT [16, 18], enabling the study of biological photoreceptors [28] and the *in situ* design of optimal fluorescence probes with tailored optical properties [32]. Among the latest extensions of this method are also the calculation of NMR chemical shifts [14].

Here, we will first give an introduction to the basic ideas underlying the Car-Parrinello method, especially addressed to complete newcomers in the field. We will then try to outline some of the recent methodological extensions, with particular emphasis on aspects with potential interest for applications in medicinal chemistry. The power and limitations of these new modeling tools will be illustrated with few selected examples.

1.2

The Car-Parrinello Approach – Basic Ideas

The Car-Parrinello approach combines an electronic structure method with a classical molecular dynamics scheme and thus unifies two major fields of computational chemistry, which have hitherto been essentially orthogonal. Through this unification a

Tab. 1.1 Comparison of the properties of quantum chemical electronic structure calculations (QC methods), classical molecular dynamics (Classical MD) based on empirical force fields and first-principles molecular dynamics (*ab initio* MD) simulations.

QC Methods	Classical MD	<i>ab initio</i> MD
⊕ High accuracy	⊖ Limited to accuracy of empirical force field	⊕ QC accuracy
⊕ First-principles approach	⊖ Parameterization effort ⊖ limited transferability	⊕ Parameter-free MD based on first-principles
⊕ Description of chemical reactions	⊖ MD simulation of chemical reactions difficult	⊕ MD simulations of chemical reactions possible
⊕ Treatment of transition metal ions possible	⊖ Treatment of transition metal ions difficult	⊕ Treatment of transition metal ions possible
⊖ Zero Kelvin	⊕ Finite temperature	⊕ Finite temperature
⊖ Mostly gas phase only	⊕ Also condensed phases	⊕ Also condensed phases
⊖ Limited to characterization of few selected points of the PES	⊕ Finite temperature sampling of the PES	⊕ Finite temperature sampling of the PES
⊖ Local geometry optimization	⊕ Simulated annealing	⊕ Simulated annealing

series of new features has become available that goes far beyond the capabilities of each of the single parts on their own. A few of these special aspects are summarized schematically in Tab. 1.1 and illustrated in Fig. 1.1 a, b.

Quantum chemical (QC) methods have the advantage of high intrinsic accuracy but are essentially limited to the treatment of small molecules in the gas phase at static nuclear configurations. In a typical QC calculation only few points of the potential energy surface (PES) of the system are characterized by localizing the stationary point (minimum or transition state) that lies closest to a given starting configuration. In this way only a limited portion of the PES can be sampled and the system might be trapped in a local minimum far from the energetically most favorable configurations.

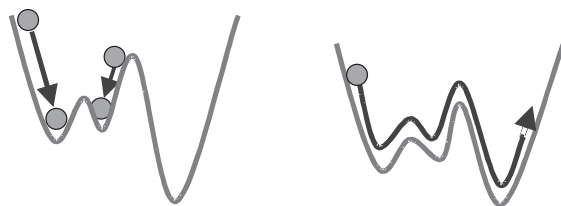


Fig. 1.1 (a) In traditional quantum chemical methods the potential energy surface (PES) is characterized in a pointwise fashion. Starting from an initial geometry, optimization routines are applied to localize the nearest stationary point (minimum or transition state). Which point of the PES results from this procedure mainly depends on the choice of the initial configuration. The system can get trapped easily in local minima without ever arriving at the global minimum struc-

ture. On the other hand, the PES can be described with high accuracy.

(b) In classical molecular dynamics, the PES is approximated via an appropriately parameterized empirical force field. The particles possess kinetic energy with which they can overcome local barriers and access a wide portion of the PES. The reliability of the approach is limited by the accuracy of the underlying empirical force field.

In classical molecular dynamics, on the other hand, particles move according to the laws of classical mechanics over a PES that has been empirically parameterized. By means of their kinetic energy they can overcome energetic barriers and visit a much more extended portion of phase space. Tools from statistical mechanics can, moreover, be used to determine thermodynamic (e.g. relative free energies) and dynamic properties of the system from its temporal evolution. The quality of the results is, however, limited to the accuracy and reliability of the (empirically) parameterized PES.

1.2.1

How It Can be Done

How can one join an electronic structure calculation with a classical MD scheme? In principle, this is possible in a straightforward manner – we can optimize the electronic wavefunction for a given initial atomic configuration (at time $t=0$) and calculate the forces acting on the atoms via the Hellman-Feynman theorem:

$$\vec{F}_I = - \left\langle \psi \left| \frac{dH}{d\vec{R}_I} \right| \psi \right\rangle$$

These forces can then be plugged into the classical equations of motion (EOM) (Newton's equations of motion) and the system can be propagated to a new configuration at a time $t + \Delta t$, at which we can repeat the whole procedure again. Implementations which use this direct procedure are usually referred to as 'Born-Oppenheimer' first-principles MD schemes. With the powerful computers available today this type of dynamics has become possible but it usually requires special care in order to optimize the efficiency of the full electronic structure calculation that has to be performed at every time step. An alternative way that does not require full blown electronic structure calculation at every time step has been proposed by Car and Parrinello in 1985 [9]. They have suggested including the electronic wavefunctions (i.e. in DFT as the underlying QC method, the Kohn-Sham one-particle states) explicitly in the calculation and propagating them in parallel to the motions of the atoms. This can be achieved by the elegant trick of considering the one-particle orbitals as fictitious classical degrees of freedom that evolve under the laws of classical mechanics. Instead of using the familiar Newton's equation of motion, such a scheme is more conveniently formulated in terms of the equivalent Lagrangian formulation of classical mechanics. In Lagrangian mechanics the system is described in terms of generalized coordinates q_i and their conjugate momenta $p_i = m \delta q_i / \delta t$. The use of generalized coordinates facilitates the introduction of the electronic variables as additional classical degrees of freedom and allows us to treat them on the same footing as the atomic motion. The central quantity that describes these dynamics is the Lagrangian L :

$$L = K - E_{\text{pot}}$$

where K is the kinetic energy and E_{pot} is the potential energy. For our combined system consisting of nuclear and electronic coordinates the extended Lagrangian, L_{ex} , can be written as:

$$L_{\text{ex}} = K_{\text{N}} + K_{\text{e}} - E_{\text{pot}}$$

where K_{N} is the kinetic energy of the nuclei, K_{e} is the analogous term for the electronic degrees of freedom, and E_{pot} is the poten-

tial energy, which depends on both the nuclear positions $\{\bar{R}_I\}$ and the electronic variables $|\varphi_i\rangle$. The Lagrangian determines the time evolution of the classical system via EOM that are given by the Euler-Lagrange equations:

$$\frac{d}{dt} \left(\frac{\delta L}{\delta \dot{q}_i^*} \right) = \frac{\delta L}{\delta q_i^*}$$

A further advantage of using Lagrangian dynamics is that we can easily impose boundary conditions and constraints by applying the method of Lagrangian multipliers. This is particularly important for the dynamics of the electronic degrees of freedom, as we will have to impose that the one-electron wavefunctions remain orthonormal during their time evolution. The L_{ex} of our extended system can then be written as:

$$\begin{aligned} L_{\text{ex}} = & \sum_I \frac{1}{2} M_I \dot{\bar{R}}_I^2 + \sum_i \mu |\dot{\varphi}_i|^2 - \langle \psi_0 | H | \psi_0 \rangle \\ & + \sum_{i,j} A_{ij} [\{ \int \varphi_i^*(\vec{r}) \varphi_j(\vec{r}) d\vec{r} \} - \delta_{ij}] \end{aligned}$$

where the A_{ij} are Lagrange multipliers that ensure orthonormality of the one-electron wavefunctions $|\varphi_i\rangle$, μ is a fictitious mass associated with the electronic degrees of freedom and the potential energy is given by the expectation value of the total (ground state) energy of the system $E = \langle \psi_0 | H | \psi_0 \rangle$. L_{ex} determines the time evolution of a fictitious classical system in which nuclear positions as well as electronic degrees of freedom are treated as dynamic variables and the equation of motion for both degrees of freedom can be derived via the Euler-Lagrange equations. The EOM for the nuclear degrees of freedom become:

$$M_I \ddot{\bar{R}}_I = - \frac{\delta E}{\delta \bar{R}_I}$$

and for the electronic ones:

$$\mu \ddot{\varphi}_i = -H\varphi_i + \sum_j A_{ij} \varphi_j$$

where the term with the Lagrange multipliers A_{ij} describes the constraint forces needed to keep the wavefunctions orthonormal dur-

ing the dynamics. The parameter μ is a purely fictitious variable and can be assigned an arbitrary value. By full analogy with the nuclear degrees of freedom, μ determines the rate at which the electronic variables evolve in time. In particular, the ratio of M_I to μ characterizes the relative speed at which the electronic variables propagate relative to the nuclear positions. For $\mu \ll M_I$ the electronic degrees of freedom adjust instantaneously to changes in the nuclear coordinates and the resulting dynamics are adiabatic. Under these conditions $K_e \ll K_N$ and the extended Lagrangian L_{ex} becomes identical to the physical Lagrangian L of the system:

$$L = K_N - E_{\text{pot}}$$

For finite values of μ the system moves within a limited width, given by the fictitious electronic kinetic energy, above the Born-Oppenheimer surface. Adiabacity is ensured if the highest frequency of the nuclear motion ω_I^{max} is well separated from the lowest frequency associated with the fictitious motion of the electronic degrees of freedom ω_e^{min} . It can be shown [30] that ω_e^{min} is proportional to the gap E_g :

$$\omega_e a \sqrt{\frac{E_g}{\mu}}$$

For systems with a finite E_g , the parameter μ can be used to shift the electronic frequency spectrum so that $\omega_e^{\text{min}} \gg \omega_I^{\text{max}}$ and no energy transfer takes place between the nuclear and electronic subsystems. For metallic systems special variations of the original method must be adopted [31–34]. In practice, it is easy to check if adiabatic conditions are fulfilled by monitoring the energy conservation of the physical Lagrangian, L . The integration of the EOM for nuclear and electronic degrees of freedom with a standard MD integration algorithm, such as Verlet [35] or velocity Verlet [36], generates classical nuclear trajectories on a quantum mechanical PES. After initial optimization of the electronic wavefunctions for a given starting configuration ionic and electronic degrees of freedom can be propagated in parallel along the Born-Oppenheimer surface without having to perform a full electronic structure calculation at each point. It is important to note that the electronic dynamics that is generated in this way does not describe the real

quantum dynamics of the electronic wavefunctions (which we would be able to describe only by solving the time-dependent electronic Schrödinger equation) but is merely an elegant way of propagating the wavefunction simultaneously to the nuclear motion, i.e. enabling the electronic structure to adjust instantaneously to every new nuclear configuration.

The Car-Parrinello method is similar in spirit to the extended system methods [37] for constant temperature [38, 39] or constant pressure dynamics [40]. Extensions of the original scheme to the canonical NVT-ensemble, the NPT-ensemble, or to variable cell constant-pressure dynamics [41] are hence in principle straightforward [42, 43]. The treatment of quantum effects on the ionic motion is also easily included in the framework of a path-integral formalism [44–47].

Most of the current implementations employ the original Car-Parrinello scheme based on DFT. The system is treated within periodic boundary conditions (PBC) and the Kohn-Sham (KS) one-electron orbitals $|\varphi_i\rangle$ are expanded in a basis set of plane waves (with wave vectors \vec{G}_m) [48–50]:

$$\varphi_i(\vec{r}) = \frac{1}{\sqrt{V_{\text{cell}}}} \sum_m c_{im} e^{i\vec{G}_m \cdot \vec{r}}$$

up to a given kinetic energy cutoff E_{cut} . Substituting this basis set expansion into the extended Lagrangian L_{ex} gives:

$$L_{\text{ex}} = \sum_I \frac{1}{2} M_I \dot{\vec{R}}_I^2 + \mu \sum_i \sum_m |\dot{c}_{im}|^2 - E_{\text{KS}} + \sum_{i,j} A_{ij} \left[\sum_m c_{im}^* c_{jm} - \delta_{ij} \right]$$

and the EOM for the electronic degrees of freedom is replaced by analogous classical equations for the plane wave coefficients c_{im} :

$$\mu \ddot{c}_{im} = -\frac{\delta E}{\delta c_{im}^*} + \sum_j A_{ij} c_{jm}$$

Due to their localized nature, core electrons can only be adequately described with \vec{G} vectors of very high frequency, which would necessitate the use of prohibitively large basis sets in a standard plane wave scheme. Consequently, only valence electrons are treated explicitly and the effect of the ionic cores is integrated out

by use of a pseudopotential formalism [51–53]. Consistent with the first-principles character of Car-Parrinello simulations, the pseudopotentials used for this purpose are *ab initio* pseudopotentials (AIPP). AIPP are derived directly from atomic all-electron calculations and different schemes are available for their construction [54–61]. One of the general recipes is to impose the condition that, for a specific atomic reference configuration, all-electron $\varphi^{\text{ae}}(r)$ and pseudo wavefunction $\varphi^{\text{ps}}(r)$ have identical eigenvalues and coincide outside a given core radius r_c . The rapidly oscillating all-electron wavefunction within the core is replaced by an arbitrary smooth, nodeless function. If the pseudo-wavefunction within the core is chosen in such a way that:

$$\int_0^{r_c} |\varphi^{\text{ps}}(r)|^2 dr = \int_0^{r_c} |\varphi^{\text{ae}}(r)|^2 dr$$

the resulting pseudopotential is norm-conserving. Norm-conservation imposes that the charge enclosed in the core region is identical to the all-electron case and is therefore an important property in ensuring the electrostatic transferability behavior of AIPP. After a pseudo-wavefunction has been constructed in this way, the radial Schrödinger equation can be inverted to find the corresponding potential that, after descreening of the Hartree and exchange-correlation contributions of the core, yields the required pseudopotential acting on the valence electrons. AIPP are, in general, of a nonlocal form, i.e. they consist of a local component at long range and a nonlocal, angular momentum-dependent contribution at short range. A variety of different types, such as standard norm-conserving [55, 57], soft norm-conserving [54, 56], and ultrasoft Vanderbilt [59] pseudopotentials are currently in use in the context of Car-Parrinello simulations.

Although the use of a pseudopotential formalism helps significantly in reducing the size of a plane wave basis set, typical expansions still include impressive numbers of 10,000–100,000 plane wave coefficients. All of these have to be propagated simultaneously during the dynamics; this makes AIMD approaches highly memory intensive.

The use of a plane wave expansion implies the presence of periodic boundary conditions (PBCs). This is a natural selection in

the description of crystalline solids and is also appropriate in the simulation of liquids. PBC are, however, a less obvious choice for calculations of finite size systems, such as molecules or clusters. These must be placed in a super cell of sufficient size, so that the distance between the periodic replicas of the system is large enough and their mutual interactions become negligible. Because of the long-range nature of Coulomb interactions, the cell size needed to completely screen the images of charged systems would be excessively large and these cases must be treated with special care. Different methods are available for this purpose [62–64].

Working with a basis set expansion of plane waves renders calculation of the exchange integral (as required when using exchange-correlation functionals of the hybrid type, e.g. B3LYP) very expensive. Due to this reason the functionals used in the context of first-principles MD are mostly of the LDA or GGA type. As we will see below, plane waves and PBCs also necessitate a special approach for the calculation of response properties within DFPT. Plane waves, however, also have a number of advantages, e.g. the convergence with regard to the size of the basis set can be probed systematically, simply by increasing the kinetic energy cutoff, and most calculations are performed essentially at the basis set limit. Plane waves are, furthermore, a space-fixed basis (as opposed to an atom-centered basis) and are therefore free of basis set superposition errors. Last, but not least, plane waves enable efficient calculation of the kinetic energy term (which is diagonal in reciprocal space) and the Hartree term by applying fast Fourier transform techniques.

Besides the traditional scheme, AIMD using semiempirical [65, 66], Hartree-Fock [67–70], generalized valence bond (GVB) [71], complete active space (CASSCF) [72], and configuration interaction (CI) [73] electronic structure methods have been realized. Several different variations of the basis set have also been implemented, including extensions to projector augmented [74] and generalized (adaptive) plane waves [75]. A hybrid basis set of atom-centered basis functions and (augmented) plane waves [76, 77] and an all-electron version of this [78] have also been introduced recently. First steps towards a dynamic treatment of excited states within DFT have been made through a grand-canonical Car-Parinello method [33] based on the finite temperature formulation of Mermin [79] and by a dynamic approach for excited singlet states

[27], and by a linear response formulation of TDDFT [18]. Recently, the Car-Parrinello scheme has also been extended into a mixed quantum/classical QM/MM approach [10–12].

1.2.2

***Ab initio* Molecular Dynamics Programs**

Several groups have implemented their own *ab initio* molecular dynamics programs. Tab. 1.2 lists some of the most frequently used software:

Tab. 1.2 Some currently available computer programs with *ab initio* molecular dynamics capabilities.

Program name	Ref.	
ABINIT	80 a	www.abinit.org
CASTEP	80 b	Molecular Simulations Inc.
CPMD	81	M. Parrinello, CSCS Manno, Switzerland and IBM Zurich Research Laboratory, Switzerland (www.cpmid.org)
Fhi98md	82	Fritz-Haber Institute Berlin, Germany fhim@fhi-berlin.mpg.de
JEEP		François Gygi, Lawrence Livermore National Laboratory, USA
NWCHEM		Pacific Northwest National Laboratory, USA
PAW	62	P.E. Blöchl, Clausthal University of Technology, Germany
SIESTA	83	P. Ordejon, Institut de Ciencia de Materials de Barcelona, Barcelona, Spain
VASP	84 a	J. Hafner, University of Vienna, Austria

1.3

Mixed Quantum Mechanical/Molecular Mechanical (QM/MM) Car-Parinello Simulations

By use of parallel computers Car-Parinello simulations can currently be performed for systems containing a few hundred to a few thousand atoms. However impressive, this is still too limited a size for most biologically relevant applications. One possible solution for the modeling of systems of several tens of thousands

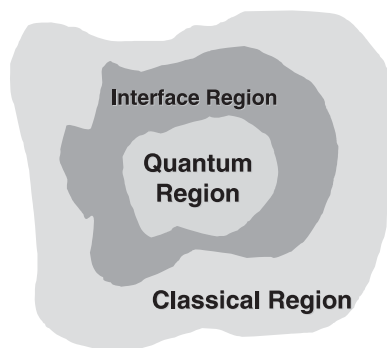


Fig. 1.2 QM/MM partitioning of a system. The quantum region (QM) is treated with a quantum chemical electronic structure method whereas the surroundings are taken into account in the framework of a classical force field

(MM). The interface region between QM and MM part of the system might be taken into account in a special way to smoothen the transition between the two rather disparate approaches.

to several hundred thousand atoms is the choice of a hierarchical hybrid approach in which the whole system is partitioned into a localized chemically active region (treated by use of a quantum mechanical method) and its environment (treated with empirical potentials). This is the so-called quantum mechanical/molecular mechanics (QM/MM) method [84 b]. In a QM/MM approach the computational effort can be concentrated on the part of the system where it is most needed whereas the effects of the surroundings are taken into account with a more expedient model (Fig. 1.2).

The Car-Parinello method can be extended into a mixed QM/MM scheme by use of a mixed Lagrangian of the form [10–12]:

$$L = \frac{1}{2} \mu \sum_i \int d\vec{r} \dot{\psi}_i^*(\vec{r}) \dot{\psi}_i(\vec{r}) + \frac{1}{2} \sum_I M_I \dot{R}_I^2 - E_{MM} - E_{QM/MM} - E_{QM} + \sum_{i,j} A_{i,j} (\int d\vec{r} \psi_i^*(\vec{r}) \psi_j(\vec{r}) - \delta_{i,j})$$

where the potential energy terms E_{MM} , $E_{QM/MM}$, and E_{QM} refer to the classical part, the interaction between the QM and MM parts, and the energy of the QM system given by the Kohn-Sham energy-density functional:

$$E_{K_s}[\psi_i, \vec{R}_I] = -\frac{1}{2} \sum_I \int d\vec{r} \psi_i^*(\vec{r}) \nabla \psi_i(\vec{r}) + \int d\vec{r} V_N(\vec{r}) \rho(\vec{r}) \\ + \frac{1}{2} \int d\vec{r} d\vec{r}' \rho(\vec{r}) \frac{1}{|\vec{r} - \vec{r}'|} \rho(\vec{r}') + E_{xc}[\rho(\vec{r})]$$

where $V_N(\vec{r})$ is the external potential, $E_{xc}[\rho(\vec{r})]$ the exchange-correlation functional and the electron density $\rho(\vec{r})$ is given by the sum of the densities of the occupied one-particle states:

$$\rho(\vec{r}) = 2 \sum_i \psi_i^*(\vec{r}) \psi_i(\vec{r})$$

The purely classical part, E_{MM} , is described by a standard biomolecular force field:

$$E_{MM} = E_{MM}^{\text{bonded}} + E_{MM}^{\text{nonbonded}}$$

where E_{MM}^{bonded} and $E_{MM}^{\text{nonbonded}}$ are of the general form:

$$E_{MM}^{\text{bonded}} = \sum_b \frac{1}{2} k_b (r_{ij} - b_0)^2 + \sum_\theta \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2 \\ + \sum_\varphi \sum_n k_n [1 + \cos(n\varphi_{ijkl} - \varphi_0)]$$

$$E_{MM}^{\text{nonbonded}} = \sum_{lm} \frac{q_l q_m}{4\pi\epsilon_0 r_{lm}} + \sum_{op} 4\epsilon_{op} \left(\left(\frac{\sigma_{op}}{r_{op}} \right)^{12} - \left(\frac{\sigma_{op}}{r_{op}} \right)^6 \right)$$

The terms in E_{MM}^{bonded} take into account harmonic bond, angle and dihedral terms and those in $E_{MM}^{\text{nonbonded}}$ electrostatic point charge and van der Waals interactions.

The intricacies of QM/MM methods lie in the challenge of finding an appropriate treatment for the coupling between QM and MM regions as described by the interaction $E_{QM/MM}$. Special care must be taken that the QM/MM interface (Fig. 1.3) is described accurately and consistently, in particular in combination with a plane wave-based Car-Parinello scheme. In the fully Hamiltonian coupling scheme developed by our group [10–12], bonds between the QM and MM parts of the system are treated with specifically designed monovalent pseudo potentials, whereas the remaining bonding interactions of the interface region, i.e. angle bending and dihedral distortions, are described on the lev-

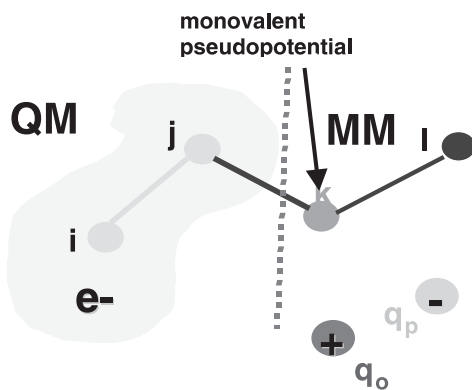


Fig. 1.3 Schematic representation of the interface region between QM and MM parts.

el of the classical force field. The same holds for the van der Waals interactions between the QM and MM parts of the system.

On the other hand, the electrostatic effects of the classical environment are taken into account in the quantum mechanical description as an additional contribution to the external field of the quantum system

$$E_{\text{QM/MM}}^{\text{ele}} = \sum_{i \in \text{MM}} q_i \int dr \rho(r) v_i(|r - r_i|)$$

where q_i is the classical point charge located at r_i and $v_i(|r - r_i|)$ is a Coulombic interaction potential modified at short-range in such a way as to avoid spill-out of the electron density to nearby positively charged classical point charges [10]. In the context of a plane wave-based Car-Parinello scheme direct evaluation of this equation is prohibitive, because it involves in the order of $N_r \times N_{\text{MM}}$ operations, where N_r is the number of real space grid points (typically ca. 100^3) and N_{MM} is the number of classical atoms (usually of the order of 10,000 or more in systems of biochemical relevance). Therefore, the interaction between the QM system and the more distant MM atoms is included via a Hamiltonian term explicitly coupling the multipole moments of the quantum charge distribution with the classical point charges. This QM/MM Car-Parinello implementation establishes an interface between the Car-Parinello code CPMD [81] and the classical force fields GROMOS96 [85] and AMBER95 [86] in combination with a particle-particle-particle mesh (P3M) treatment of the long-range electrostatic interactions [87].

In this way, efficient and consistent QM/MM Car-Parinello simulations of complex extended systems of several tens of thousands to several hundred thousand atoms can be performed in which the steric and electrostatic effects of the surrounding are taken explicitly into account.

1.3.1

Gly-Ala Dipeptide in Aqueous Solution – Polarizable versus Nonpolarizable Force Fields

As a simple example of a QM/MM Car-Parinello study, we present here results from a mixed simulation of the zwitterionic form of Gly-Ala dipeptide in aqueous solution [12]. In this case, the dipeptide itself was described at the DFT (BLYP [88, 89 a]) level in a classical solvent of SPC water molecules [89 b]. The quantum solute was placed in a periodically repeated simple cubic box of edge 21 au and the one-particle wavefunctions were expanded in plane waves up to a kinetic energy cutoff of 70 Ry. After initial equilibration, a simulation at 300 K was performed for 10 ps.

As a first comparison between the QM/MM description of the dipeptide and its analogs in terms of standard biomolecular force fields we compared the quantum electrostatic field with the equivalent quantities resulting from purely classical MD runs with the AMBER95 [86] and the GROMOS96 [85] force fields. For thirty-six configurations of the QM/MM trajectory, an optimum set of atomic point charges was fitted to the quantum electrostatic field by applying the recently developed technique of dynamically generated restrained electrostatic potential-derived charges (D-RESP charges) [12]. The resulting values are compared in Tab. 1.3 with the corresponding Hirshfeld [89 c], AMBER [86], and GROMOS [85] values.

If separate D-RESP charge sets are fitted for every single one of the 36 frames, the standard deviation of the electrostatic field generated varies between 3.5 and 5% with respect to the full quantum reference. This accuracy is the best (in the least-squares sense) that can be obtained if the system is modeled with *time-dependent* atomic point-charges and represents the accuracy limit for a fluctuating point charge model of the dipeptide.

In order to quantify the importance of polarization effects in describing the electrostatic field of the system we compared this

Tab. 1.3 Comparison between different sets of atomic point charges for a zwitterionic Gly-Ala dipeptide in aqueous solution. D-RESP: electrostatic potential derived charges [12] fitted to all 36 configurations. Hirshfeld: average value of the Hirshfeld charges [89c] along the full trajectory, Amber: AMBER 1995 force field [86], Gromos: GROMOS96 force field [85]. The charges of equivalent atoms are imposed to be equal.

	D-RESP	Hirshfeld	AMBER	GROMOS
HN ^{ter}	0.226	0.202	0.164	0.248
N ^{ter}	0.165	0.152	0.294	0.129
Ca	0.043	0.028	-0.010	0.000
Ha	0.090	0.066	0.089	0.000
C	0.082	0.098	0.616	0.380
O	-0.503	-0.265	-0.572	-0.380
N	0.117	0.031	-0.382	-0.280
H	0.257	0.100	0.268	0.280
Ca	0.096	0.038	-0.175	0.000
Ha	0.048	-0.014	0.107	0.000
C β	-0.049	-0.067	-0.209	0.000
H β	0.035	0.008	0.076	0.000
C ^{ter}	-0.023	-0.032	0.773	0.270
O ^{ter}	-0.599	-0.416	-0.806	-0.635

result with the standard deviation obtained using a single *time-independent* set of charges to reproduce the field of all the 36 configurations. The standard deviation of the electrostatic field of this optimal “non-polarizable” model varies between 5.5 to 7%, indicating that the gain in accuracy that can be obtained for this system using a polarizable model would be rather small (only 2–3% improvement). It should be noted that the above analysis has been performed on conformations extracted from a 10-ps trajectory, and that the system during this time does not undergo significant conformational change. We repeated the same analysis on eight configurations belonging to different conformers (obtained from an enhanced sampling [90] classical MD run generating several dihedral transitions with successive 0.3 ps of QM/MM thermalization). In this case, the standard deviation ranges between 3 and 7% for *time-dependent* point charges, whereas the corresponding quantity for a *time-independent* charge set ranges between 4 and 14%, indicating some configurational dependence of the charge distribution.

It is also interesting to compare the accuracy of the D-RESP [12] point charges with the standard deviation of the field obtained using atomic charges from the GROMOS96 [85] and AMBER 1995 [86] force fields. Remarkably, standard non-polarizable force-field models can reproduce the field all along the MD trajectory very well. The standard deviation for the MD trajectory exploring only one conformer is between 6 and 13% for the AMBER charge set and only slightly higher, between 9 and 16%, for the GROMOS charge set, which is a united atom model and hence includes fewer degrees of freedom to reproduce the field. We consider this a first indication of the consistency between a density-functional theory-based QM/MM approach and the point charge representation of the electrostatic field used in standard biomolecular force fields.

The QM dipole moment of the Gly-Ala dipeptide is, furthermore, reproduced by the D-RESP set within 2%. The predictivity of any D-RESP set on the dipole along the full trajectory is approximately 4%. This compares with predictivities of $\sim 6\%$ and $\sim 7\%$ for the dipoles computed with the AMBER 1995 and GROMOS96 force field charges, respectively.

1.4

Density-functional Perturbation Theory and the Calculation of Response Properties

1.4.1

Introduction to Density-functional Perturbation Theory

Study of the quantum mechanical ground state and its properties can provide much information about the physics of a system. DFT has been used with great success for a computational modeling of the properties of a wide range of systems of chemical [7] and, increasingly, also biological [8] relevance. DFT electronic structure calculations have been shown to give a realistic description of molecular systems from first principles, i.e. without adjustable parameters. Many experimentally accessible quantities are, however, related to excited electronic states, whose description requires further theoretical effort. Density functional perturbation theory (DFPT) is increasingly used to obtain such

data with moderate computational effort. In this section the basic ideas of DFPT are outlined, and several examples of applications are given to illustrate the power of this method.

Most spectroscopic properties are related to second derivatives of the total energy. As a simple illustrative example, vibrational modes, which arise from the harmonic oscillations of atoms around their equilibrium positions, are characterized by the quadratic variation of the total energy as a function of the atomic displacements δR_I :

$$\delta E[\delta R_I] = \sum_{I,J} \sum_{a,\beta} \delta R_{I,a} H_{I,a,J,\beta} \delta R_{J,\beta}$$

Here I and J are indices for numbering atoms, and a and β denote Cartesian coordinates x, y, z . The so-called Hessian matrix $H_{I,a,J,\beta}$ is the second derivative of the total energy

$$H_{I,a,J,\beta} = \frac{\partial^2 E^{\text{tot}}}{\partial R_{Ia} \partial R_{J\beta}}$$

From the eigenvalues of this matrix, the harmonic vibrational frequencies can be obtained, and the corresponding eigenvectors describe the vibrational modes.

Many second-order quantities like this Hessian matrix involve the response of the electronic structure of the system, and their accurate calculation is difficult in standard DFT. Instead, DFPT as outlined below provides a powerful tool enabling access to these derivatives with moderate computational effort but very good precision.

In standard DFT, the variational principle is used to obtain the electronic structure of the ground state of a system. The variational principle states that the true ground state of a system is the one which minimizes its total energy. In perturbation theory, this variational approach is transferred to the *change* in electronic structure arising from the presence of an external perturbation, where an additional potential slightly modifies the energy landscape. This change in the electronic wavefunctions will be such that it leads to a new minimum of the total energy. In other words, the electrons adapt to the presence of the perturbation potential by adjusting their orbitals, in order to reach the new minimum energy wavefunction.

1.4.2

Basic Equations of Density-functional Perturbation Theory

The starting point is the Kohn-Sham energy functional:

$$E^{\text{KS}} = -\frac{1}{2} \sum_i \int dr \psi_i(r) \nabla^2 \psi_i(r) + \frac{1}{2} \int dr dr' \frac{\rho(r)\rho(r')}{|r-r'|} \\ + E_{\text{xc}}[\rho(r)] + \int dr V(r)\rho(r)$$

with the electronic density:

$$\rho(r) = \sum_i |\psi_i(r)|^2$$

the exchange-correlation energy functional E_{xc} and the external potential $V(r)$ created by the ions.

According to the variational principle, the ground state of the system is described by those electronic wavefunctions ψ_i which minimize the Kohn-Sham functional. The presence of an external perturbation is represented by a perturbation functional, E^{P} , that is added to the unperturbed Kohn-Sham functional:

$$E^{\text{tot}} = E^{\text{EE}} + \lambda E^{\text{P}}$$

Here, λ is a differentially small parameter which quantifies the strength of the external perturbation, but which will not appear in the final equations. It serves only to separate terms of different magnitudes of the resulting expressions. For instance, a sum of a term which contains a part linear in λ and one quadratic in λ can be simplified – because λ is supposed to be differentially small, the quadratic term will be infinitely smaller than the linear term, so it can be neglected. If, however, the linear term happens to be zero, the quadratic term must be taken into account.

In principle, the new orbitals could be calculated by repeating the energy minimization for the new energy functional E^{tot} with a small but non-zero perturbation strength λ . This so-called *finite-difference* method has the advantage of a relatively straightforward implementation, but also presents several drawbacks. Its accuracy is limited and sensitive to the choice of the parameter λ . Further, some types of perturbation, like magnetic fields, significantly increase the computational cost of the calculation, because they break certain symmetries of the system.

Instead, a more elegant way consists in separating analytically the energy contribution resulting from the perturbation from the unperturbed parts. This method is numerically more stable and also results in more accurate determination of the system's reaction to the perturbation.

The electronic orbitals that minimize the total perturbed energy functional E^{tot} will be different from those obtained for the unperturbed functional E^{KS} . Because the perturbation is only differentially small, however, they will be arbitrarily close to the old ones. Denoting the unperturbed orbitals by $\psi^{(0)}$ and the new ones (which will minimize E^{tot}) by ψ^{tot} , the latter can thus be expanded in a power series of λ :

$$\psi_i^{\text{tot}}(r) = \psi_i^{(0)}(r) + \lambda\psi_i^{(1)}(r) + \dots$$

The variation ψ in results in a change in the charge density:

$$\rho^{\text{tot}}(r) = \rho^{(0)}(r) + \lambda\rho^{(1)}(r) + \dots$$

where the perturbation density $\rho^{(1)}$ can be derived from the perturbed electronic orbitals:

$$\rho^{(1)}(r) = \sum_i \bar{\psi}_i^{(0)}(r)\psi_i^{(1)}(r) + \bar{\psi}_i^{(1)}(r)\psi_i^{(0)}(r)$$

The orbitals $\psi^{(1)}$ and the density $\rho^{(1)}$ represent the reaction of the electronic system to the external perturbation. Because, in both cases, the correction due to the perturbation is limited to the linear terms $\psi^{(1)}$ and $\rho^{(1)}$, this approach is also called linear response theory. Consequently, $\psi^{(1)}$ and $\rho^{(1)}$ are called linear response orbitals and density, respectively.

In complete analogy with this procedure, the energy functional can be expanded in powers of the differential perturbation parameter λ :

$$\begin{aligned} E^{\text{tot}} &= E^{\text{tot}}[\psi_i^{(0)} + \lambda\psi_i^{(1)} + K] \\ &= E^{\text{EE}}[\psi_i^{(0)} + \lambda\psi_i^{(1)} + K] + \lambda E^{\text{P}}[\psi_i^{(0)} + \lambda\psi_i^{(1)} + K] \\ &= E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + K \end{aligned}$$

It can be shown that the linear term $E^{(1)}$ vanishes if the variational principle is satisfied for the original orbitals $\psi^{(0)}$, so that

the subsequent $E^{(2)}$ must also be taken into account. The idea of the perturbation theory method is that for the newly perturbed situation the variational principle must still hold. Because the unperturbed orbitals $\psi^{(0)}$ are already a good approximation to the final electronic structure, they are held frozen, and only the perturbation-dependent electronic response orbitals $\psi^{(1)}$ will be allowed to change. To fulfil the variational principle, they must minimize the above expression, which corresponds to a stationary point of E^{tot} with respect to $\psi^{(1)}$:

$$\frac{\delta E^{\text{tot}}}{\delta \psi_i^{(1)}(r)} = 0$$

The calculation itself is somewhat lengthy, since it involves second derivatives of the Kohn-Sham functional with respect to the orbitals, and does not provide much insight into the physics of the problem. We therefore refer the interested reader to related references [13, 91]. The final stationarity equation reads:

$$(\hat{H}^{\text{KS}} - \varepsilon_i)\psi_i^{(1)} + (\int dr' K(\hat{r}, r')\rho^{(1)}(r')\psi_i^{(0)}) = -\frac{\delta E^p}{\delta \psi_i} \Big|_{\psi=\psi^{(0)}}$$

with the Kohn-Sham Hamiltonian \hat{H}^{KS} and its eigenvalues ε_i :

$$\hat{H}^{\text{KS}} = -\frac{1}{2}\nabla^2 + \int dr' \frac{\rho^{(0)}(r')}{|\hat{r} - r'|} + v_{\text{xc}}[\rho^{(0)}](\hat{r}) + v_{\text{ion}}(\hat{r})$$

$$\varepsilon_i = \langle \psi_i | \hat{H}^{\text{KS}} | \psi_i \rangle$$

Here, $\rho^{(0)}$ is the unperturbed density, v_{xc} the exchange-correlation potential and v_{ion} represents the atomic potential. $K(r, r')$ is the Hartree-exchange-correlation kernel, defined by:

$$K(r, r') = \frac{1}{|r - r'|} + \frac{\delta v_{\text{xc}}[\rho](r)}{\delta \rho(r')}$$

The set of Kohn-Sham-like linear equations above represents the working equations of DFPT. They are usually solved by iterative linear algebra algorithms (conjugate-gradient minimization).

1.4.3

NMR Chemical Shieldings within DFPT**1.4.3.1 Introduction to Nuclear Magnetic Resonance
Chemical Shifts**

When a magnetic field is applied to a medium it interacts with the electrons in the system. The electrons respond by creating a current distribution which itself induces an additional magnetic field. Even if the external field is homogeneous in space, the induced one is not so, the sum of both fields is inhomogeneous as well. The actual spatial shape of the induced field depends very sensitively on the electronic structure in the system. In particular, the amplitudes at the atomic positions of non-equivalent nuclei are different. Since this amplitude determines the resonance frequency of the nuclear spin, the electronic structure of a system can be analyzed in great detail by probing the spectrum of its nuclear spins. This technique is called nuclear magnetic resonance (NMR) spectroscopy.

The simplest experimental arrangement to measure such a spectrum consists of irradiating with a radio frequency, and varying this frequency until resonance occurs. This setup is called sweeping or continuous-wave technique. Although nowadays it is used only in tutorials, it nevertheless represents the fundamental resonance experiment. Modern spectrometers instead measure with the Fourier spectroscopy technique in which the sample is submitted to a pulse of large frequency bandwidth. This pulse turns the nuclear spins into a plane orthogonal to the homogeneous external field, which creates a precession motion. Because the angular velocity of this precession depends on the strength of the local magnetic field, which is inhomogeneous, the individual spins have different precession frequencies.

This effect induces a free induction decay (FID) signal in the detection circuit. The FID can be measured, and the normal absorption spectrum can be obtained by means of an inverse Fourier transform. A variety of experimental extensions have been developed for this approach. By means of particular pulse sequences it is possible to detect spin resonances selectively on the basis of a broad ensemble of properties such as spatial proximity and dipolar coupling strengths. The central fundamental quantity of interest is, however, still the energy spectrum of the nuclear spin,

which manifests itself in its resonance frequency. This frequency can be translated into the dimensionless proportionality factor between external and induced magnetic fields, which is called chemical shielding. The difference between the chemical shielding of a sample and that of a known reference substance is called chemical shift.

Theoretical calculation of NMR chemical shifts is usually done by first considering the electronic current density which is induced by the external magnetic field. Once this current has been calculated the chemical shift can be obtained by application of the Biot-Savard law, which describes the magnetic field created by it. The strength of this field at the position of an atom represents the NMR chemical shielding of this atom.

1.4.3.2 NMR Chemical Shieldings

We consider a molecular system, placed in a strong homogeneous magnetic field B^{ext} . The total magnetic field throughout space will be equal to this external field plus a small correction B^{ind} induced by the system's electronic structure, which will be inhomogeneous in space:

$$B^{\text{tot}}(R) = B^{\text{ext}} + B^{\text{ind}}(R)$$

Under the assumption that the induced field is much smaller than the external one, and that the latter is still small on the atomic energy scale, the induced field will be proportional to the external field. Both assumptions are valid for experimentally accessible field strengths, i.e. up to more than 10 Tesla. The proportionality coefficient between them depends on the chosen position in space, and is called chemical shielding, commonly denoted by $\sigma_{a\beta}(R)$:

$$B_a^{\text{ind}}(R) = - \sum_{\beta} \sigma_{a\beta}(R) B_{\beta}^{\text{ext}} + \dots$$

where a and β denote Cartesian indices. The chemical shielding is a tensor quantity, which means that an external field in a certain direction will induce a field B^{ind} that may also have components in other directions. Thus, B^{ext} and B^{ind} are not necessarily parallel. However, the main component normally is, and most ex-

periments measure only the trace of the shielding tensor, i.e. its orientational average. In the linear regime, the total magnetic field reads:

$$B^{\text{tot}}(R) = (1 - \sigma(R))B^{\text{ext}} + \dots$$

and the chemical shielding can formally be written as the negative derivative of the electronically induced field with respect to the externally applied one, taken at the position of the atomic nucleus considered, R :

$$\sigma_{\alpha\beta}(R) = -\frac{\partial B_{\alpha}^{\text{ind}}(R)}{\partial B_{\beta}^{\text{ext}}}$$

The goal of the calculation is, therefore, to compute the strength of the induced magnetic field in space, relative to the strength of the external one. Because one of the main assumptions is that all fields are small on the typical energy scale of atoms or molecules, this calculation can be performed in the framework of perturbation theory.

Magnetic fields cannot be incorporated directly into the Hamiltonian of the system by means of their field strength, B . Instead, they must be transformed into a vector potential $A(r)$ which has to satisfy the relationship:

$$B(r) = \nabla \times A(r)$$

For periodic systems special care must be taken, because it is not possible to define properly a vector potential for a constant magnetic field in an infinite system. In finite systems, however, this problem does not exist, and a given magnetic field B can be described by a vector potential:

$$A(r) = -\frac{1}{2}(r - R_{\text{g}}) \times B$$

where R_{g} is the so-called gauge origin. Its choice is theoretically not relevant, but it turns out that the accuracy of actual calculations is strongly linked to the value of R_{g} . A badly chosen gauge can result in significant deterioration of the final results. This problem is due to the use of finite basis sets. Several methods have been developed in the quantum chemistry community to treat this problem, of these

the IGLO, GIAO, and CSGT approaches [92] have been shown to be of comparable quality. In the current implementation the CSGT method is used exclusively. The exact formalism of the actual implementation is given elsewhere [14].

This vector field can be inserted directly into the quantum mechanical Hamiltonian by a variable change in the momentum operator:

$$\hat{p} \rightarrow \hat{p} - \frac{1}{c} A(\hat{r})$$

where c is the speed of light. Inserted into the DFT energy functional, this change leads to a perturbation functional:

$$E^P = -\frac{1}{2c} [\langle \psi | \hat{p} \cdot A(\hat{r}) | \psi \rangle + \langle \psi | A(\hat{r}) \cdot \hat{p} | \psi \rangle] + \frac{1}{2c^2} \langle \psi | A(\hat{r}) \cdot A(\hat{r}) | \psi \rangle$$

This functional can directly be used in the DFPT equations above. Further, for the special case of a magnetic perturbation it can be shown that the response density $\rho^{(1)}$ analytically vanishes, making the calculation significantly easier. The working equations simplify to:

$$(\hat{H}^{\text{KS}} - \varepsilon_i) \psi_i^{(1)} = - \left. \frac{\delta E^P}{\delta \psi_i} \right|_{\psi = \psi^{(0)}}$$

After solving this system of equations for the response orbitals $\psi^{(1)}$, the induced electronic ring current density is obtained from the $\psi^{(1)}$, using the expectation value of the standard quantum mechanical current density operator:

$$j(r) = \frac{1}{2c} \sum_i [\bar{\psi}_i^{(1)} \nabla \psi_i^{(0)} + \bar{\psi}_i^{(0)} \nabla \psi_i^{(1)}] + \rho(r) A(r)$$

When the electronic current density $j(r)$ has been calculated, the Biot-Savart law for the induced magnetic field reads:

$$B^{\text{ind}}(R) = \int dr \frac{r - R}{|r - R|^3} \times j(r)$$

1.4.3.3 Calculation of NMR Chemical Shifts in QM/MM Car-Parinello Simulations

When NMR chemical shielding tensors are calculated in a QM/MM embedding scheme, it must be ensured that the quantum region contains all particles which are in chemical contact with the atoms of interest. For instance, although the overall properties of H-bonded systems are generally well reproduced in our QM/MM approach [10], the subtle electronic effects induced on the chemical shielding, are unlikely to be correctly reproduced by the simple point charge representation of an empirical, non-polarizable force field. When considering a hydrogen atom in a molecule that forms such a bond, the corresponding donor atom and its nearest neighbors still interact with the proton in a genuinely quantum mechanical way. In such circumstances replacing the donor atoms by point charges yields only a qualitative value for the proton NMR chemical shielding. Instead, atoms and molecules which are further apart from the region of interest can safely be treated by the force field description. Their effect on the quantum part is still significant, but occurs rather on the level of polarization, which can be reproduced well by a simple charge model.

To determine the required size of the quantum region in more detail, the shielding of a typical extended system was investigated within the QM/MM framework. The system we chose is liquid water, represented by a periodic box containing 64 water molecules. The geometry was taken from an *ab initio* Car-Parinello-type molecular dynamics simulation at ambient conditions [93]. From this single configuration, isotropic NMR shielding constants of all 128 protons have been calculated in the fully periodic quantum mechanical framework according to Ref. [14], in order to obtain full QM reference values.

The chemical shieldings were then recalculated in this same system using the QM/MM method [10]. To this end, each molecule was considered individually. The water molecule of interest and its first solvation shell were treated quantum mechanically, whereas the surrounding water molecules were taken into account with an empirical force field representation (MM molecules). The first solvation shell was defined via a distance criterion on the oxygen–oxygen distance. As a threshold, the first minimum of the O–O pair correlation function was taken; this occurs at 3.5 Å [93]. All

water molecules with oxygen atoms within this distance of the oxygen of the molecule of interest were included in the QM region. All the remaining molecules were represented by MM atoms, using the coordinates from the *ab initio* molecular dynamics snapshot. Standard partial charges of $+0.4 e$ and $-0.8 e$ were assigned to the hydrogen and oxygen atoms, respectively.

These QM/MM calculations are in contrast to a standard evaluation of chemical shielding for gas phase water clusters where the classical point charge environment is omitted entirely. The same solvation shell criterion as before was applied, and the system was treated purely quantum mechanically.

The resulting data are shown in Fig. 1.4, in which is plotted the isotropic NMR chemical shift of all 128 protons, obtained from the QM/MM and the isolated cluster calculations as a function of the fully periodical quantum mechanical results.

Figure 1.4 shows a significant deviation between the isolated cluster calculations and the full calculation. The situation is, however, considerably improved by the presence of the classical point charges in the QM/MM calculation. Here the whole bandwidth of chemical shielding constants is present, and correlation with the reference values is excellent.

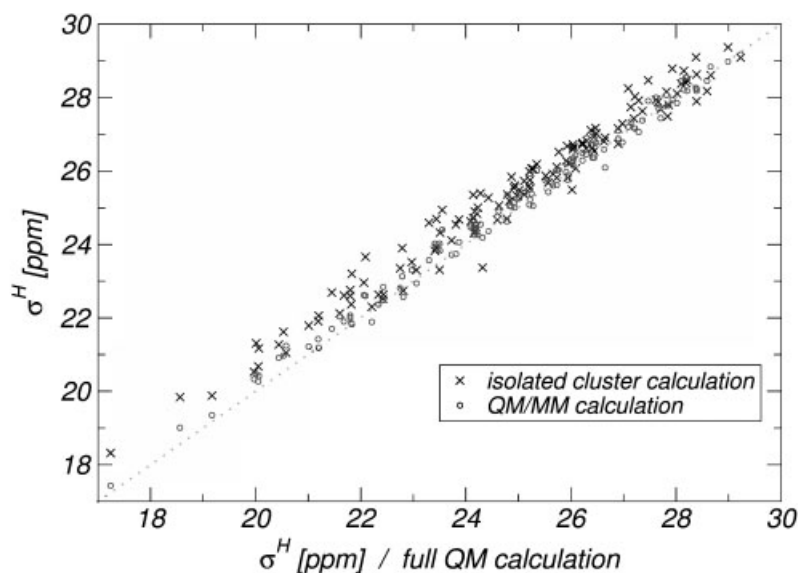


Fig. 1.4 Proton NMR chemical shieldings of liquid water in the

QM/MM approach – comparison with a simple cluster calculation.

This shows that the second and further solvation shells still have a non-negligible effect on NMR chemical shielding constants through the long-range electrical field they create. The approximation of an isolated molecular cluster *in vacuo* is valid for large clusters only; this eventually makes determination of the shieldings of all protons computationally much more expensive than the fully periodic *ab initio* calculation.

For a water molecule with its first solvation shell only, chemical shielding differs considerably from the complete *ab initio* calculation. Although this discrepancy is not completely removed by the inclusion of the electrical field of the remaining molecules through the QM/MM approach, it is strongly reduced.

1.5

Introduction to Time-dependent Density-functional Theory (TDDFT)

One important application of TDDFT is to compute low lying excited electronic states and energies. Simpler approaches, in which the virtual ground state Kohn-Sham orbitals and energies are determined as an estimate for excited states are often not sufficiently accurate for chemical applications and can only be used as a rather qualitative indication.

There are several possible ways of deriving the equations for TDDFT. The most natural way departs from density-functional perturbation theory as outlined above. Initially it is assumed that an external perturbation is applied, which oscillates at a frequency ω . The linear response of the system is then computed, which will be oscillating with the same imposed frequency ω . In contrast with the standard static formulation of DFPT, there will be special frequencies ω_v for which the solutions of the perturbation theory equations will persist even when the external field vanishes. These particular solutions for orbitals and frequencies describe excited electronic states and energies with very good accuracy.

1.5.1

Basic Equations of TDDFT

For description of a time-dependent problem, the Kohn-Sham equations:

$$(H^{\text{KS}} - \varepsilon_i)\psi_i = 0$$

must be modified. By analogy with the time-dependent Schrödinger equation, they transform into:

$$(H^{\text{KS}} - \varepsilon_i)\psi_i = i\frac{\partial}{\partial t}\psi_i$$

For a stationary (time-independent) set of wavefunctions, the time derivative on the right hand side vanishes and the equation obviously transforms back into the previous one. Now the system is perturbed by an external perturbation \hat{H}^{P} , which is chosen to have a well-defined temporal periodicity:

$$\hat{H}^{\text{P}} = \lambda\hat{h}^{\text{P}}(e^{-i\omega t} + e^{+i\omega t})$$

The actual form of the Hamiltonian operator \hat{h}^{P} does not have to be defined at this moment. As in standard perturbation theory, it is assumed that the solution of the electronic structure problem of the combined Hamiltonian $\hat{H}^{\text{KS}} + \hat{H}^{\text{P}}$ can be described as the solution $\psi^{(0)}$ of \hat{H}^{KS} , corrected by a small additional linear-response wavefunction $\lambda\psi^{(1)}$. Only these response orbitals will explicitly depend on time – they will follow the oscillations of the external perturbation and adopt its time dependency. Thus, the following Ansatz is made for the solution of the perturbed Hamiltonian $\hat{H}^{\text{KS}} + \hat{H}^{\text{P}}$:

$$\psi_i(r, t) = \psi_i^{(0)}(r) + \lambda(\psi_i^{(1-)}(r)e^{-i\omega t} + \psi_i^{(1+)}(r)e^{+i\omega t})$$

Because of the separation into a time-independent unperturbed wavefunction and a time-dependent perturbation correction, the time derivative on the right-hand side of the time-dependent Kohn-Sham equation will act only on the response orbitals. From this perturbed wavefunction the first-order response density follows as:

$$\begin{aligned}\rho_{\text{TDDFT}}^{\text{tot}}(r, t) &= \rho^{(0)}(r) + \lambda\rho^{(1)}(r, t) + \dots \\ \rho^{(1)}(r, t) &= \sum_i (\psi_i^{(1-)} \bar{\psi}_i^{(0)} + \psi_i^{(0)} \bar{\psi}_i^{(1+)}) e^{-i\omega t} \\ &\quad + (\psi_i^{(1+)} \bar{\psi}_i^{(0)} + \psi_i^{(0)} \bar{\psi}_i^{(1-)}) e^{+i\omega t} \\ &= \rho^{(1-)}(r) e^{-i\omega t} + \bar{\rho}^{(1-)}(r) e^{+i\omega t}\end{aligned}$$

which ensures that the response density remains a real number. After insertion of the Ansatz for ψ and ρ into the density functional perturbation equations, they can be separated into terms in $e^{-i\omega t}$ and $e^{+i\omega t}$ which are then solved separately. Following the procedure of standard DFPT for the part in $e^{-i\omega t}$, one obtains basically the same working equations, containing an additional term:

$$(\hat{H}^{\text{KS}} - \varepsilon_i) \psi_i^{(1-)} + (\int dr' K(\hat{r}, r') \rho^{(1-)}(r')) \psi_i^{(0)} = -\hat{h}^{\text{P}} \psi_i^{(0)} + \omega \psi_i^{(1-)}$$

This equation describes the electronic reaction to the oscillating external perturbation. In principle, it has a solution for any frequency ω . One special class of solutions is, however, of particular interest. If at a frequency ω , there is a solution $\psi^{(1-)} = \psi^{(\omega)}$ that satisfies the above equation also at zero perturbation strength ($\hat{h}^{\text{P}} = 0$), then the unperturbed system will also be stable in the state described by this particular solution $\psi^{(0)} + \lambda\psi^{(\omega)}$. In such circumstances the term λ is no longer bound to the perturbation strength. Instead, it can take any value, as long as it is sufficiently small to remain in the linear response region of the system. Such a new state, however, is nothing but an excited state.

Thus, the goal is to find frequencies ω and corresponding orbitals $\psi^{(\omega)}$ for which the DFPT equation without the external perturbation Hamiltonian:

$$(\hat{H}^{\text{KS}} - \varepsilon_i) \psi_i^{(\omega)} + (\int dr' K(\hat{r}, r') \rho^{(\omega)}(r')) \psi_i^{(0)} = \omega \psi_i^{(\omega)}$$

is valid. This equation is an eigenvalue problem for the $\{\omega, \psi^{(\omega)}\}$. When the ground-state wavefunctions $\psi^{(0)}$ are known, it can be solved by means of iterative diagonalization schemes (Davidson algorithm). The resulting eigenvalues ω represent the vertical excitation energies of the system.

1.5.2

Applications of TDDFT within the QM/MM Framework – Solvent Shift of the S_0/S_1 Transition of Acetone in Water

The combination of TDDFT with a QM/MM approach is, in principle, straightforward. The surrounding system of point charges modifies the electrostatic potential of the system, which enters the perturbation equations through the Kohn-Sham Hamiltonian \hat{H}^{KS} . This causes a change in the excitation wavelenghts which reflects the influence of the environment.

As a first test system chosen to probe the capabilities of a QM/MM approach in reproducing environment-induced shifts in excitation spectra, we have calculated the solvent shift of the first optical excitation ($n \rightarrow \pi^*$) of an acetone molecule in the gas phase and in water (Fig. 1.5) [16].

The vertical excitation energies were calculated for different configurations of a QM/MM trajectory using the approximate ROKS [27] method as well as TDDFT. The effect of the size of the quantum region was tested systematically by including (i) only the solute in the quantum region or (ii) the solute and its first solvation shell (defined as the 12 water molecules closest to the acetone molecule).

Using the ROKS method a blue shift of 0.23 eV (experimental value: 0.21 eV [94]) is calculated for the case in which only acetone itself is included in the QM region. Addition of the first solvation shell has only a tiny effect (a shift of 0.03–0.04 eV) indicating that the solvent shift is basically converged with respect to

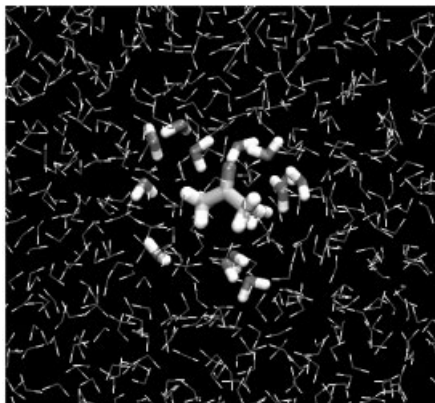


Fig. 1.5 Acetone in water. The QM region is shown as thick cylinders, the MM region as thin sticks.

the size of the QM part. The excellent reproduction of the experimentally measured value suggests QM/MM Car-Parinello simulation as a promising technique for predicting solvent shifts for spatially localized electronic transitions.

The absolute values of the absorption energy for different configurations in solution can vary substantially ($\Delta=0.3\text{--}0.4$ eV). Comparison of ROKS and TDDFT values shows that both techniques predict very similar relative variations although their absolute values are shifted by a constant amount.

1.6

Acknowledgments

Both authors would like to thank Michele Parrinello for his vital role in all the work described here. Contributions of Alessandro Laio, Joost VandeVondele and Ute Röhrig are gratefully acknowledged. Part of this work has been supported by the Swiss National Science Foundation (Grant No. 21-57250.99) and the ETH Research Foundation.

1.7

References

- 1 P. HOHENBERG, W. KOHN, *Phys. Rev. B* **1964**, *136*, 864–871.
- 2 W. KOHN, L. J. SHAM, *Phys. Rev. A* **1965**, *140*, 1133–1138.
- 3 W. KOHN, P. VASHISTA, in: *Theory of the Inhomogeneous Electron Gas*, N. H. MARCH, S. LUNDQVIST (eds), Plenum, New York, **1983**.
- 4 R. G. PARR, W. YANG, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, **1989**.
- 5 R. M. DREIZLER, E. K. U. GROSS, *Density-Functional Theory*, Springer, Berlin, **1990**.
- 6 W. KOHN, *Rev. Mod. Phys.* **1999**, *71*, 1253–1266.
- 7 J. K. LABANOWSKI J. ANDZELM, *Density Functional Methods in Chemistry*, Springer, Heidelberg, **1991**.
- 8 L. ERIKSSON, *Theoretical Biochemistry – Processes and Properties of Biological Systems*, Elsevier Science, **2001**.
- 9 R. CAR, M. PARRINELLO, *Phys. Rev. Lett.* **1985**, *55*, 2471.
- 10 A. LAIO, J. VANDEVONDELE, U. RÖTHLISBERGER, *J. Chem. Phys.* **2002**, *116*, 6941–6948.

- 11 M. C. COLOMBO, L. GUIDONI, A. LAIO, A. MAGISTRATO, P. MAURER, S. PIANA, U. RÖHRIG, K. SPIEGEL, M. SULPIZI, J. VANDEVONDELE, M. ZUMSTEIN, U. RÖTHLISBERGER, *CHIMIA* **2002**, *56*, 11–17.
- 12 A. LAIO, J. VANDEVONDELE, U. RÖTHLISBERGER, *J. Phys. Chem. B* **2002**, *106*, 7300–7307.
- 13 A. PUTRINO, D. SEBASTIANI, M. PARRINELLO, *J. Chem. Phys.* **2000**, *113*, 7102–7109.
- 14 D. SEBASTIANI, M. PARRINELLO, *J. Phys. Chem. A* **2001**, *105*, 1951–1958.
- 15 S. PIANA, D. SEBASTIANI, P. CARLONI, M. PARRINELLO, *J. Am. Chem. Soc.* **2001**, *123*, 8730.
- 16 U. RÖHRIG, A. LAIO, J. VANDEVONDELE, J. HUTTER, U. RÖTHLISBERGER, *Chem. Phys. Chem.*, submitted.
- 17 J. HUTTER, M. ODELIUS in: *Annual Report 2000/2001 of the Competence Center for Computational Chemistry*, DAURA, X. (ed.), ETH, Zürich, **2001**, pp. 28–55.
- 18 J. HUTTER, manuscript in preparation
- 19 E. RUNGE, E. K. U. GROSS, *Phys. Rev. Lett.* **1984**, *52*, 997.
- 20 W. ANDREONI, P. BALLONE, *Phys. Scripta* **1987**, *T19A*, 289.
- 21 R. CAR, M. PARRINELLO, *Phys. Rev. Lett.* **1988**, *60*, 204.
- 22 U. RÖTHLISBERGER, *Computational Chemistry: Reviews of Current Trends*, J. Leszczynski **2001**, *6*, 33–68.
- 23 P. CARLONI, U. RÖTHLISBERGER, *Theoretical Biochemistry – Processes and Properties of Biological Systems*, ERIKSSON, L. (ed.), Elsevier Science, **2001**, p. 215.
- 24 P. CARLONI, U. RÖTHLISBERGER, M. PARRINELLO, *Acc. Chem. Res.* **2002**, *35*, 455–464.
- 25 P. CARLONI, F. ALBER, *Perspect. Drug Discov.* **1998**, *9/11*, 169–179.
- 26 M. SULPIZI, A. LAIO, J. VANDEVONDELE, A. CATTANEO, U. RÖTHLISBERGER, P. CARLONI, *Proteins* **2002**, submitted for publication.
- 27 I. FRANK, J. HUTTER, D. MARX, M. PARRINELLO, *J. Chem. Phys.* **1998**, *108*, 4060–4069.
- 28 U. RÖHRIG, A. LAIO, J. VANDEVONDELE, U. RÖTHLISBERGER, **2002**, manuscript in preparation.
- 29 M. SULPIZI, P. CARLONI, U. RÖTHLISBERGER, **2002**, to be published.
- 30 G. PASTORE, E. SMARGIASSI, F. BUDA, *Phys. Rev. A* **2002**, *44*, 6334.
- 31 P. E. BLÖCHL, M. PARRINELLO, *Phys. Rev. Lett.* **2002**, *45*, 9412.
- 32 G. KRESSE, J. HAFNER, *J. Non Cryst. Solids* **2002**, *956*, 156.
- 33 A. ALAVI, J. KOHANOFF, M. PARRINELLO, D. FRENKEL, *Phys. Rev. Lett.* **1994**, *73*, 2599.
- 34 J. VANDEVONDELE, A. DEVITA, *Phys. Rev. B* **1999**, *60*, 13241.
- 35 L. VERLET, *Phys. Rev.* **1967**, *159*, 98.
- 36 W. C. SWOPE, H. C. ANDERSEN, P. H. BERENS, K. R. WILSON, *J. Chem. Phys.* **1982**, *76*, 637.
- 37 H. C. ANDERSEN, *J. Chem. Phys.* **1980**, *72*, 2384.
- 38 S. NOSÉ, *Mol. Phys.* **1984**, *52*, 255.
- 39 W. G. HOOVER, *Phys. Rev. Lett.* **1985**, *31*, 1695.
- 40 S. MELCHIONNA, G. CICCOTTI, B. L. HOLIAN, *Phys. Rev.* **1993**, *78*, 533.
- 41 M. PARRINELLO, A. RAHMAN, *Phys. Rev. Lett.* **1980**, *45*, 1196.
- 42 P. FOCHER, G. L. CHIAROTTI, M. BERNASCONI, E. TOSATTI, M. PARRINELLO, *Europhys. Lett.* **1994**, *26*, 345.

- 43 M. BERNASCONI, G. L. CHIAROTTI, P. FOCHEP, S. SCANDALO, E. TOSATTI, M. PARRINELLO, *J. Phys. Chem. Solids* **1995**, *56*, 501.
- 44 D. MARX, M. PARRINELLO, *Z. Phys. B* **1994**, *94*, 143.
- 45 D. MARX, M. PARRINELLO, *J. Chem. Phys.* **1996**, *104*, 4077.
- 46 M. E. TUCKERMAN, D. MARX, M. L. KLEIN, M. PARRINELLO, *J. Chem. Phys.* **1996**, *104*, 5579.
- 47 G. J. MARTYNA, A. HUGHES, M. E. TUCKERMAN, *J. Chem. Phys.* **1999**, *110*, 2810.
- 48 J. IHM, A. ZUNGER, M. L. COHEN, *Solid State Phys.* **1979**, *12*, 4409.
- 49 M. T. YIN, M. L. COHEN, *Phys. Rev. B* **1982**, *25*, 3259.
- 50 G. P. SRIVASTAVA, D. WEAIRE, *Adv. Phys.* **1987**, *36*, 463.
- 51 W. E. PICKETT, *Comput. Phys. Rep.* **1989**, *9*, 115.
- 52 D. J. SINGH, *Plane waves, Pseudopotentials and the LAPW Method*, Kluwer, Dordrecht, **1994**.
- 53 M. FUCHS, M. SCHEFFLER, *Comp. Phys. Commun.* **1999**, *119*, 67.
- 54 G. P. KERKER, *J. Phys. C* **1980**, *13*, L189.
- 55 G. B. BACHELET, D. R. HAMANN, M. SCHLÜTER, *Phys. Rev. B* **1982**, *26*, 4199.
- 56 N. TROUILLER, J. L. MARTINS, *Phys. Rev. B* **1991**, *43*, 1993 and 8861.
- 57 X. GONZE, R. STUMPF, M. SCHEFFLER, *Phys. Rev. B* **1991**, *43*, 8503.
- 58 J. S. LIN, A. QTEISH, M. C. PAYNE, V. HEINE, *Phys. Rev. B* **1993**, *47*, 4174.
- 59 D. VANDERBILT, *Phys. Rev. B* **1985**, *32*, 8412.
- 60 S. GOEDECKER, M. TETER, J. HUTTER, *Phys. Rev. B* **1996**, *54*, 1703.
- 61 C. HARTWIGSEN, S. GOEDECKER, J. HUTTER, *Phys. Rev. B* **1998**, *58*, 3661.
- 62 P. E. BLÖCHL, *J. Chem. Phys.* **1995**, *103*, 7422.
- 63 D. MARX, E. FOIS, M. PARRINELLO, *Int. J. Quantum Chem.* **1996**, *57*, 655.
- 64 G. J. MARTYNA, M. E. TUCKERMAN, *J. Chem. Phys.* **1999**, *110*, 2810.
- 65 J. BROUGHTON, F. KAHN, *Phys. Rev. B* **1990**, *40*, 12098.
- 66 S. HAMMES-SCHIFFER, H. C. ANDERSEN, *J. Chem. Phys.* **1993**, *99*, 2810.
- 67 B. HARTKE, E. A. CARTER, *Chem. Phys. Lett.* **1992**, *189*, 358.
- 68 S. A. MALUENDES, M. DUPUIS, *Int. J. Quantum Chem.* **1992**, *42*, 1327.
- 69 A. HEIDENREICH, J. SAUER, *Z. Phys. D* **1995**, *35*, 279.
- 70 J. JELLINEK, V. BONACIC-KOUTECKY, P. FANTUCCI, M. WIECHERT, *J. Chem. Phys.* **1994**, *101*, 10092.
- 71 B. HARTKE, E. A. CARTER, *J. Chem. Phys.* **1992**, *97*, 6569.
- 72 A. J. R. DA SILVA, M. R. RADEKE, E. A. CARTER, *Surf. Sci. Lett.* **1997**, *381*, L628.
- 73 Z. LIU, L. E. CARTER, E. A. CARTER, *J. Phys. Chem.* **1995**, *99*, 4355.
- 74 P. E. BLÖCHL, *Phys. Rev. B* **1994**, *50*, 17953.
- 75 F. GYGI, *Europhys. Lett.* **1992**, *19*, 617.
- 76 G. LIPPERT, J. HUTTER, M. PARRINELLO, *Phys. Rev.* **1997**, *92*, 477.
- 77 G. LIPPERT, J. HUTTER, M. PARRINELLO, *Theor. Chem. Acc.* **1999**, *103*, 124.
- 78 M. KRACK, M. PARRINELLO, *Chem. Phys. Phys. Chem.* **2000**, *2*, 2105.

- 79 N. D. MERMIN, *Phys. Rev.* **1965**, 137, A1441.
- 80a ABINIT is a common project of the Université Catholique de Louvain, Corning Inc., and other contributors (www.abinit.org)
- 80b M. C. PAYNE, M. P. TETER, D. C. ALLAN, T. A. ARIAS, J. D. JOANNOPULOS, *Rev. Mod. Phys.* **1992**, 64, 1945.
- 81 J. HUTTER, A. ALAVI, T. DEUTSCH, M. BERNASCONI, S. GOEDECKER, D. MARX, M. TUCKERMAN, M. PARRINELLO, *MPI für Festkörperforschung and IBM Zurich Research Laboratory*. <http://www.cpmd.org>
- 82 M. BOCKSTEDTE, A. KLEY, J. NEUGEBAUER, M. SCHEFFLER, *Comput. Phys. Commun.* **1997**, 107, 187.
- 83 P. ORDEJON, E. ARTACHO, J. M. SOLER, *Phys. Rev. B* **1996**, R10441.
- 84a G. KRESSE, J. FURTHMÜLLER, *Phys. Rev. B* **1996**, 54, 11169.
- 84b For a recent review see: P. SHERWOOD in *Modern Methods and Algorithms of Quantum Chemistry*, NIC Series Vol. 1, J. GROENDORST (ed), Jülich **2000**.
- 85 W. R. P. SCOTT, P. H. HÜNENBERGER, I. G. TIRONI, A. E. MARK, J. FENNEN, A. E. TORDA, T. HUBER, P. KRÜGER, W. F. VAN GUNSTEREN, *J. Phys. Chem. A* **1999**, 103, 3596.
- 86 W. D. CORNELL, P. CIEPLAK, C. I. BAYLY, K. M. GOULD, K. M. MERZ, D. M. FERGUSON, G. L. SEIBEL, U. C. SINGH, P. K. WEINER, P. A. KOLLMANN, *J. Am. Chem. Soc.* **1995**, 117, 5179.
- 87 P. HÜNENBERGER, *J. Chem. Phys.* **2000**, 23, 10464.
- 88 A. D. BECKE, *Phys. Rev. A* **1988**, 38, 3098.
- 89a C. LEE, W. YANG, R. G. PARR, *Phys. Rev. B* **1988**, 37, 785.
- 89b H. J. DERENDSEN, J. P. M. POSTMA, W. F. VON GUNSTEREN, J. HERTMANS in *Intermolecular Forces*, B. PULLMAN (ed) Dordrecht, **1981**.
- 89c F. L. HIRSHFELD, *Theor. Chim. Acta* **1977**, 44, 129.
- 90 J. VANDEVONDELE, U. RÖTHLISBERGER, *J. Phys. Chem. B* **2002**, 106, 203–208.
- 91 X. GONZE, *Phys. Rev. A* **1995**, 52, 1096.
- 92 T. HELGAKER, M. JASZUNSKI, K. RUUD, *Chem. Rev.* **1999**, 99, 293.
- 93 P. L. SILVESTRELLI, M. PARRINELLO, *J. Chem. Phys.* **1999**, 111, 3572.
- 94 N. S. BAYLISS, G. WILLS-JOHNSON, *Spectrochim. Acta* **1968**, 24A, 55.

2

Density-functional Theory Applications in Computational Medicinal Chemistry

ANDREA CAVALLI, GERD FOLKERS, MAURIZIO RECANATINI,
and LEONARDO SCAPOZZA

2.1

Introduction

From genomics, proteomics, and functional and structural genomics, the scientific community and the community in general are expecting answers to crucial issues related to health and the quality of life. In this context, medicinal chemistry and the rational mechanism-based approach of developing new therapeutics play a major role [1]. The fundamental assumption of the 'rational' approach is that the event that produces the beneficial effects of drugs is the molecular recognition and binding of ligands to the active site of specific targets, such as enzymes, receptors, and nucleic acids. The effect of binding can be promotion or inhibition of signal transduction, of enzymatic activity, or of molecular transport. The design of small molecules able to affect the biological functions of latter is, therefore, one of the major aims in the future of medicinal chemistry.

The history of drug discovery is characterized by systematic searching for compounds endowed with biological activity by use of animal models for human diseases. Nowadays combination of experimental methods for structure determination, e.g. crystallography and NMR, with theoretical procedures known as computer-aided molecular design (CAMD) is essential for the development of new drugs aimed at new targets, and thus for medicinal chemistry [2]. A variety of computational chemistry methods is used in CAMD, which comprises mainly two categories of approach – ligand-based (e.g. receptor mapping, QSAR, pseudoreceptor) and structure-based (e.g. *de novo* design, virtual screening) methods [3–6]. Within this scenario, theoretical methods based on density-

functional theory (DFT) are going to play an increasingly prominent role in many applications of computational chemistry to drug discovery. Several questions dealing with the electronic structure of the matter are increasingly tackled by means of DFT-based calculations, such that it seems appropriate to review briefly here the contribution already provided by DFT-based methods to the central issue of theoretical medicinal chemistry, that can be broadly designated as the study of drug-target interaction.

In this chapter, a short introduction to DFT and to its implementation in the so-called *ab initio* molecular dynamics (AIMD) method will be given first. Then, focusing mainly on our own work, applications of DFT to such fields as the definition of structure-activity relationships (SAR) of bioactive compounds, the interpretation of the mechanism of enzyme-catalyzed reactions, and the study of the physicochemical properties of transition metal complexes will be reviewed. Where possible, a case study will be examined, and other applications will be described in less detail.

2.2 Density-functional Theory and Related Methods

In the following text we present a very short synopsis both of the DFT approach and the *ab initio* molecular dynamics (AIMD) method that can by no means be considered as an introduction to the use of the computational tools based on them. The interested reader will find exhaustive treatment of these arguments elsewhere in this book (Chapter 1).

2.2.1 Density-functional Theory

Density-functional theory provides a framework to deal with the ground-state energy of the electrons in many-atom systems. In general, the problem of finding the ground-state of a many-electron system consists in finding the lowest energy eigenvalue E and the corresponding eigenstate ψ of the time-independent Schrödinger equation:

$$\hat{H}\psi = E\psi \quad (1)$$

To calculate E , one must solve Eq. (2):

$$E = \text{Min}_{\psi} \langle \psi | \hat{H} | \psi \rangle \quad (2)$$

which is minimized by means of a normalized electronic wavefunction:

$$\psi(\{r\}) = \psi(r_1, r_2, K, r_n) \quad (3)$$

Here, the difficulty is that ψ is a many-body electronic wavefunction, which depends on the coordinates of all the n electrons.

In DFT, a different approach is followed. Rather than focusing on ψ one focuses on the single particle density $\rho(r)$, which is a quantity related to ψ by the equation:

$$\rho(r) = n \int dr_1, dr_2, K dr_{n-1} |\psi(r, r_1, r_2, K r_{n-1})|^2 \quad (4)$$

The density ρ is a much simpler quantity than ψ , because it depends on one spatial coordinate only.

In this context, the crucial point is how to associate the ground-state energy of the electrons with their density. This issue was addressed by Hohenberg and Kohn [7], who demonstrated that the ground-state energy of a system of interacting electrons subject to an external local potential $V(r)$ is a unique functional of the electron density. The actual ground-state energy and the corresponding density can be found by minimizing the energy functional with respect to the density. Formally:

$$E = \text{Min}_{\rho} E_V[\rho] \quad (5)$$

where the minimization is performed with respect to density distributions that preserve n , the total number of electrons. Conceptually Eq. (5) is a great simplification compared with Eq. (2). In practice, following Hohenberg and Kohn [7], one can write:

$$E_V[\rho] = \int dr V(r)\rho(r) + \frac{1}{2} \int dr dr' \frac{\rho(r)\rho(r')}{|r - r'|} + F[\rho] \quad (6)$$

The first term on the right hand side of Eq. (6) represents the energy of interaction of the electrons with the external potential

V. The second term is the classical Coulomb energy of a density distribution ρ . The quantity $F[\rho]$ is a universal functional of the density, which means that it is uniquely specified by the density ρ of the interacting electrons and does not depend on the particular external potential V acting on the electrons. The functional F contains whatever is necessary to make the energy in Eq. (6) equal to the expected value in Eq. (2).

Kohn and Sham provided a further contribution to make the DFT approach useful for practical calculations, by introducing the concept of *fictitious non-interacting electrons* with the same density as the true interacting electrons [8]. Non-interacting electrons are described by orthonormal single-particle wavefunctions $\psi_i(r)$ and their density is given by:

$$\rho(r) = 2 \sum_i |\psi_i(r)|^2 \quad (7)$$

Here, the factor 2 accounts for spin degeneracy, and the sum is over the occupied single-particle states. In terms of ψ_i , the energy functional given by Eq. (6) can be written as:

$$E_V[\rho] = \int dr V(r)\rho(r) + \frac{1}{2} \int dr dr' \frac{\rho(r)\rho(r')}{|r-r'|} + 2 \sum_i \left\langle \psi_i \left| -\frac{\nabla^2}{2} \right| \psi_i \right\rangle + E_{xc}[\rho] \quad (8)$$

The third term on the right hand side of this expression is the single-particle kinetic energy of the noninteracting electrons whereas the functional $E_{xc}[\rho]$ contains the additional contribution to the energy that is needed to make Eq. (8) equal to Eq. (6).

The term $E_{xc}[\rho]$ is called the exchange-correlation energy functional and represents the main problem in the DFT approach. The exact form of the functional is unknown, and one must resort to approximations. The *local density approximation* (LDA), the first to be introduced, assumed that the exchange and correlation energy of an electron at a point r depends on the density at that point, instead of the density at all points in space. The LDA was not well accepted by the chemistry community, mainly because of the difficulty in correctly describing the chemical bond. Other approaches to $E_{xc}[\rho]$ were then proposed and enable satisfactory prediction of a variety of observables [9].

Going back to Eq. (8), the problem now is that one must minimize E_V with respect to the ψ_i , instead of directly minimizing it with respect to ρ . This can, however, be achieved by solving Eq. (9), called the Kohn-Sham equation:

$$\left(-\frac{\nabla^2}{2} + V(r) + \int dr' \frac{\rho(r')}{|r-r'|} + V_{xc}(r) \right) \psi_i(r) = \varepsilon_i \psi_i(r) \quad (9)$$

Here, $V_{xc} \equiv \frac{\delta E_{xc}[\rho]}{\delta \rho(r)}$ is a local potential called the exchange and correlation potential. The ε_i are eigenvalues of the matrix of Lagrange multipliers and are called the Kohn-Sham eigenvalues. The Kohn-Sham equations have the form of self-consistent-field equations of the Hartree type, but, in contrast with the Hartree equations, which are approximate, are formally exact. The self-consistent potential acting on the electrons in Eq. (9) is given by a sum of the external potential, the Hartree potential, and the exchange and correlation potential. Under the action of this self-consistent potential, the non-interacting electrons acquire the same density of the true interacting electrons. After solving Eq. (9), the electron density is calculated from Eq. (7), and the corresponding ground-state energy (of the true interacting electrons) is calculated from Eq. (8).

2.2.2

Ab Initio Molecular Dynamics

The modeling of atoms in motion has become an essential requirement in every study aimed at elucidating the characteristics of a ligand-target interaction. For this reason, molecular dynamics (MD) simulations are almost routine tasks in research groups involved in rational drug design. Usually, in MD simulations, empirical force fields are used to describe the interatomic interactions. Although the accuracy of the force fields is often remarkable, it cannot, however, match the accuracy and predictive power of quantum mechanical calculations based on DFT, particularly when changes in the electronic structure play a crucial role, e.g. in chemical reactions. To address this issue, in the mid-eighties Car and Parrinello formulated a new molecular dynamics scheme [10] in which the potential energy surface is generated from the instantaneous ground state of the electrons with-

in DFT. This approach, called *ab initio* (or first-principles) molecular dynamics (AIMD), is now used with success to model chemical processes in chemistry and biology.

The basic idea underlying AIMD is to compute the forces acting on the nuclei by use of quantum mechanical DFT-based calculations. In the Car-Parrinello method [10], the electronic degrees of freedom (as described by the Kohn-Sham orbitals $\psi_i(r)$) are treated as dynamic classical variables. In this way, electronic-structure calculations are performed “on-the-fly” as the molecular dynamics trajectory is generated. Car and Parrinello specified system dynamics by postulating a classical Lagrangian:

$$L = \frac{1}{2} \sum_I M_I \dot{R}_I^2 + 2\mu \sum_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \tilde{E}[\{R\}, \{\psi\}] \quad (10)$$

In this expression μ is a “mass” parameter associated to the electronic fields, i.e. it is a parameter that fixes the time scale of the response of the classical electronic fields to a perturbation. The factor 2 in front of the classical kinetic energy term is for spin degeneracy. The functional $\tilde{E}[\{R\}, \{\psi\}]$ plays the role of potential energy in the extended parameter space of nuclear and electronic degrees of freedom. It is given by:

$$\tilde{E}[\{R\}, \{\psi\}] = E_V[\rho] + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|R_I - R_J|}, \quad (11)$$

where $E_V[\rho]$ is the Kohn-Sham functional shown in Eq. (8). A constraint of orthonormality between the Kohn-Sham orbitals at each instant of time is also added.

The following equations of motion can then be derived:

$$M_I \ddot{R}_I = - \frac{\partial \tilde{E}}{\partial R_I} \quad (12)$$

$$\mu \ddot{\psi}_i = -\hat{H}_{KS} \psi_i + \sum_j \lambda_{ij} \psi_j \quad (13)$$

Here, $\hat{H}_{KS} = -\frac{\nabla^2}{2} + V + V_H + V_{xc}$ is the Kohn-Sham Hamiltonian of Eq. (9) in which V_H denotes the Hartree potential. When the ψ_i are at the minimum of \tilde{E} for a given nuclear configura-

tion $\{R\}$, they are solutions of the Kohn-Sham equation, and $\ddot{\psi}_i = 0$ in Eq. (13). Correspondingly, $\tilde{E} = [\{R\}, \{\psi\}] = \Phi(\{R\})$ and Eq. (12) coincides with:

$$M_I \ddot{R}_I = - \frac{\partial \Phi}{\partial R_I} \quad (14)$$

which represents the classical Newton's equations of motion of the nuclei. By appropriate choice of μ the electronic dynamics computed from Eq. (13) is considerably faster than the nuclear dynamics from Eq. (12). If at time $t=0$ the electrons are in the ground state corresponding to a nuclear configuration $\{R(t=0)\}$, at $t>0$ the nuclei evolve according to Eq. (12), and the electrons correspondingly evolve according to Eq. (13). The evolution of the electrons consists of rapid oscillations around their slowly changing instantaneous ground state. The oscillations have frequencies, $\sqrt{\frac{\Delta \epsilon}{\mu}}$, where $\Delta \epsilon$ is a Kohn-Sham excitation energy, i.e. the energy difference between HOMO and LUMO. If the electronic frequency is much higher than the nuclear frequency, decoupling of the nuclear and electronic motions is observed and the electrons follow the nuclei adiabatically.

Today, AIMD is extensively applied in different fields of computational chemistry and, certainly, because cell functions occur at approximately 310 K, its use in life sciences seems particularly appropriate, because of the importance of temperature effects in biological systems. In this respect AIMD, which accounts explicitly for the dynamic behavior at finite temperature, can be considered one of the methods of choice when performing DFT-based simulations in theoretical medicinal chemistry. Further details about AIMD and its applications in biological chemistry are available elsewhere [11].

Finally, it must be remembered that DFT and AIMD can be incorporated into the so-called mixed quantum mechanical/molecular mechanical (QM/MM) hybrid schemes [12, 13]. In such methods, only the immediate reactive region of the system under investigation is treated by the quantum mechanical approach – the effects of the surroundings are taken into account by means of a classical mechanical force field description. These DFT/MM calculations enable realistic description of atomic processes (e.g. chemical reactions) that occur in complex heterogeneous envir-

oments, e.g. homogeneous catalytic processes in solution, or enzymatic reaction cycles in an explicit protein environment.

2.3

SAR Studies of Ligand-Target Interactions

In medicinal chemistry, a key question to be answered when studying a class of bioactive compounds is what determines the variation of biological potency within a set of analogs. The answer, if obtained, defines the SAR of the series. It might be thought that nowadays advanced and high-throughput methods for determination of the structure of ligand-macromolecule complexes should be able to provide the information needed to build the SAR, at least for systems such as enzyme-inhibitor or DNA-ligand complexes. Indeed, this is often so, but it can happen that details at the atomic and electronic levels are needed to interpret experimental data on the ligand-target interaction. In this section we will present a case study showing how DFT-based quantum chemistry calculations can help address issues that remained unrevealed after structural elucidation.

2.3.1

The Case Study: Herpes Simplex Virus Type 1 Thymidine Kinase Substrates and Inhibitors

Thymidine kinase (TK, EC 2.7.1.21) is the key enzyme in the pyrimidine salvage pathway catalyzing the phosphorylation of thymidine (dT) to thymidine monophosphate (dTMP) in the presence of Mg^{2+} and ATP [14]. In contrast with cellular TK, Herpes simplex virus type 1 thymidine kinase (HSV1 TK) accepts a broad range of substrates (Fig. 2.1) [15–17]. This peculiarity is the basis of several therapeutic and diagnostic applications, e.g. antiviral therapy, in which HSV1 TK is involved [15], suicide gene therapy in stem cell transplantation (SGT in SCT) [18], and *in-situ* molecular imaging [19]. For these applications, prodrugs (compounds that have to be activated *in vivo* to achieve the desired pharmacological effect) are used. These prodrugs, e.g. aciclovir (ACV) and ganciclovir (GCV), shown in Fig. 2.1, are selec-

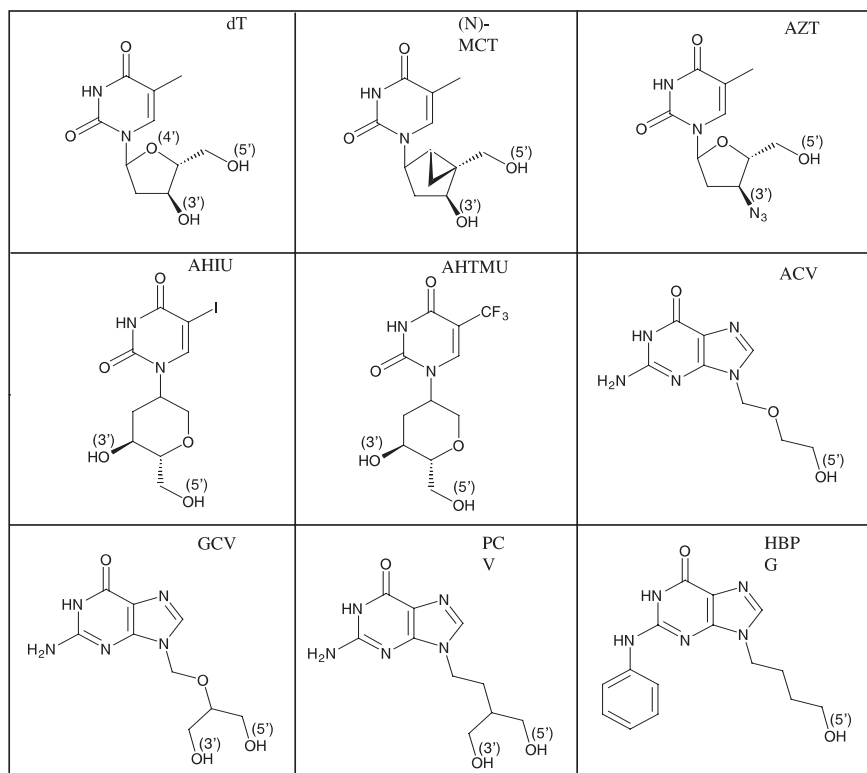


Fig. 2.1 Chemical formulas of selected (fraudulent) substrates and inhibitors of HSV1 TK. Thymidine (dT) is the natural substrate. (N)-MCT: 2-*exo*-methanocarbathymidine, (North)-methanocarbathymidine; ACV: aciclovir; PCV: penciclovir; GCV: ganciclovir; AHIU: 5-iodouracil anhydrohexitoluridine;

AHTMU: 5-trifluoromethylanhydrohexitoluridine; AZT: 3-azidothymidine; HBPG: 9-(4-hydroxybutyl)-*N*²-phenylguanine. (N)-MCT, ACV, PCV, GCV, AHIU, AHTMU, and AZT are prodrugs; HBPG is an inhibitor. The 5-OH and 3-OH groups belonging to dT and their mimics are labeled.

tively activated through phosphorylation by HSV1 TK to act in their triphosphorylated form as DNA polymerase inhibitors and DNA chain terminators [15, 16] stopping viral and cell proliferation, when used as antiviral compound and prodrug in SGT and SCT [20], respectively. Despite the clinical success of the antiviral therapy, resistance has emerged as a relevant problem [21]. In SGT, moreover, some limitations related to the suicide gene HSV1 TK were observed [22]. To address the clinical limitations, a search for a new prodrug and the design of HSV1 TK mutants

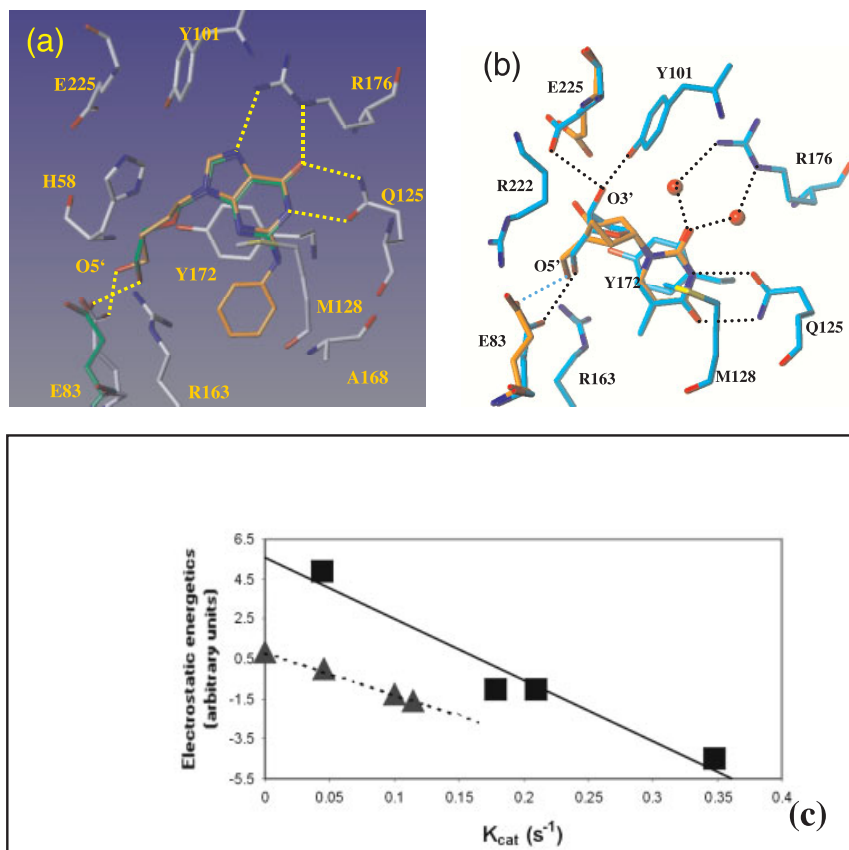


Fig. 2.2 Active site superposition of representative ligands of the guanine (a) (ACV: C atoms in green, O in red, N in blue; HBPG: C atoms in orange, O in red, N in blue) and thymine (b) (thymidine: C atoms in cyan, O in red, N in blue); (N)-MCT: C atoms in orange, O in red, N in blue) series. Hydrogen bonds are shown as dashed lines. (c) Correlation between k_{cat} and the sugar-moiety-dipole-Glu225 electrostatic energy (electrostatic energetics). Interaction energy was

calculated as $E_{\text{charge-dipole}} = (\zeta\mu R) / (4\pi\epsilon_0 R^3)$, where R is the distance vector between the centers of charges of the sugar moiety and the residue of charge ζ and μ is the dipole moment for the sugar moiety. Thymine and guanine derivatives are displayed as squares and triangles, respectively. Linear fits are also plotted (R^2 values are 0.954 and 0.994 for thymine and guanine, respectively). Adapted from Ref. [33].

with improved specificity towards the prodrug is, therefore, a useful endeavor [23]. A prerequisite for achieving this aim is knowledge at molecular level of substrate binding and catalysis. For this reason, several structures of HSV1 TK in complex with different prodrugs have been solved [24–30]. The enzyme three dimensional (3D) structure is characterized by a typical $\alpha\beta$ fold with a central five-stranded parallel β -sheet surrounded by twelve helices and is highly conserved, despite the ligands. The X-ray structures clearly revealed that prodrugs (substrates) are bound to the nucleoside binding site via both a sandwich-like complex formed by Met128 and Tyr172 and a sophisticated hydrogen-bonding network, including a Watson-Crick like interaction between Gln125 and the base, and that the 5-OH group points towards the catalytic base Glu83 (Fig. 2.2a,b). Although this information is essential for understanding steric accessibility and the role of hydrogen bonds [21], several questions remained open concerning substrate diversity and catalysis [26]. To address these issues, DFT-based calculations were recently performed, and the results verified experimentally [31–33].

2.3.1.1 Rationalizing Substrate Diversity – SAR of HSV1 TK Ligands

The Role of Met128 and Tyr172 in Nucleobase Fixation

From a multiple alignment study of type I thymidine kinases (long TK) of different species, e.g. herpes simplex virus type 1,2 (HSV 1,2), marmoset herpes virus (MHV), equine herpes virus type 4 (EHV), varicella zoster virus (VZV), and Epstein-Barr virus (EBV), it was found that Gln125 was conserved over all species whereas the triad His58/Met128/Tyr172 was conserved only by strains with broad substrate diversity (HSV1, HSV2, and MHV). In contrast, strains with narrow substrate specificity (EHV, VZV, and EBV) had X58/Phe128/Tyr172 as consensus triad, where X is a hydrophobic amino acid but never His [32]. To gain a better understanding of the nature of the interactions of thymine with Tyr172 and Met128, *ab initio* DFT calculations were performed using the generalized gradient approximation to the exchange-correlation energy functional [9]. Several model complexes, com-

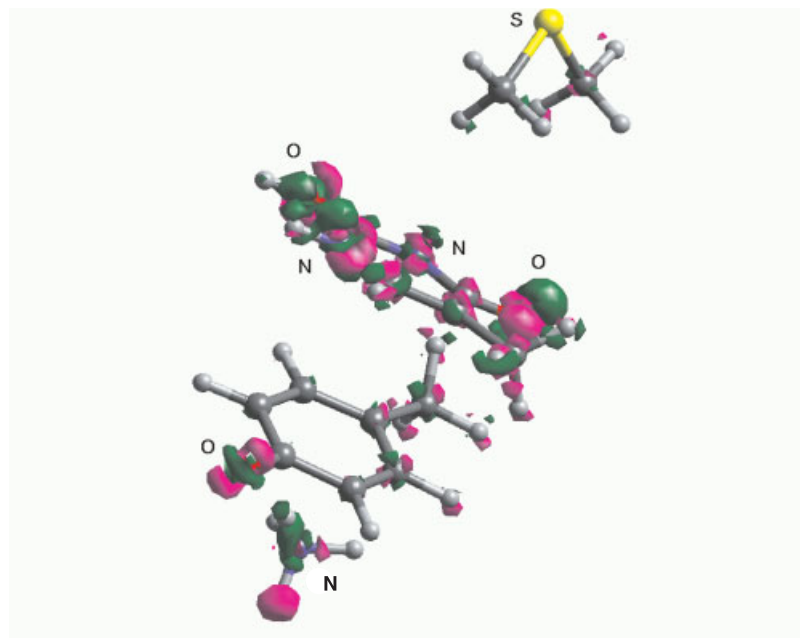


Fig. 2.3 Isodensity contours of the difference electron density $\Delta\rho$ ($\Delta\rho = \rho_{\text{complex}} - \rho_{\text{fragments}} - \rho_{\text{substrates}}$) of complex with total charge +1. Green: $+0.054 \text{ e } \text{\AA}^{-3}$; magenta: $-0.054 \text{ e } \text{\AA}^{-3}$. Atoms are shown color coded (C in black, O in red, N in blue, H in gray). Adapted from Ref. [31].

posed of Met128, Tyr172, Arg163, and thymine, were constructed by protonating the thymine ring and Tyr172 differently, and by taking tautomerism into account [31]. The complexes were built starting from the X-ray structures [25] and the calculations were performed with a scalar version of the Car-Parrinello code [10]. AIMD simulations were performed to assess the stability of the complex. The calculations revealed strong polarization on thymine, but no polarization on the sulfur atom of Met128 (Fig. 2.3). The Met128 and Tyr172 thymine interactions were further investigated by calculating the Kohn-Sham levels at the region of the HOMO-LUMO gap of the complexes. Neither overlapping between the molecular orbitals of Met128 and those of the substrate, nor π - π interactions between Tyr172 and thymine could be observed. This indicated that Tyr172-thymine interactions are dominated by electrostatics, and that the role of Met128 is purely steric and hydrophobic [31].

On the basis of this information, the mechanisms of substrate binding of HSV 1 TK were studied by means of site-directed mutagenesis combined with thermodynamic measurements performed using isothermal titration calorimetry [32]. The experimental results showed that Met128 could be replaced by Ile without losing activity, and showed also the link between the exceptionally broad substrate diversity of HSV 1 TK and the presence of structural features such as the residue triad His58/Met128/Tyr172. Thus, the theoretical results were confirmed by thermodynamic studies suggesting that conformational changes are essential for binding, and that accommodation of the base within the plane formed by Met128 and Tyr172 is indeed a prerequisite for the correct formation of hydrogen bonds. These results were confirmed also by the crystal structure of the Q125N mutant complexed with dT, solved at 2.4 Å resolution [30], which clearly showed that the nucleobase binds exactly as in the wild-type enzyme, whereas rearrangement of the hydrogen-bond network occurs [30].

HSV1 TK and Substrate Diversity at the Sugar Moiety Level

In solution, the sugar ring of nucleosides and nucleotides equilibrates between two extreme forms, a 2-*exo*/3-*endo* (North) conformation and a 2-*endo*/3-*exo* (South) conformation [34]. On binding to the active site one particular conformation is fixed, resulting in an unfavorable entropy contribution. Combined studies involving *ab initio* calculation and biochemical and structural characterization performed using thymine derivatives with a conformationally restricted sugar moiety of the bicyclohexane carbocyclic type [34, 35] clearly pointed out that HSV1 TK has no selectivity for the conformation of the sugar moiety [29]. So, despite the elucidation of the key aspects of nucleobase binding described above, fundamental questions about the nature of the interactions between the ribose-like moiety and the enzyme were still unanswered.

Since the first structure of HSV1 TK in complex with dT and ATP was elucidated in 1995 [24], several structures of HSV1 TK have shed light on the binding mode of a variety of ligands. A striking finding was that the inhibitor 9-(4-hydroxybutyl)-*N*²-phenylguanine (HBPG, Fig. 2.1) has the same binding mode as structurally related substrates such as aciclovir [26] (Fig. 2.2 a, b).

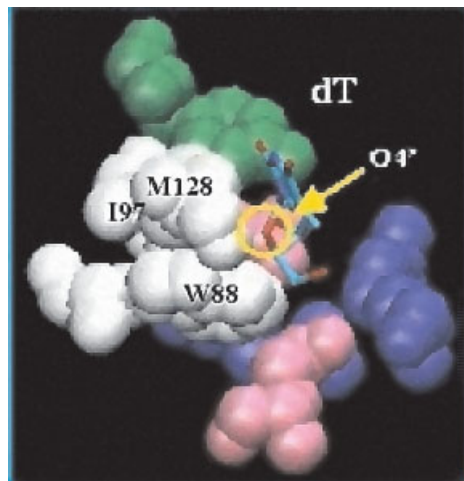


Fig. 2.4 Zoom within the thymidine binding site of the dT-HSV1 TK complex. The orientation of O₄ in its hydrophobic pocket is indicated. Adapted from Ref. [33].

It was, furthermore, not known why the k_{cat} values of prodrugs were much smaller than that of the natural substrate [36], although the protein-sugar mimicking H-bonding interactions are very similar (Fig. 2.2 a, b). Thus, an intriguing question arises as to what is the structural basis for such different properties. Also, from the structural data, it is rather surprising that the environment accommodating the inner sugar ring C₁-O₄-C₄ group of the natural substrate is totally hydrophobic (Fig. 2.4). The question raised by this observation was: what is the role of the O₄-oxygen in the hydrophobic environment?

To address both issues, *ab initio* DFT-based calculations were performed. Starting from crystallographic information about the HSV1 TK in complex with several prodrugs and an inhibitor [23–29], calculations were performed on a group of ligands carrying a representative set of the large spectrum of sugar-mimicking moieties (Fig. 2.1). On the experimental side, the catalytic constants (k_{cat}) of these compounds were measured under the same experimental conditions, using a UV-spectrophotometric assay [36] to obtain a complete and homogenous set of k_{cat} values.

The results were indicative of the crucial role of the electric dipole moment of ligands and its interaction with the negatively charged residue Glu225 and, moreover, the dipole proved to be an useful observable for discriminating between substrate and inhibitor (Fig. 2.5). A striking correlation was found between the energetics associated with this interaction and the k_{cat} values

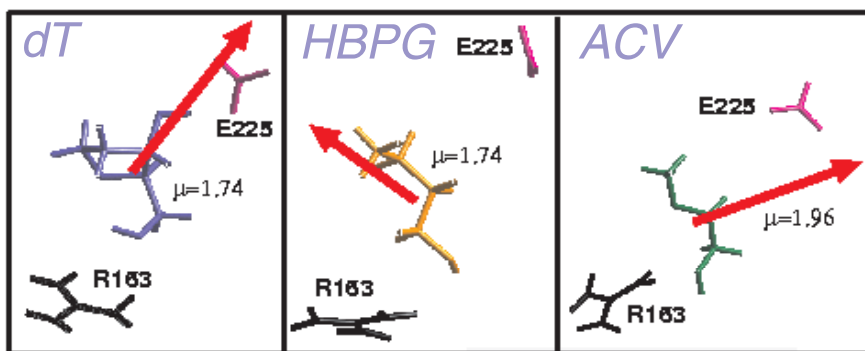


Fig. 2.5 Electric dipoles of HSV1 TK sugar-like chains. dT and ACV are substrates and are phosphorylated by HSV1 TK, whereas HBPG is an inhibitor. The dipole of both

substrates is oriented with the positive moiety towards the negatively charged E225, whereas the dipole of the inhibitor assumes a different orientation. Adapted from Ref. [33].

(Fig. 2.2c), whereas no correlation with binding affinities was observed [33]. The difference between the guanine and thymine analogs corresponds to the additive dipole contribution of the base deriving from the Tyr172-nucleobase electrostatic interactions [31]. Experimental validation using site-directed mutagenesis revealed that replacement of the charge with a non polar residue, e.g. E225L mutation, led to a 20-fold decrease of k_{cat} and confirmed the theoretical results [32].

In view of these results, we addressed the issue of the role of the O_4 oxygen within a purely hydrophobic environment. More precisely the following question was addressed: in the natural substrate, is the polar $C_1-O_4-C_4$ (which faces a hydrophobic pocket; Fig. 2.4) important for correct orientation of the dipole? To answer this question, we calculated the change of the electric dipole associated with replacement of O_4 with the apolar group CH_2 . The resulting dipole was both smaller and different in orientation relative to the sugar moiety, and thus it seemed that the polar function was essential for correct alignment of the dipole to the Glu225 charge.

The protein field might be very important to the chemistry of the active site of this and other enzymes [11]. The effect of the environment was estimated by comparing the electronic structure of the complexes in vacuum with those in the presence of the protein. The Wannier functions [37], the centers (WFC) of which represent

chemical concepts such as lone pairs and chemical bonds, was used to represent the electronic structure. The results of the comparison showed that there is no appreciable displacement of the WFC, so it was possible to conclude that the main contributions to the interaction were included in the model chosen for quantum mechanical calculations. It is, however, important to keep in mind that this might not be so if significant conformational changes of the target occur during ligand binding.

2.3.1.2 What Can be Learned from this Case Study – From SAR to Drug Design

Within the framework of rational design, virtual screening is an emerging tool for developing new therapeutics. Ligands are screened from a library of compounds by docking them into the target using *scoring functions*, which enable crude and rapid estimation of protein-drug interactions. Between the simplest scoring function estimating binding free energies from the atomic coordinates [38, 39] and the CPU-intensive more accurate free energy perturbation techniques [40, 41], the force-field-based empirical scoring function [42, 43], 3D QSAR [44], and continuum methods [45] can be used to estimate binding free energy differences. Although all the scoring functions mentioned enable more or less reliable estimation of the binding free energy none enables discrimination between substrate and inhibitors.

The case presented shows it is crucial to know the electronic details of catalysis and the factors (e.g. protein environment) influencing it to understand binding and catalysis beyond the structural information that is indispensable but not sufficient [26, 33]. Using the dipole of the ligand within the binding site of the target as observable, the DFT-based scoring function was shown to enable quantitative discrimination between substrate and inhibitors, and thus to overcome the limitation of other scoring functions dedicated to prediction of binding free energy. Thus, the discovery of new drugs in the context of prodrug-based therapies would profit from a combination of the DFT-based scoring function addressing the issue of catalysis with a good scoring function for predicting binding free energy.

This study also provides evidence of the applicability of DFT to other therapeutic fields involving enzymes and prodrugs, e.g.

antibody-directed enzyme-prodrug therapy [46], in which both binding and efficiency of catalysis are important for the clinical success of the therapy.

2.4

Theoretical Studies of Enzymatic Catalysis

Enzymes are therapeutic targets of utmost importance in many areas of pharmacological intervention, and many drugs act at the molecular level by inhibiting their action. The search for enzyme inhibitors is quite active in a variety of fields in medicinal chemistry, and takes advantage from all the techniques used to elucidate both the structure and the catalytic mechanisms of the selected targets. It is immediately apparent that a knowledge of the 3D structure of the active site of an enzyme gives the opportunity to design molecules able to fit into the site and consequently to block access of the substrate to the catalytic center. Competitive inhibitors work in this way but other, sometimes more efficient, inhibitors act in a way that takes into account or even exploits the reaction catalyzed by the enzyme; these can be defined as mechanism-based inhibitors. A third class of enzyme inhibitor, not yet thoroughly explored in medicinal chemistry but which might provide highly potent and specific compounds potentially ideal for the use in therapy, is the transition state (TS) analog.

According to the Pauling theory, the powerful catalytic action of the enzymes might be explained by specific binding of the enzymes to the TS [47]. This binding is supposed to be much tighter than that of the substrate in the Michaelis complex. Transition states are believed to bind to the enzymes with a dissociation constant in the order of 10^{-14} – 10^{-23} M [48]. It therefore seems reasonable to assume that compounds that resemble the TS structure and can thus capture even a small fraction of the TS binding energy might be very potent enzyme inhibitors. Remarkably, physical methods furnishing information about the TS in enzymatic reactions at the atomic level are not yet available. *Ab initio* quantum chemical calculations might help overcome this problem, enabling theoretical estimation of the TS of the enzymatic reaction.

In this section, we present some recent applications of DFT to the study of enzymatic reactions of pharmaceutical interest. One of the major aims of these studies was to identify TS features that, because the biological targets investigated are involved in the pathogenesis of diseases, could be exploited when designing inhibitors of these enzymes.

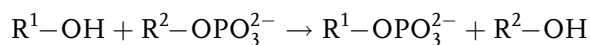
2.4.1

The Phosphoryl Transfer Reaction

2.4.1.1 Cdc42-catalyzed GTP Hydrolysis

Phosphoryl transfer reactions play a crucial role in several fundamental functions of the cell, ranging from the signaling pathways involving proteins kinases to the proinflammatory cytokines pathways and the biosynthesis of nucleic acids [49]. This type of reaction is also involved in some pathology, e.g. that of cancer and Alzheimer's disease [50, 51], and is currently being investigated as a potential target for intervention in anti-tumor chemotherapy and in gene therapy-based anticancer research [52]. In biological systems this reaction is catalyzed by several different enzymes, e.g. hydrolases, kinases, and nucleoside monophosphate kinases.

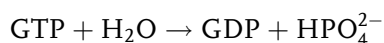
In every type of biological phosphoryl transfer reaction, a P–O bond of the leaving group ($R^2\text{--OH}$) is broken and a new P–O bond is formed between phosphorus and the attacking group $R^1\text{--OH}$:



The phosphoryl transfer reaction between guanosine triphosphate (GTP) and a water molecule, catalyzed by the G-protein Cdc42, was recently investigated [53]. Cdc42 is a small GTP binding protein with GTPase activity and belonging to the Rho GTPases subgroup of the Ras super-family [54]. Cdc42 is involved in a variety of biological processes including organization of actin cytoskeleton, vesicular trafficking, cell cycle progression, and control of transcription [55, 56]. It has, furthermore, been shown that Cdc42 plays an important role in pathological cell growth control – the F28L mutation induces malignant transformation [57]. This protein is a hydrolase that acts as molecular

switch, being “on” in the GTP-bound state and “off” in the GDP-bound state [58]. The intrinsic GTPase activity is therefore fundamental in the production of bound GDP and turning the cell signaling off. This enzymatic activity can be strongly enhanced by physical interaction with GTPase-activating proteins [59] (GAP; here GAP indicates the specific protein Cdc42GAP).

The purpose of the study was theoretical investigation, at the atomic level, of the mechanism of the GTP hydrolysis catalyzed by the Cdc42-GAP enzymatic complex:



to provide evidence about the suitability of DFT-based methods for investigation of the phosphoryl transfer reaction in biological systems. In particular, an AIMD simulation of the reaction was performed by means of the DFT-based Car-Parrinello (CPMD) method [10], which, because it takes into account the effects of finite temperature, can be considered the method of choice when studying reactions in biological systems.

The study was performed on a model system based on the crystal structure of Cdc42-Cdc42GAP complexed with GDP and AlF_3 [60], which can be considered a TS mimic of phosphoryl transfer [61, 62]. A large model system (Fig. 2.6) was required to properly take into account the effect on the reagents of the electrostatic field of the protein. It comprised all the amino acids directly interacting with the triphosphate moiety, the Mg^{2+} cation with its own coordination shell, and AlF_3 replaced by the PO_3^- moiety.

Classical electrostatic modeling based on the Coulomb equation demonstrated that the model system chosen could account for as much as 85% of the effect of the protein electric field on the reactants. Several preliminary computations were, moreover, required to establish the correct H-bond pattern of the catalytic water molecule (WAT in Fig. 2.6). Actually, in the crystal structure of Cdc42-GAP complex [60] the resolution of 2.10 Å did not enable determination of the positions of the hydrogen atoms. Thus, in principle, the catalytic water molecule could establish several different H-bond patterns with the amino acids of the protein-active site. Both classical and quantum mechanical calculations showed that WAT, in its minimum-energy conformation,

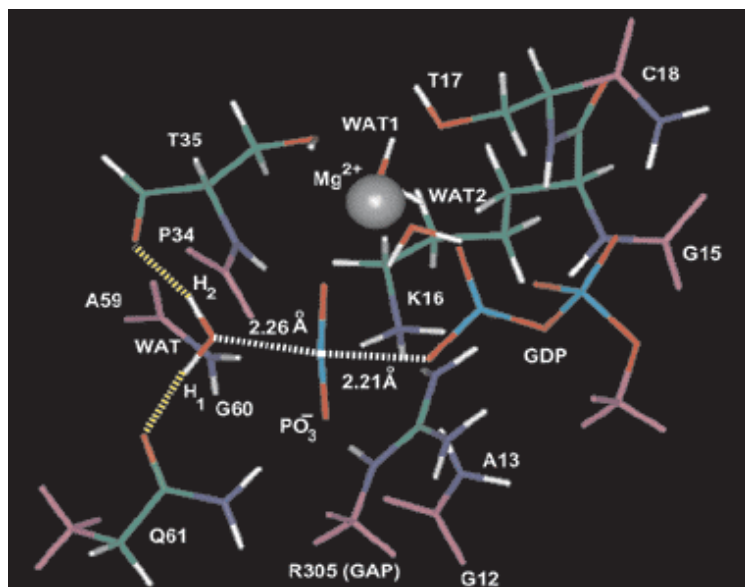


Fig. 2.6 The model system of the Cdc42-GAP enzymatic complex on which *ab initio* molecular dynamics simulations were performed. The catalytic water molecule is in-

dicated as WAT and the H-bonds are drawn in yellow. The atoms which remain in their crystallographic positions are colored magenta. Adapted from Ref. [53].

formed hydrogen bonds with the side chain of Gln61 and with the backbone of Thr35. WAT was, moreover, much more polarized in the protein than in the bulk, consistent with the hypothesis that the enzyme cavities act as a kind of supersolvent [63, 64]. When the suitability of the model system had been confirmed and the WAT conformation univocally determined, a reaction coordinate study was performed. To this end a process of constrained MD simulation [65] of the nucleophilic attack of WAT on the GTP γ -phosphate was conducted.

Starting from the transition state it was expected the reaction would evolve either forward to the products or backward to the reactants. During the unconstrained CPMD simulations, however, the system was always found to evolve towards the reactants. Because of this it was necessary to apply constrained dynamics to the principal coordinate reaction (the distance between WAT oxygen and GTP γ -phosphorus); this enabled investigation of the system evolution towards the products (Fig. 2.7).

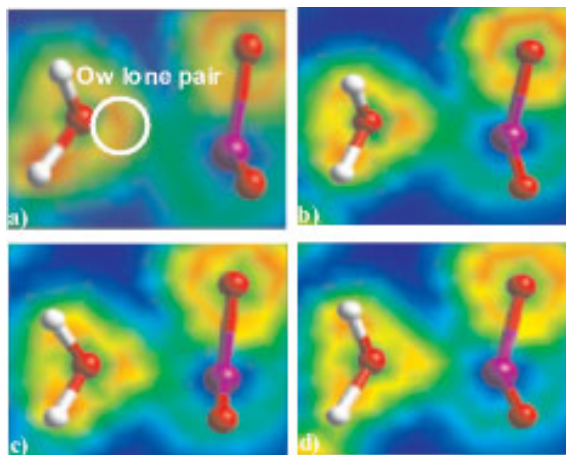


Fig. 2.7 Electron localization function (ELF) between the WAT oxygen atom and the γ -phosphorus in the different steps of constrained AIMD at 2.2 Å (a), 2.1 Å (b), 2.0 Å (c), and 1.9 Å (d). The red and yellow areas, indicating high electron density, clearly show the movement of WAT lone pairs from the oxygen to the phosphorus.

During the constrained CPMD simulations formation of the products was observed when the distance between the WAT and the γ -phosphate was 1.8 Å. One key step of the reaction was proton transfer from WAT to the Gln61 side chain. This process stabilized the highly nucleophilic species OH^- , which could complete nucleophilic attack of WAT on GTP. This finding was consistent with the X-ray structure of the Cdc42/TS mimic complex, which clearly shows that the NH_2 of the Gln61 side chain hydrogen bonds to the γ -phosphate, maintaining Gln61 very close to the catalytic water molecule (Fig. 2.6) and might also explain the 60–80% decrease in activity of the Q61E mutant [66]. Indeed, although glutamate is a stronger base than Gln, it is expected to be located further from the catalytic water than Gln, because of electrostatic repulsion between the carboxylate group of Glu and the triphosphate moiety of GTP. It should, however, be remarked that the simulations did not rule out the possibility of proton transfer from WAT to the solvent before the catalytic water reaches its reactive conformation. The WAT molecule is, in fact, highly polarized within the active site, because of the presence of many positively charged residues (i.e., Lys16, Mg^{2+} and GAP/Arg305) in the catalytic site of the pro-

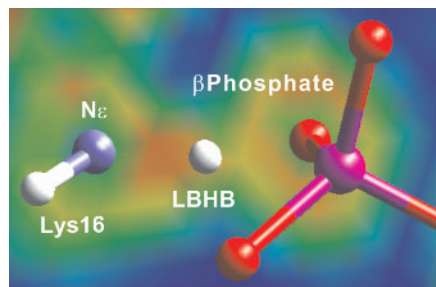


Fig. 2.8 The low-barrier hydrogen bond between Lys16 and an oxygen atom of GTP β -phosphate group. The electron localization function (ELF) is projected on the plane containing the three atoms involved in the LBHB. The red and yellow areas located between the

three atoms are a strong evidence of a three-center covalent bond. The covalent contribution of the LBHB might be responsible for further stabilization of the reaction transition state. Adapted from Ref. [53].

tein. In this context it was not surprising that simulations performed using the OH^- anion instead of a water molecule showed that the reaction occurred readily and quickly, bringing the system directly to the reaction products (i.e. GDP and inorganic phosphorus). This suggests that in the crystal structure of Cdc42-GAP with AlF_3 , the catalytic water might be considered deprotonated, thus, being an OH^- ion rather than an H_2O molecule. In addition, during the simulations a low-barrier hydrogen bond (LBHB) between Lys16 and β -phosphate was detected (Fig. 2.8). This LBHB turned out to be partly covalent in nature, in agreement with a previous suggestion [67]. The LBHB could, in principle, be very important in stabilizing the transition state, because it might provide a very large stabilization energy [68, 69]. It could, furthermore, enhance nucleophilic attack by water by significantly reducing the electron density on the γ -phosphate and consequently increasing its electrophilicity.

Summarizing, nucleophilic attack by water on GTP in the Cdc42-Cdc42GAP enzymatic complex has been investigated by use of DFT-based MD simulations. The model system enabled consideration of the most relevant physicochemical interactions between GTP, WAT, and the biological complex. In particular, the system seemed to include the atoms forming the pre-organized catalytic site and took into account the major contribution of the electric

field of the protein on the reactants. Quantum chemical and classical calculations suggested that the catalytic water forms hydrogen bonds to Gln61 and Thr35 during its nucleophilic attack. The computations provided no evidence for the postulated proton transfer from WAT to the γ -phosphate. Instead, the highly nucleophilic species OH^- could be stabilized by proton transfer to Gln61. This event showed that WAT was highly polarized within the active site without excluding the possibility that proton transfer occurred before WAT reached the final stable conformation. A second key event was the formation of a low-barrier hydrogen bond between Lys16 and β -phosphate. During the constrained dynamics, when the distance between WAT and GTP was 1.8 Å, a chemical bond was formed between γ -phosphorus and the catalytic water oxygen. In this simulation, removal of the constraint caused spontaneous evolution of the system towards the products (GDP and inorganic phosphate). These findings are fully consistent with X-ray data on the TS analog complex [60] and site-directed mutagenesis data on the Q61E mutant of Cdc42 [66].

2.4.1.2 HIV-1 Integrase

Retroviral integrases are a class of enzymes that catalyze the insertion of viral DNA into the host-cell nucleus. A deep understanding of the molecular aspects of the mechanism of action of the integrases might be relevant for medicinal chemists involved in anti-viral research, considering the continuous need to explore new viral targets, to counter the increasing problem of drug resistance [70]. Moreover, as stated above, an accurate description of the reaction TS might aid the rational design of TS analog inhibitors of such enzymes.

A paper reporting a DFT-based study aimed at determining the molecular mechanism of hydrolysis of phosphodiester bonds in HIV-1 integrase has recently appeared in the literature [71]. The study, by Bernardi et al., was conducted within a DFT framework using the hybrid DFT-HF exchange and correlation functional B3LYP as implemented in Gaussian 98 (Gaussian, Pittsburgh PA, USA, 1998) software. Locally dense basis sets (LBDS) [72], describing only the atoms directly involved in the reaction at 6-31G(d,p) level, were used to cope with the high computational demand of *ab initio* calculation. The paper pointed out that the

hydrolysis reaction catalyzed by HIV-1 integrase occurred through a metaphosphate TS, supporting a dissociative mechanism. It was remarkable that all attempts to identify an energetically stable intermediate failed. The hydrolysis of the phosphodiester bond catalyzed by this enzyme occurred through a typical S_N2 mechanism. Another important issue raised by the paper is that the electrostatic contribution of the protein bulk to the TS stabilization was as low as $0.4 \text{ kcal mol}^{-1}$, showing that protein regions apart from the active site did not participate directly in reducing the reaction-free energy. Although the system used was highly limited in size, the main points of the reaction mechanism seem to be well described by the DFT-based approach.

2.5

Studies on Transition Metal Complexes

The modeling of transition metal complexes is a difficult task that cannot be tackled properly by computational methods based on classical molecular mechanics. Determination of the effect of electrostatic contribution, which is taken into account well by the available force fields, fails when attempting to describe the geometry of transition metal complexes. Indeed, such geometries are heavily influenced by a high contribution of directionality, a typical feature of transition metal complexes. This task must, therefore, be tackled by means of *ab initio* calculations, because many important chemical and physical features of transition metal complexes can be predicted by first principles [73]. Of the quantum mechanical methods available those based on DFT enable better and more reliable description of the geometries and relative energies than traditional HartreeFock or MöllerPlesset perturbation theory at the second-order (MP2) methods, except for some weak bonding interactions [74]. Moreover, accounting for the effects at finite temperatures by means of DFT-based MD simulations enables proper description of the highly fluctuating behavior of such metal centers.

In this section, we present some applications of DFT-based methods to the characterization of the structural, electronic, and dynamic properties of metal complexes of pharmaceutical inter-

est. In this respect, besides the compounds taken into consideration here, and referred to as radiopharmaceuticals, study of the anti-cancer drug cisplatin, and its derivatives, should be mentioned; these electronic and geometric properties of these compounds are reviewed in Chapter 4 of this book.

2.5.1

Radiopharmaceuticals

Radiopharmaceuticals are metal complexes with radioactive nuclei that are widely used in medicine to monitor biological functions and to enable the imaging of tissues and organs. Nowadays, radioactive imaging techniques have become an indispensable tool in cancer diagnosis, and, furthermore, they can also be used therapeutically for *in-situ* treatment of cancerous tissues [75]. The pharmacokinetics of these agents, e.g. their travel through the body from the site of administration to the final location, is crucial for the efficiency of the diagnostic or therapeutic procedure, and it is strongly affected by the chemical environment encountered by the molecules. In this context, the physicochemical properties of the radiopharmaceutical play a critical role, but they are not usually well characterized. Accurate determination of such properties would clearly be desirable for rational design of new compounds, and DFT-based techniques can be quite useful at this regard. This is apparent from two papers published recently.

Investigation of crown thioether complexes of Re, Tc, and Ru (Fig. 2.9) in different redox environments has been performed by the group of Röthlisberger [76]. Their studies showed that these systems have interesting redox properties. Treatment of the doubly charged hexathioether complexes containing Tc or Re with mild reducing agents such as ascorbic acid, Zn, or SnCl_2 resulted in immediate carbon-sulfur bond cleavage and release of ethene. This bond fission did not occur for any of the other transition metals investigated. This observation was explained in terms of the stronger π -back donation from metal d-orbitals into empty C-S σ^* ligand orbitals by Tc and Re. Several experimental studies, including solution experiments, crystallographic data, and electron spray mass spectrometry, revealed the same trend [77, 78].

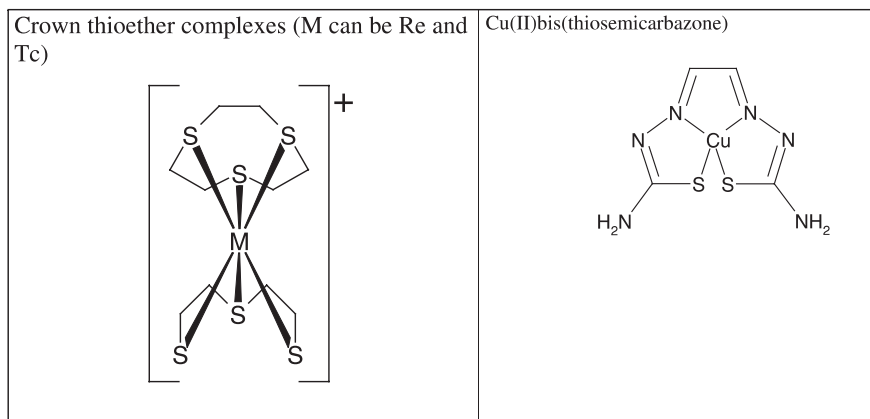


Fig. 2.9 Chemical formulas of crown thioether complexes and Cu(II)bis-thiosemicarbazone.

DFT-based calculations on the electronic structure of Cu(II)bis(thiosemicarbazone) (Fig. 2.9) analogs are reported in another recent paper [79]. This class of compound is gaining increasing importance as radionuclide agents for both positron emission tomography (PET) and targeted therapy. This DFT study addressed four fundamental properties of bithiosemicarbazones – the structure determined by geometry optimization, the vibrational properties derived from the harmonic frequencies, the protonation state evaluated by electrostatic potential and energetic calculations, and the redox properties correlated to molecular orbital energies. The chemical and electrochemical results presented by the authors support the hypothesis that intracellular reduction of the Cu(II) complexes can lead to two distinct patterns of chemical behavior. One would consist of rapid acid-catalyzed dissociation; the second is indicative of resistance to dissociation, enabling subsequent reoxidation by molecular oxygen.

This kind of study can be of great interest in medicinal chemistry (as is also apparent from the publication of the results from one such study in the *Journal of Medicinal Chemistry*), because it can be used to predict the behavior of the radiopharmaceutical on injection in the human body. These investigations can also be of relevance in the rational design of new radiopharmaceuticals performed with a view to designing more stable and less toxic compounds.

2.6 Conclusions and Perspectives

In recent times DFT-based methods have received increasing attention from the chemistry community [80]. Although the use of DFT in computational approaches to drug design is in its infancy, the studies described above indicate that it shows great promises in terms of both reliability and suitability for tackling issues the practical solution of which was previously considered difficult. In theoretical medicinal chemistry (as in other chemical disciplines that rely on computational models of molecules) one seeks the best compromise between accuracy of description and computational intensity. One might observe that in the design of drugs, rapidity in solving the problems is an added advantage, and in this sense simulations performed by means of DFT-based procedures produce high-level results in shorter computing time than other approaches. It is our opinion that the use of DFT-based methods should be strongly encouraged within the community of theoretical medicinal chemists and molecular modelers.

Of course, several aspects of the implementation of DFT in computational procedures need further improvement, especially the form of the exchange-correlation energy functional, which, as outlined above, is approximated in different ways. It is a matter of considerable debate whether current DFT approximations are good enough to deal with issues such as activation energies in chemical reactions, because the TS might not be described as accurately as the reactant or product states. The question is far from being settled, because experimental and precise quantum chemical knowledge of TS is rare, particularly in complex molecular and condensed-phase environments. It is, however, the opinion of many authors that DFT description of the TS might at least be qualitatively correct and it often seems to be in good quantitative agreement with experiment [81].

Within the framework of genomics and the rational development of new therapeutic strategies, a general question one might ask is “What is the role of quantum chemistry in medicinal chemistry in the world of macromolecular systems and in the era of exponentially increasing data on these systems?” Looking at the examples illustrated above we can postulate that details at the electronic level will definitely play a role in the development

of new molecular-based therapeutic strategies; *ab initio* calculations can therefore be considered as relevant complementary tools to structural and functional genomics for addressing clinical relevant issues. On the other hand, to increase the use of quantum chemical calculation within medicinal chemistry the size of the systems studied must be enlarged so it might be representative of the whole macromolecule or complex. In this context, DFT is a very promising approach, and even more so if implemented in the QM/MM approach that combines the precision of quantum chemical calculation with the power of molecular mechanics for handling large systems and their dynamics.

2.7

References

- HÖLTJE, H.-D.; FOLKERS, G. in: *Methods and Principle in Medicinal Chemistry*. R. MANNHOLD, H. K. H. TIMMERMAN (eds), VCH, Weinheim, 1997; Vol. 5.
- LEACH, A. R. in: *Molecular modelling. Principles and applications*. Addison Wesley, 1996.
- VEDANI, A.; ZBINDEN, P.; SNYDER, J. P.; GREENIDGE, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4987–4994.
- BOHM, H. J. *J. Comput. Aid. Mol. Des.* **1992**, *6*, 61–78.
- BISSANTZ, C.; FOLKERS, G.; ROGNAN, D. *J. Med. Chem.* **2000**, *43*, 4759–4767.
- KUBINYI, H. *Drug Discov. Today* **1997**, *2*, 457–467.
- HOHENBERG, P.; KOHN, W. *Phys. Rev.* **1964**, *136*, 864.
- KOHN, W.; SHAM, L. J. *Phys. Rev.* **1965**, *140*, 1133.
- KOCH, W.; HOLTHAUSEN, M. C. *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, Germany, 2001.
- CAR, R.; PARRINELLO, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
- CARLONI, P.; RÖTHLISBERGER, U. in: *Theoretical Biochemistry – Processes and Properties of Biological Systems*, Elsevier, 2001.
- EICHINGER, M. T. P.; HUTTER, J.; PARRINELLO, M. *J. Chem. Phys.* **1999**, *110*, 10452–10467.
- LAIO, A.; VAN DE VONDELE, J.; RÖTHLISBERGER, U. *J. Chem. Phys.* **2002**, *116*, 6941–6947.
- OKAZAKI, R.; KORNBERG, A. J. *Biol. Chem.* **1964**, *239*, 275–284.
- ELION, G. B.; FURMAN, P. A.; FYFE, J. A.; DE MIRANDA, P.; BEAUCHAMP, L.; SCHAEFFER, H. J. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5716–5720.
- FYFE, J. A.; KELLER, P. M.; FURMAN, P. A.; MILLER, R. L.; ELION, G. B. *J. Biol. Chem.* **1978**, *253*, 8721–8727.
- REARDON, J. E.; SPECTOR, T. J. *Biol. Chem.* **1989**, *264*, 7405–7411.
- BONINI, C.; FERRARI, G.; VERZELETTI, S.; SERVIDA, P.; ZAPPONE, E.; RUGGIERI, L.; PONZONI, M.; ROSSINI, S.; MAVILIO, F.;

- TRAVERSARI, C.; BORDIGNON, C. *Science* **1997**, *276*, 1719–1724.
- 19 GAMBHIR, S. S.; HERSCHMAN, H. R.; CHERRY, S. R.; BARRIO, J. R.; SATYAMURTHY, N.; TOYOKUNI, T.; PHELPS, M. E.; LARSON, S. M.; BALATONI, J.; FINN, R.; SADELAIN, M.; TJUVAJEV, J.; BLASBERG, R. *Neoplasia* **2000**, *2*, 118–138.
- 20 CULVER, K. W.; RAM, Z.; WALLBRIDGE, S.; ISHII, H.; OLDFIELD, E. H.; BLAESE, R. M. *Science* **1992**, *256*, 1550–1552.
- 21 KUSSMANN-GERBER, S.; KUONEN, O.; FOLKERS, G.; PILGER, B. D.; SCAPOZZA, L. *Eur. J. Biochem.* **1998**, *255*, 472–481.
- 22 VERZELETTI, S.; BONINI, C.; MARKTEL, S.; NOBILI, N.; CICERI, F.; TRAVERSARI, C.; BORDIGNON, C. *Hum. Gene Ther.* **1998**, *9*, 2243–2251.
- 23 CHRISTIANS, F. C.; SCAPOZZA, L.; CRAMERI, A.; FOLKERS, G.; STEMMER, W. P. *Nat. Biotechnol.* **1999**, *17*, 259–264.
- 24 WILD, K.; BOHNER, T.; AUBRY, A.; FOLKERS, G.; SCHULZ, G. E. *FEBS Lett.* **1995**, *368*, 289–292.
- 25 WILD, K.; BOHNER, T.; FOLKERS, G.; SCHULZ, G. E. *Protein Sci.* **1997**, *6*, 2097–2106.
- 26 BENNETT, M. S.; WIEN, F.; CHAMPNESS, J. N.; BATUWANGALA, T.; RUTHERFORD, T.; SUMMERS, W. C.; SUN, H.; WRIGHT, G.; SANDERSON, M. R. *FEBS Lett.* **1999**, *443*, 121–125.
- 27 BROWN, D. G.; VISSÉ, R.; SANDHU, G.; DAVIES, A.; RIZKALLAH, P. J.; MELITZ, C.; SUMMERS, W. C.; SANDERSON, M. R. *Nat. Struct. Biol.* **1995**, *2*, 876–881.
- 28 CHAMPNESS, J. N.; BENNETT, M. S.; WIEN, F.; VISSÉ, R.; SUMMERS, W. C.; HERDEWIJN, P.; DE CLERQ, E.; OSTROWSKI, T.; JARVEST, R. L.; SANDERSON, M. R. *Proteins* **1998**, *32*, 350–361.
- 29 PROTA, A.; VOGT, J.; PEROZZO, R.; PILGER, B.; WURTH, C.; MARQUEZ, V.; RUSS, P.; SCHULZ, G.; FOLKERS, G.; SCAPOZZA, L. *Biochemistry* **2000**, *39*, 9597–9603.
- 30 VOGT, J.; PEROZZO, R.; PAUTSCH, A.; PROTA, A.; SCHELLING, P.; PILGER, B.; FOLKERS, G.; SCAPOZZA, L.; SCHULZ, G. E. *Proteins* **2000**, *41*, 545–553.
- 31 ALBER, F.; KUONEN, O.; SCAPOZZA, L.; FOLKERS, G.; CARLONI, P. *Proteins* **1998**, *31*, 453–459.
- 32 PILGER, B. D.; PEROZZO, R.; ALBER, F.; WURTH, C.; FOLKERS, G.; SCAPOZZA, L. *J. Biol. Chem.* **1999**, *274*, 31967–31973.
- 33 SULPIZI, M.; SCHELLING, P.; FOLKERS, G.; CARLONI, P.; SCAPOZZA, L. *J. Biol. Chem.* **2001**, *276*, 21692–21697.
- 34 SAENGER, W. *Principles of Nucleic Acid Structure*, Springer, New York, **1984**.
- 35 MARQUEZ, V. E.; EZZITOUNI, A.; SIDDIQUI, M. A.; RUSS, P.; IKEDA, H.; GEORGE, C. *Nucleosides Nucleotides* **1997**, *16*, 1431–1434.
- 36 SCHELLING, P.; FOLKERS, G.; SCAPOZZA, L. *Anal. Biochem.* **2001**, *295*, 82–87.
- 37 MARZARI, N.; VANDERBILT, D. *Phys. Rev. B* **1997**, *56*, 12847–12865.
- 38 BOHM, H. J. *J. Comput. Aided Mol. Des.* **1998**, *12*, 309–323.
- 39 ROGNAN, D.; LAUEMOLLER, S. L.; HOLM, A.; BUUS, S.; TSCHINKE, V. *J. Med. Chem.* **1999**, *42*, 4650–4658.
- 40 GILSON, M. K.; GIVEN, J. A.; BUSH, B. L.; MCCAMMON, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.

- 41 KOLLMAN, P. A. *Acc. Chem. Res.* **1996**, *29*, 461–469.
- 42 HOLLOWAY, M. K.; WAI, J. M.; HALGREN, T. A.; FITZGERALD, P. M. D.; VACCA, J. P.; DORSEY, B. D.; LEVIN, R. B.; THOMPSON, W. J.; CHEN, L. J.; DESOLMS, S. J.; GAFFIN, N.; GHOSH, A. K.; GIULIANI, E. A.; GRAHAM, S. L.; GUARE, J. P.; HUNGATE, R. W.; LYLE, T. A.; SANDERS, W. M.; TUCKER, T. J.; WIGGINS, M.; WISCOUNT, C. M.; WOLTERS-DORF, O. W.; YOUNG, S. D.; DARKE, P. L.; ZUGAY, J. A. *J. Med. Chem.* **1995**, *38*, 305–317.
- 43 PITERA, J.; KOLLMAN, P. *J. Am. Chem. Soc.* **1998**, *120*, 7557–7567.
- 44 HEAD, R. D.; SMYTHE, M. L.; OPREA, T. I.; WALLER, C. L.; GREEN, S. M.; MARSHALL, G. R. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.
- 45 SHEN, J.; QUIOCHO, F. A. *J. Comput. Chem.* **1995**, *16*, 445–448.
- 46 XU, G.; MCLEOD, H. L. *Clin. Cancer Res.* **2001**, *7*, 3314–3324.
- 47 PAULING, L. *Am. Sci.* **1948**, *36*, 50–58.
- 48 SCHRAMM, V. L. *Annu. Rev. Biochem.* **1998**, *67*, 693–720.
- 49 GRAVES, J. D.; KREBS, E. G. *Pharmacol. Ther.* **1999**, *82*, 111–121.
- 50 COHEN, P. *Curr. Opin. Chem. Biol.* **1999**, *3*, 459–465.
- 51 PATRICK, G. N.; ZUKERBERG, L.; NIKOLIC, M.; DE LA MONTE, S.; DIKES, P.; TSAI, L. H. *Nature* **1999**, *402*, 615–622.
- 52 MEIJER, L.; LECLERC, S.; LEOST, M. *Pharmacol. Ther.* **1999**, *82*, 279–284.
- 53 CAVALLI, A.; CARLONI, P. *J. Am. Chem. Soc.* **2002**, *124*, 3763–3768.
- 54 SHINJO, K.; KOLAND, J. G.; HART, M. J.; NARASIMHAN, V.; JOHNSON, D. I.; EVANS, T.; CERIONE, R. A. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9853–9857.
- 55 COSO, O. A.; CHIARIELLO, M.; YU, J. C.; TERAMOTO, H.; CRESPO, P.; XU, N.; MIKI, T.; GUTKIND, J. S. *Cell* **1995**, *81*, 1137–1146.
- 56 OLSON, M. F.; ASHWORTH, A.; HALL, A. *Science* **1995**, *269*, 1270–1272.
- 57 LIN, R.; BAGRODIA, S.; CERIONE, R.; MANOR, D. *Curr. Biol.* **1997**, *7*, 794–797.
- 58 SYMONS, M. *Trends Biochem. Sci.* **1996**, *21*, 178–181.
- 59 LAMARCHE, N.; HALL, A. *Trends Genet.* **1994**, *10*, 436–440.
- 60 NASSAR, N.; HOFFMAN, G. R.; MANOR, D.; CLARDY, J. C.; CERIONE, R. A. *Nat. Struct. Biol.* **1998**, *5*, 1047–1052.
- 61 WITTINGHOFFER, A. *Curr. Biol.* **1997**, *7*, R682–R685.
- 62 SCHLICHTING, I.; REINSTEIN, J. *Nat. Struct. Biol.* **1999**, *6*, 721–723.
- 63 WARSHEL, A. *J. Biol. Chem.* **1998**, *273*, 27035–27038.
- 64 WARSHEL, A.; PAPAZYAN, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211–217.
- 65 CARLONI, P.; SPRIK, M.; ANDREONI, W. *J. Phys. Chem. B* **2000**, *104*, 823–835.
- 66 LERM, M.; SELZER, J.; HOFFMEYER, A.; RAPP, U. R.; AKTORIES, K.; SCHMIDT, G. *Infect. Immun.* **1999**, *67*, 496–503.
- 67 ALBER, F.; CARLONI, P. *Protein Sci.* **2000**, *9*, 2535–2546.
- 68 CLELAND, W. W. *Biochemistry* **1992**, *31*, 317–319.
- 69 CLELAND, W. W. *Arch. Biochem. Biophys.* **2000**, *382*, 1–5.
- 70 YOUNG, S. D. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 402–410.

- 71 BERNARDI, F.; BOTTONI, A.; DE VIVO, M.; GARAVELLI, M.; KESERU, G.; NARAY-SZABO, G. *Chem. Phys. Lett.* **2002**, 362, 127.
- 72 WRIGHT, J. S.; JOHNSON, E. R.; DILABIO, G. A. *J. Am. Chem. Soc.* **2001**, 123, 1173–1183.
- 73 DAVIDSON, R. (ed.) *Chemical Reviews* **2000**, 100, 351–818.
- 74 COTTON, F. A.; FENG, X. J. *J. Am. Chem. Soc.* **1997**, 119, 7514–7520.
- 75 DILWORTH, J. R.; PARROTT, S. J. *Chem. Soc. Rev.* **1998**, 27, 43.
- 76 MAGISTRATO, A.; MAURER, P.; RÖTHLISBERGER, U. *J. Phys. Chem. B* **2002**, in press.
- 77 MULLEN, G. E. D.; FÄSSLER, T. F.; WENT, M. J.; HOWLAND, K.; STEIN, B.; BLOWER, P. J. *J. Chem. Soc., Dalton Trans.* **1999**, 3759–3766.
- 78 MULLEN, G. E. D.; BLOWER, P. J.; PRICE, D. J.; POWELL, A. K.; HOWARD, M. J.; WENT, M. J. *Inorg. Chem.* **2000**, 39, 4093–4098.
- 79 MAURER, R. I.; BLOWER, P. J.; DILWORTH, J. R.; REYNOLDS, C. A.; ZHENG, Y.; MULLEN, G. E. *J. Med. Chem.* **2002**, 45, 1420–1431.
- 80 CAVALLI, A., Ed. *Quant. Struct.–Act. Relat.* **2002**, 21.
- 81 CAR, R. *Quant. Struct.–Act. Relat.* **2002**, 21, 97–104.

3

Applications of Car-Parrinello Molecular Dynamics in Biochemistry – Binding of Ligands in Myoglobin

CARME ROVIRA

Abstract

This chapter summarizes our work on the binding of ligands in myoglobin using density-functional theory (DFT) combined with molecular dynamics (MD), within the Car-Parrinello approach. The first part of this investigation will be devoted to analysis of the interplay between the structure, energy, and dynamics of the binding of O₂, CO, and NO to the heme active center. The calculations show that the heme porphyrin substituents (–CH=CH₂, –CH₃, and –CH₂CH₂COOH) do not affect the structural and electronic properties of the porphyrin, e.g. the deformation from planarity or energy splitting among spin states. Instead, the proximal histidine amino acid (His93) substantially increases the binding strength of the Fe–CO and Fe–O₂ bonds, whereas the opposite is true for Fe–NO, in agreement with thermodynamic studies of heme proteins and biomimetics. Calculations for larger porphyrin complexes, e.g. the *picket-fence-oxygen* myoglobin biomimetic, show that the structure of the Fe–ligand bonds is insensitive to the presence of bulky porphyrin substituents. Room-temperature molecular dynamics simulations reveal that the dynamics of CO is characterized by small, albeit extremely complex, displacements around its equilibrium position, with small fluctuations of the FeCO tilt and bend angles ($\leq 13^\circ$). This, and the observation that the energy cost of small deformations of the FeCO bond is marginal, confirm that such small deformations should not have any bearing on the protein discrimination for CO. In contrast, the dynamics of the O₂ ligand is quite anharmonic and characterized by frequent jumps (every 4–6 ps) among the porphyrin quadrants. Hybrid QM/MM calculations based on DFT combined with a classical force field (CHARMM) show that

the heme–CO structure is quite rigid and not affected by the conformation of the distal pocket. This excludes any relationship between FeCO distortion and the different CO absorption bands observed in the IR spectra of MbCO. In contrast, both the CO stretch frequency and the strength of the CO ··· His64 interaction are highly dependent on the orientation and tautomerization of His64. Our results imply that His64 is protonated at N_ε and show that both O₂ and CO are stabilized by interaction with His64. The larger interaction found for O₂ supports the conclusion that hydrogen bonding is the origin of protein discrimination of CO.

3.1

Introduction

The oxygen-carrying proteins hemoglobin (Hb) and myoglobin (Mb) have often been used as examples of protein conformation, dynamics, and function [1]. Their biological function, the binding and release of oxygen (O₂), occurs in the active center and is modulated by a large polypeptide framework. The latter is engineered to control the binding of O₂ and discriminate against the binding of endogenous ligands such as carbon monoxide (CO) [2].

Figure 3.1 shows the molecular structure of heme, the active center of myoglobin and hemoglobin. It is built from an iron-porphyrin (FeP) substituted with two propionate, two vinyl, and four methyl groups. As shown in Fig. 3.2, there is a covalent bond between the heme iron and the nitrogen atom of the His93 residue (the so-called *proximal histidine*). The opposite face of the iron-porphyrin, denoted the *distal side*, is free and ready to bind oxygen and other ligands such as carbon and nitric monoxides (CO, NO).

According to X-ray and neutron diffraction structures [3, 4] the binding of CO to the heme leads to a bent FeCO unit. The Fe–C–O angle is, however, found to be linear in synthetic models of the protein (*biomimetic* molecules). Because of this, it was originally thought that the FeCO distortion was responsible for the well known discrimination of the protein against CO – the affinity ratio CO/O₂ is lower in the protein than in biomimetic systems [1]. In

Fig. 3.1 The molecular structure of heme b (also called protoporphyrin IX), the active center of myoglobin, hemoglobin, catalases, and peroxidases, among other heme proteins.

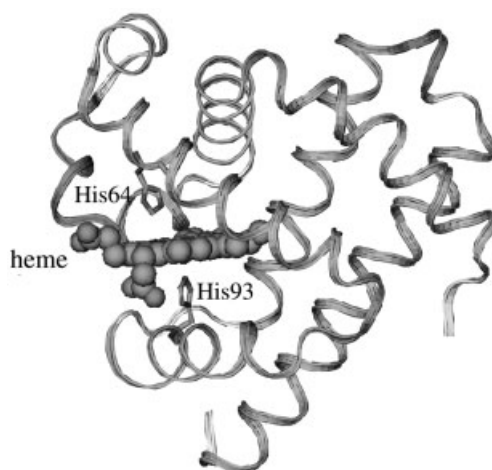
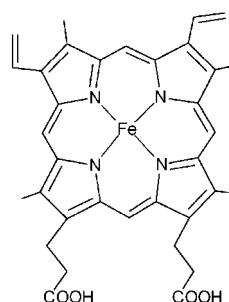


Fig. 3.2 The structure of myoglobin (deoxy form, PDB entry 1AGN, at 1.15 Å resolution [3f]). The heme active center is highlighted (van der Waals spheres), as are the proximal and distal histidines (His93 and His64, respectively, shown as sticks).

other words, the protein weakens the interaction of the CO ligand with the heme compared with that of the O₂ ligand. It was, however, found that the relationship between the distortion of FeCO and the protein discrimination against CO was not straightforward. First, the wide range of FeCO angles (120°–173°) reported in X-ray studies [3] seems to indicate that precise quantification of the heme–CO structure requires higher X-ray resolution than that currently available. Secondly, kinetic and thermodynamic measurements for a variety of proteins did not find an unambiguous relationship between CO affinity and FeCO distortion [2b]. Spec-

troscopic studies [5], on another hand, predicted just a small distortion ($\Delta \text{FeCO } 173^\circ$) which, according to recent theoretical investigations, has negligible energetic cost [6]. As a consequence, whether or not the FeCO distortion has a functional role is nowadays regarded as very questionable [7] and the ligand binding control is believed to have a different origin, possibly hydrogen-bond stabilization or electrostatic interactions [2, 8].

A sensitive probe of electrostatic interactions in the distal pocket is provided by the structural and vibrational properties of the Fe–CO unit [9]. The bound CO ligand exhibits three main infrared (IR) absorption bands, denoted A_0 , A_1 , and A_3 , with vibrational frequencies 1965 cm^{-1} , 1949 cm^{-1} , and 1933 cm^{-1} , respectively. These bands, which change relative intensity and wavenumber depending on temperature, pressure, pH, or solvent [10], are used to identify functionally different conformational substrates of MbCO, denoted taxonomic substates [11]. Nevertheless the relationship between the A states and specific structural features of the protein has not yet been clarified.

The A_0 component is observed on reducing the pH or mutating the distal histidine residue (His64) [9a, 12]. In addition, an X-ray study of MbCO at low pH [3b] has demonstrated that His64 is far from the ligand and out of the heme pocket. On this basis the A_0 state is usually associated with a protein substate in which the CO is in an apolar environment.

It was proposed early on that the A_1 and A_3 states correspond to different degrees of distortion of the FeCO unit caused by steric interactions in the heme pocket [13a]. Because large CO distortion would weaken the Fe–CO bond, a relationship was assumed between the A states and protein CO discrimination. Several spectroscopic studies, however, support the hypothesis that it is the polarity of the heme pocket that determines the ligand vibrations [2b, 8a, 9, 12, 13b]. In particular, it is assumed that the A states arise from different rotational conformations and/or tautomerization states of the distal histidine. Nevertheless, a correspondence between each CO stretch frequency and a specific heme-pocket conformation has not yet been found. Another relevant question that remains unsolved concerns the tautomerization of His64. Although neutron diffraction study of MbCO assigned a proton to the delta nitrogen of His64 (N_δ), with the lone pair of the epsilon nitrogen (N_ϵ) pointing towards the CO [4],

other experimental studies have suggested that His64 is instead protonated at N_e [9b, 13b]. This latter is in analogy with MbO₂ [14] and MbCN [15]. The structural and vibrational properties of the FeCO bond in MbCO are, therefore, not yet fully understood.

In contrast with CO, the binding of O₂ results in a bent Fe–O₂ unit, as indicated by X-ray and neutron structures of MbO₂ [3e, 14, 16], HbO₂ [17], and biomimetic systems [18]. Nevertheless, a wide range of values has been reported for the FeO₂ unit [14, 16, 17]. Several studies also suggest that the O₂ is hydrogen-bonded with the N_e proton of His64, both in MbO₂ and in α -Hb [14, 16, 17]. There is, however, no evidence of hydrogen bonding in β -HbO₂, in which free rotation of the ligand around its equilibrium position is expected [17]. Recent EPR measurements in cobalt-substituted Hb have also found evidence of O₂ rotation [19]. In addition, the fourfold disorder found in the crystal structure of biomimetic systems [18] by both Mössbauer and NMR experiments [20] has been interpreted as dynamic O₂ motion.

In comparison with CO and O₂, there is less structural information available on the binding of NO to myoglobin. In the first X-ray structure of nitrogen monoxymyoglobin (MbNO) [21] the FeNO angle was bent (Fe–N–O=112°), very different from the values reported for synthetic models (150°). In addition, the NO ligand is known to have a unique effect on binding to an iron-porphyrin derivative – it has a tendency to weaken the axial *trans* ligand bond [22]. This has been observed in MbNO and its biomimetics, and in other heme proteins such as guanylate cyclase. In this instance the effect is so strong that the binding of NO to the heme induces the release of the *trans* axial histidine [23].

Synthetic Models or Biomimetics of Myoglobin and Hemoglobin

Synthetic models of myoglobin and hemoglobin are complex molecules that mimic the stereochemical properties of the protein active center [24] and have oxygen affinities similar to those measured for the protein [25–27]. The first heme model that reversibly binds oxygen (i.e. the *picket-fence-oxygen* complex Fe(T_{piv}PP)(1,2-MeIm)(O₂), shown in Fig. 3.3) was obtained in the early nineteen-seventies by Collman and coworkers (T_{piv}PP=tetrapivalaminophenyl porphyrin; 2-meIm=2-methylimidazole) [18]. Research on synthetic models of the protein has led to a deeper understand-

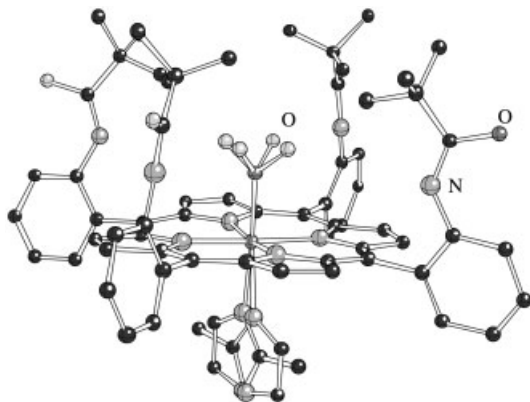


Fig. 3.3 The myoglobin biomimetic molecule $\text{Fe}(\text{T}_{\text{piv}}\text{PP})(2\text{-Melm})(\text{O}_2)$, known as *picket-fence-oxygen* [18d].

ing of the ligand-binding properties of myoglobin. For instance, several studies have shown that structural differences among heme models can vary the equilibrium constant (K_{eq}) of the O_2 and CO binding reaction. These changes have been attributed mainly to hydrogen-bonding and polar interactions [8a, 20, 28, 29], but steric interactions, porphyrin distortions and the interplay of various factors have also been proposed [25, 30, 31].

Despite all these studies of proteins and synthetic models, many essential aspects of the function of myoglobin and hemoglobin, e.g. the way the protein controls the binding of ligands (O_2 , CO, and NO), the precise structure of the Fe–ligand bonds and the structure-spin-energy relationships at the active center, are a topic of debate [2].

To enable understanding of all these issues, a precise knowledge of the intrinsic structural and dynamic properties of the heme–ligand bonds is of great interest. Theoretical studies could be very valuable in providing these data. It would also be extremely interesting to transcend a purely static point of view and examine fully the influence of thermal fluctuations. In this respect, studies based on classical molecular dynamics techniques have given considerable insight into dynamic features of myoglobin [32, 33], complementing the information obtained by use of structural techniques. For instance, the most likely pathways for ligand entry and exit from the active center of myoglobin have

been identified by means of MD simulations [33]. Subtle electronic/structural/spin changes (and chemical reactions) occurring at the active center cannot, however, be described with the force fields currently available. These processes are traditionally studied by quantum chemistry methods using simplified models of the active center, often rearranged to make them as symmetric as possible. These computations are typically performed on a fixed structure [34]. Nevertheless, in the last few years, the rapid development of efficient first-principles methods based on density-functional theory (DFT) has enabled modeling of the active center beyond the “frozen structure approximation”, thus capturing most of the chemistry of these systems [8b–d, 35]. Among all DFT-based approaches first principles molecular dynamics techniques such as the Car-Parrinello method are emerging as a useful tool to model the reactivity of reactive processes in systems of biological interest [36].

In this chapter we will summarize our work in the modeling of the interaction of ligands with myoglobin by use of Car-Parrinello molecular dynamics. In a first step, we will analyze the interplay between structure, energy, and dynamics in the binding of O₂, CO, and NO to the heme. In a second step, we will investigate the interaction between the heme and the rest of the protein, using hybrid quantum chemistry/molecular mechanics (QM/MM) methodology.

3.2

Computational Details

All calculations presented here are based on density-functional theory [37] (DFT) within the LDA and LSD approximations. The Kohn-Sham orbitals [38] are expanded in a plane wave (PW) basis set, with a kinetic energy cutoff of 70 Ry. The Ceperley-Alder expression for correlation and gradient corrections of the Becke-Perdew type are used [39]. We employ *ab initio* pseudopotentials, generated by use of the Troullier-Martins scheme [40]. The core-radii used, in au, were: 1.23 for the s, p atomic orbitals of carbon, 1.12 for s, p of N, 0.5 for the s of H, and 1.9, 2.0, 1.5, 1.97,

respectively, for the s, p, d, f atomic orbitals of Fe. The nonlinear core-correction [41] was used (with core-charge radius of 1.2 au).

The Car-Parrinello method [42, 43], based on a combination of a molecular dynamics (MD) algorithm with electronic-structure calculations using density-functional theory (DFT), has been used with success in the study of different systems of biological interest [36]. Successive use of quenching and annealing performed for approximately one picosecond was necessary to reach a final convergence of 10^{-5} and 5×10^{-4} au for electronic and ionic gradients, respectively. Structure optimizations were performed with no constraints starting from nonsymmetric structures. The convergence of our results with the energy cutoff in the PW expansion was investigated for an iron-porphyrin. The ordering of spin states was found to be insensitive to the PW cutoff value, and the energy differences changed only very slightly (within $0.5 \text{ kcal mol}^{-1}$). Structural data were found to be even less sensitive than energy differences to the PW cutoff [35f]. Molecular dynamics simulations at room temperature were performed using a time step of 0.12 fs, with the fictitious mass of the Car-Parrinello Lagrangian set to 700 au. The deuterium mass for the hydrogen atoms was used. The systems were enclosed in supercells of $16 \text{ \AA} \times 16 \text{ \AA} \times 20 \text{ \AA}$, periodically repeated in space. They were allowed to evolve for 2 ps in order to achieve vibrational equilibration. The MD was performed for total periods of 18 ps and 15.5 ps for FeP(Im)-CO and FeP(Im)-O₂, respectively. Hybrid QM/MM calculations were performed using the EGO-CPMD code [44], which is an interface between the EGO classical code, based on the CHARMM force field [46], and the CPMD code [45]. The interface between the QM and MM regions was treated using the link atom approximation [44]. Harmonic ligand stretch frequencies were computed from the diagonalization of the Hessian matrix obtained by numerical energy derivatives.

3.3 Myoglobin Active Center

3.3.1

Structure, Energy, and Electronic State

Figure 3.4 shows the models used in the calculations of the active center. They are of the type FeP–AB (FeP=iron-porphyrin, AB=CO, NO, O₂), Heme–AB (AB=CO, NO, O₂), and FeP(Im)–AB (Im=imidazole). The axial imidazole ligand mimics the effect of the proximal histidine amino acid (Fig. 3.2).

As a first step in this investigation the structures of the systems FeP, FeP–AB (AB=O₂, CO, NO) and FeP(Im)AB (Im=imidazole) were optimized. The lowest energy spin-state of each system was considered. This was found to be a triplet for FeP, singlet for FeP–CO, FeP–O₂, FeP(Im)CO, and FeP(Im)–O₂, and doublet for FeP–NO and FeP(Im)–NO, in agreement with experiments [18–26]. As shown in Fig. 3.5 for the O₂ example (top), the pentacoordinated FeP–AB complexes are characterized by having a curved porphyrin. This type of distortion reinforces the bonding between the Fe(*d_{z²}*) orbital and the 3σ_g orbital of the diatomic molecule (the *d_{z²}* orbital becomes hybridized by mixing some *s* and *z* character [47]).

The main structural data defining the optimized structures of the FeP–AB models are listed in Tab. 3.1. The Fe–CO bond is linear whereas Fe–CO and Fe–NO bonds are angular. The angle increases in the direction O₂ < NO < CO, which is also the same tendency found in structural analysis of myoglobin [14–21]. The binding energy with regard to dissociation of the diatomic ligand

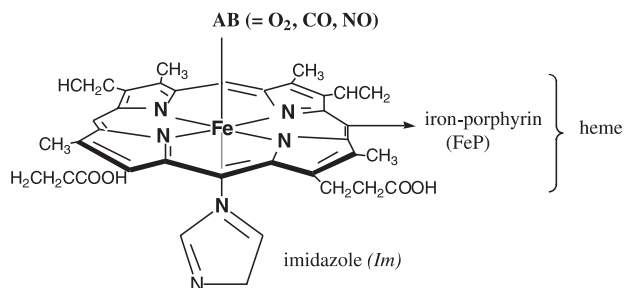


Fig. 3.4 Active center model used in the calculations.

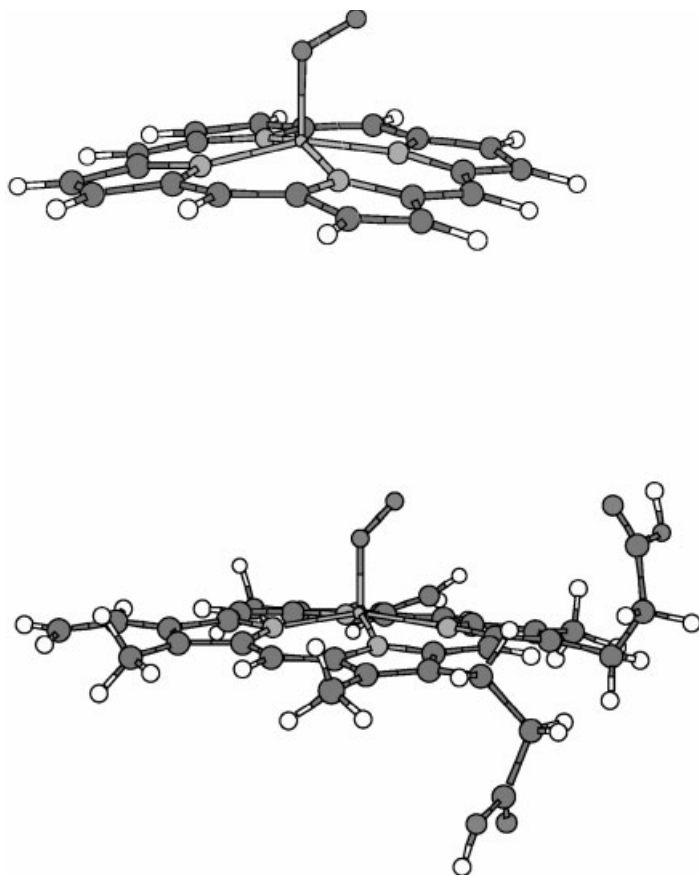
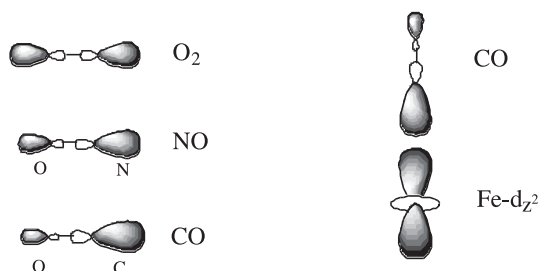


Fig. 3.5 (a) Optimized structure of the iron-porphyrin- O_2 model ($FeP-O_2$). (b) Optimized structure of the full heme b molecule complexed with oxygen ($heme-O_2$).

amounts to 9 kcal mol^{-1} for $FeP-O_2$, 26 kcal mol^{-1} for $FeP-CO$, and 35 kcal mol^{-1} for $FeP-NO$. The enhanced binding of CO and NO compared with O_2 can be understood in terms of the variation of the $Fe(d_{z^2})AB(3\sigma_g)$ interaction (Scheme 3.1). For CO and NO, the $3\sigma_g$ orbital is more polarized towards the C and N atoms (those binding to Fe), whereas for O_2 it is shared among both oxygen atoms. It is also worth mentioning that the energy of these systems changes little on rotation of the AB ligand relative to the Fe-A bond (less than 2 kcal mol^{-1}); this indicates that rotational motion of the ligand is likely to occur at room temperature.



Scheme 3.1

To relate these structure and/or energy changes to the properties of the protein, it is necessary to exclude any possible influence from the chemical groups closest to the iron-porphyrin (the $-(\text{CH}_2)_3\text{COOH}$, $-\text{CH}=\text{CH}_2$, and $-\text{CH}_3$ porphyrin substituents, and the proximal histidine residue). Additional calculations including the porphyrin substituents (Fig. 3.5b) revealed [35a] that neither the local structure of the Fe-diatomic bonds nor the binding energy of the ligand were different from those of models which did not include the porphyrin substituents.

In contrast, addition of an imidazole (Im) axial ligand (Fig. 3.4) does cause major changes in the binding of ligands. These changes are shown schematically in Fig. 3.6 and the corresponding optimized structures are listed in Tab. 3.1. Variations in the internal porphyrin structure have not been detailed, because these are minor. The only structural change resulting from Im is the loss of the porphyrin curvature. The binding energy of the Fe–O₂ and Fe–CO bonds is, however, substantially enhanced (left → center in Fig. 3.6). This can be understood in terms of the increase in the σ -donor character of the imidazole orbital which interacts with the d_{z^2} orbital of the iron [47]. In contrast, the energy of the Fe–NO bond hardly changes on addition of the imidazole, even though similar enhancement of the σ -donor character of the ligand-binding energy is expected. In this case, an antibonding orbital with strong Fe(d_{z^2}) character becomes occupied and, as a consequence, the Fe–imidazole bond becomes weaker. The balance between both effects results in the insensitivity of the Fe–NO bond to *trans*-axial ligation.

Similar trends were observed when the diatomic molecule was attached to the FeP–Im system (right → center in Fig. 3.6). The binding energy of the Fe–N_ε bond increases on binding of CO or O₂, but changes little (it even weakens) on NO binding. These

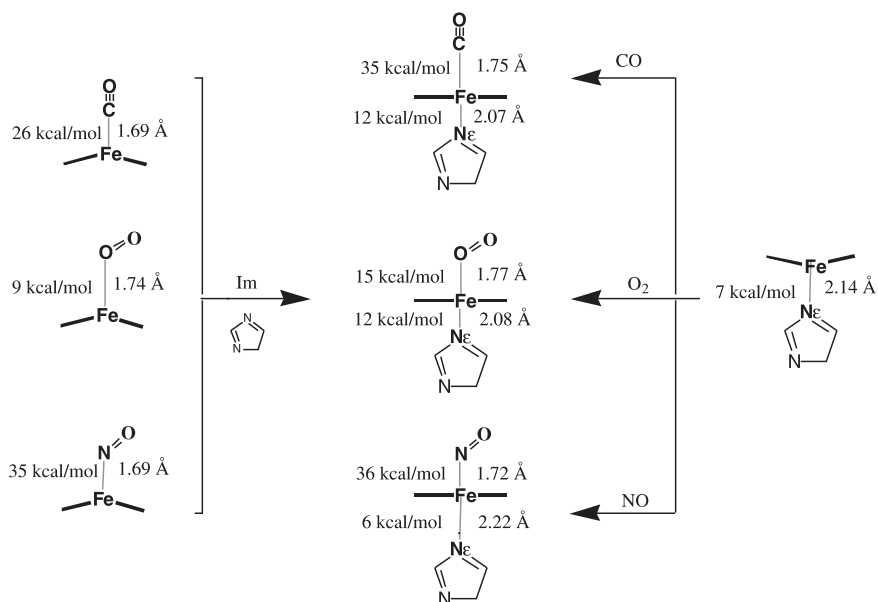


Fig. 3.6 Schematic diagram of the structural and energy changes on addition of an imidazole axial ligand to the FeP(AB) systems (left to right) or addition of a diatomic AB molecule to the FeP(Im) system (right to left), where AB=CO, O₂, NO.

types of change are consistent with the well-known *trans*-repulsive effect of the NO molecule when binding to iron-porphyrin derivatives. Studies of reactions of heme models with imidazole, CO, O₂, and NO show [22, 23] that addition of NO to an imidazole-bound iron-porphyrin weakens the Fe–Im bond whereas the reverse is true for CO and O₂. An extreme case of this is provided by the mechanism of activation of guanylate cyclase [23]. It has been proposed that binding of NO to the heme leads to the release of the proximal histidine amino acid, which in turn induces a conformational change in the protein [22].

It is also worth noting that the binding energy values obtained are in agreement with thermodynamic measurements of oxygen binding to myoglobin and biomimetic heme models, for which ΔH_0 values in the range 10–19 kcal mol⁻¹ are reported [25]. Recent experiments on FeT_{pyr}PH₂–NO have reported a binding enthalpy of 29 kcal mol⁻¹ [48], a value similar to that we obtained for FeP–NO. In addition, the large imbalance between the binding energies of CO and O₂ obtained in the gas phase empha-

Tab. 3.1 Main data defining the optimized structure of the FeP(AB) and FeP(Im)(AB) models investigated (AB=CO, NO, O₂). Distances are given in Å, angles in degrees, and energies in kcal mol⁻¹. Porphyrin nitrogens are denoted N_p and N_ε refers to one of the nitrogen atoms of the axial imidazole. The experimental values correspond to X-ray structures of heme models [18].

Structure	Fe–A	A–B	∠ Fe–A–B	Fe–N_p	Fe–N_ε	ΔE_{Fe-AB}
FeP–CO calc.	1.69	1.17	180	1.99	–	26
Expt.	–	–	–	–	–	–
FeP–NO calc.	1.69	1.19	150	2.00	–	35
Expt.	1.71(1)	1.12(1)	149(1)	2.02–1.99	–	–
FeP–O ₂ calc.	1.74	1.28	123	2.02–1.99	–	9
Expt.	–	–	–	–	–	–
FeP(Im)–NO calc.	1.72	1.20	138	2.02–2.01	2.22	36
Expt.	1.74(1)	1.12(1)	140	2.01(1)	2.18(1)	–
FeP(Im)–CO calc.	1.72	1.17	180	2.02	2.10(1)	35
Expt. ^{a)}	1.77(2)	1.12(2)	179(2)	2.02(1)	2.08	–
FeP(Im)–O ₂ calc.	1.77	1.30	121	2.02–2.01	2.08	15
Expt. ^{b)}	1.75(2)	1.2(1)	131(2)	1.98(2)	2.07(2)	–

a) The experimental model [18e] contains pyridine as an axial ligand instead of imidazole

b) The O–O bond distance of the crystal structure [18d, h] is very imprecise

sizes that the protein environment plays a major role in modulating the relative binding between both ligands.

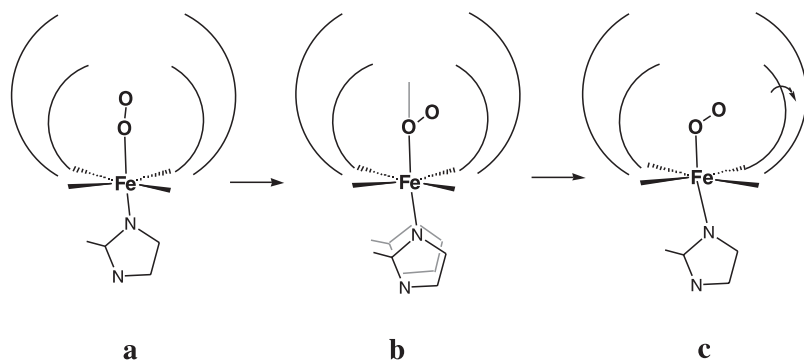
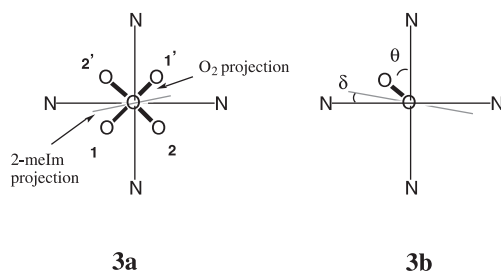
In summary, the calculations show that whereas the porphyrin substituents do not play a role in the structure/energy/spin properties of the heme-active center of myoglobin, the axial ligand has a major role in modulating the strength of the bond between the iron atom and the diatomic molecule – the binding of CO is strengthened by 66% and that of O₂ by 33%. In contrast, and because of the occupation of the Fe{d_{z²}} orbital, the binding of NO hardly changes in the presence of the imidazole axial ligand.

3.3.2

The Picket-fence-oxygen Biomimetic Complex**3.3.2.1 Interplay Structure/Electronic State**

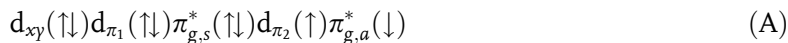
Structural optimization of the *picket-fence*-oxygen molecule (Fig. 3.3) was performed by taking its X-ray structure as a reference [18d]. To investigate the orientation preferences of the O₂ ligand a linear Fe–O–O angle was set initially. The computation was performed on the S=0 state within the LSD approximation. As shown in Scheme 3.2, the structure evolves rapidly toward a bent Fe–O–O angle (121°) with the O₂ axis projection located in the same porphyrin quadrant as the imidazole (Scheme 3.3b). This orientation corresponds to one of the four positions found experimentally (Scheme 3.3a) and suggests that orientation 1 could be the global minimum of the system.

A small displacement of the porphyrin substituents away from the O₂ molecule occurs during full relaxation of the molecular

**Scheme 3.2****Scheme 3.3**

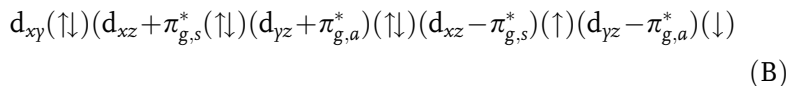
structure (b \rightarrow c in Scheme 3.2). This “opening of the picket cage”, is not equivalent for the four $T_{\text{piv}}P$ substituents but is more pronounced for the one closer to the terminal oxygen. Other changes that occur on bending are an increase of the O–O distance (from 1.28 Å in a to 1.30 Å in b, c) and of the Fe– N_{ϵ} distance (from 2.03 Å in a to 2.11 Å in b, c). As a consequence of the longer Fe– N_{ϵ} distance the steric interaction between the imidazole methyl and the porphyrin ring is partially relieved, which reduces the tilting of the imidazole by 3° [49]. These calculations indicate that changes in the Fe– N_{ϵ} distance are closely related to changes in the Fe–O–O angle and, in turn, to the energy of the Fe–O bond. Before discussing this aspect in detail, it is useful to analyze the changes in the electronic structure of the system.

The origin of the small structural differences observed when changing the Fe–O–O angle are related to changes in the electronic structure. The d-orbital configuration of Fe in the bent structure (b or c) is the same as that found for a small FeP(Im)(O₂) system [6b], an open-shell singlet; the filling of the higher occupied orbitals of this species can be shown schematically as:



This electron distribution follows the semiempirical model proposed by Hoffman et al. in the late nineteen-seventies, with some variation because of the spin polarization [51]. In this example one can clearly differentiate these orbitals as having either iron or oxygen character [52] (although a small d_{z^2} component appears in the third orbital, $\pi_{g,s}^*$, its relative contribution is very small). It follows from the above orbital assignment (A) that the Fe–O₂ bond can be formally described as Fe^{III}–O₂[−].

The electron distribution in the Fe–O–O linear conformation follows instead the scheme:



Here, because of the strong Fe/O₂ mixing, classification of these orbitals as Fe or O₂ character is not straightforward. Hence, neither the Fe^{III}–O₂[−] nor the Fe^{II}–O₂ formal description is applicable here. The change in electronic configuration (A \rightarrow B),

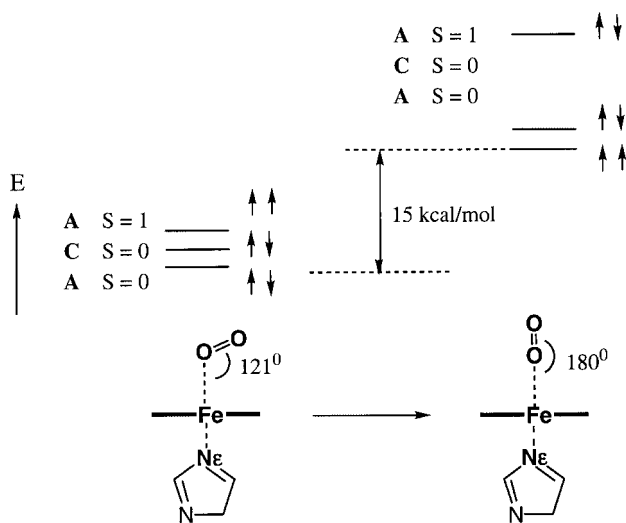


Fig. 3.7 Qualitative picture of the energies of different spin states of $\text{Fe}(\text{T}_{\text{pivPP}})(1,2\text{-Melm})(\text{O}_2)$, as a function of the Fe-O-O angle. Small energy differences are enlarged to aid visualization (the exact values are given in the text).

with the total disappearance d_{z^2} of contribution in the orbitals of B, is the cause of the shorter Fe-N_e distance in the linear structure. The charge transfer to oxygen also disappears, as is apparent from the shorter O-O distance.

All the structure and energy changes discussed so far refer to the $S=0$ state. Additional calculations were, however, also performed for higher spin states. The relative energy of these states with regard to Fe-O-O bending is illustrated in Fig. 3.7.

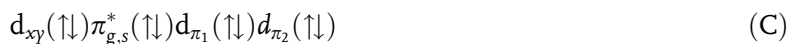
It is found that triplet ($S=1$) is the ground state for a linear Fe-O-O conformation, with the same electronic configuration as found for the open-shell singlet:

$$d_{xy}(\uparrow\downarrow)(d_{xz} + \pi_{g,s}^*)(\uparrow\downarrow)(d_{yz} + \pi_{g,a}^*)(\uparrow\downarrow)(d_{xz} - \pi_{g,s}^*)(\uparrow)(d_{yz} - \pi_{g,a}^*)(\uparrow)$$

thus following Hund's rule [53]. The structural features of this linear Fe-O-O triplet state are the same as those described for the linear Fe-O-O open-shell singlet state (B). For the bent Δ Fe-O-O , the lowest triplet state lies only 3 kcal mol^{-1} above the open-shell singlet. Its electronic configuration is very similar to that of the open-shell singlet state (A), with small mix-

ing of the d_{π_2} orbital with d_{xy} as a consequence of the loss of symmetry which occurs as a result of the bending. The total energy cost of distorting the FeO_2 moiety from the bent global minimum ($S=0$, open-shell) structure to the linear Fe-O-O conformation ($S=1$) is 15 kcal mol^{-1} . Because of this quite high value and the change in spin associated with it, a linear Fe-O-O is unlikely to occur as a result of room-temperature fluctuation of the atomic positions. This excludes it as a transition state for the mechanism of O_2 internal motion among the porphyrin quadrants.

The minimum of the triplet state corresponds to a larger Fe-O-O angle (131°) than the $S=1$ state (121°), as we also found for a small $\text{FeP(Im)(O}_2\text{)}$ system [6b]. On the other hand, an $S=0$ closed-shell state is well separated in energy in the linear conformation (22 kcal mol^{-1} relative to the ground triplet state), but becomes very close to the ground state in the bent conformation ($1.4 \text{ kcal mol}^{-1}$). Its electronic configuration can be shown schematically as:



in both conformations. For such a large system, a difference of $1.4 \text{ kcal mol}^{-1}$ is at the limit of the accuracy of the method used. There is, moreover, experimental evidence of a spin-paired electronic state by NMR experiments [25, 26, 29e, 29g]. Together with the fact that solvation or condensed-matter effects are not included in our treatment, it is not possible to differentiate between the two singlet states as being the ground state of the bent FeO_2 on the basis of these calculations.

In summary, three spin states ($S=0$ open-shell, $S=0$ closed shell and $S=1$) are in competition as the Fe-O-O angle bends. Although well separated in energy for a linear Fe-O-O angle, the three spin states become very close (within 3 kcal mol^{-1}), and two of them reverse in energetic order, when the Fe-O-O angle bends. There should, therefore, be a spin-crossing region along the Fe-O-O reaction coordinate where the three spin states could be mixed by spin-orbit interaction. Because these energy-spin relationships are shared with the small $\text{FeP(Im)(O}_2\text{)}$ complex, the existence of these three energetically close spin states seems to be a peculiarity of O_2 binding to a FeP(Im) derivative.

On the other hand, the results obtained show that variations in the Fe–O–O angle and the Fe–N_e distance, weakening the Fe–O bond, are closely related to changes in the spin state of the system – a triplet state is favored by a short Fe–N_e distance and large Fe–O–O angle, whereas the opposite favors a singlet state. In the context of the protein, this suggests that appropriate tension through the proximal histidine and/or steric hindrance opening the Fe–O–O angle could change the spin state of heme.

3.3.2.2 Optimized Structure and Energy of O₂ Binding

The most important structural data defining the optimized structure of the FeT_{piv}PP(2-meIm)(O₂) molecule are reported in Tab. 3.2. Only the most important structural data are listed, which are those defining the orientation of the O₂ and 2-meIm ligands. The computed values are compared with the experimental X-ray structure and with the structure obtained for FeP(Im)(O₂) (Section 3.3.1).

The computed structure is in good agreement with that found experimentally. Slight discrepancies arise only in parts of the structure not precisely known, e.g. the FeO₂ internal geometry (O–O_{exp} > 1.22, Δ Fe–O–O_{exp} < 129°) and the C(CH₃)₃ groups of

Tab. 3.2 Main parameters defining the optimized structure of the *picket-fence*-oxygen molecule, FeT_{piv}PP(2-meIm)O₂, in comparison with the FeP(Im)O₂ model. Distances are given in Å and angles in degrees.

Structural data	Calculated <i>picket-fence</i>-O₂ Fe(T_{piv}PP)(2-meIm)O₂	Calculated FeP(Im)-O₂	Experimental [18d] <i>picket-fence</i>-O₂ Fe(T_{piv}PP)(2-meIm)O₂
Fe–O	1.78	1.77	1.898(7)
O–O	1.30	1.30	>1.22(2) ^{a)}
Δ Fe–O–O	121.0	121.0	<129(2) ^{a)}
Fe–N _p	2.01	2.01	1.996(4)
Fe–N _e	2.11	2.08	2.107(4)
NH···O	3.09	–	–
$\delta^b)$	4.5	0.0	7.1

a) The Fe–O–O angle and the O–O distances are reported as upper and lower bounds, respectively, in the experimental determination

b) Deviation of the Fe–N_e bond from the heme perpendicular, because of the steric interaction of the 2-methyl substituent

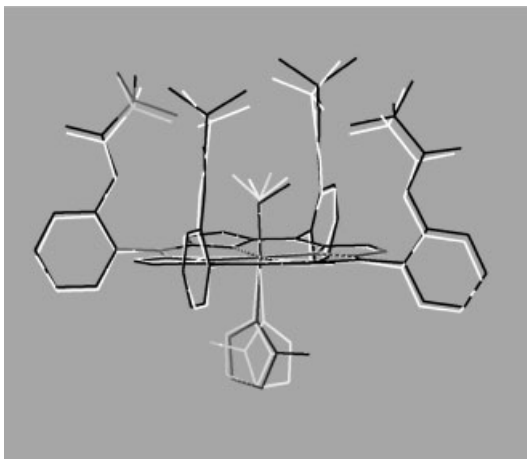


Fig. 3.8 Comparison of the *picket-fence* optimized structure (black line) with the X-ray structure (white line).

the $T_{\text{piv}}\text{P}$ substituents, which are affected by thermal motion and/or irresolvable disorder [18d]. A substantial increase of the O–O distance (1.30 Å) from its gas phase value (1.21 Å) is observed; this is a consequence of the charge transfer associated with the π -back-bonding $\text{Fe}\{d\} \rightarrow \text{O}_2\{\pi_g^*\}$. Fig. 3.8 contrasts the experimental structure (all positions of the disordered O_2 are displayed) with that calculated. The experimental “effective” C_2 symmetry structure for the $\text{Fe}(T_{\text{piv}}\text{PP})$ fragment results from an average over the four different orientations of the O_2 ligand (Scheme 3.3). As a consequence, the displacement of one of the $T_{\text{piv}}\text{P}$ substituents when the terminal oxygen gets close is not taken into account. This explains why the differences between the calculated and experimental values are larger for one of the $T_{\text{piv}}\text{P}$ substituents; it also explains the differences in the tilting of the 2-meIm axial ligand.

Additional calculations performed for different orientations of the oxygen molecule relative to the $\text{Fe}-\text{N}_p$ bonds (N_p = porphyrin nitrogen atom) reveal that there is an equivalent minimum in which O–O is rotated 180° around $\text{Fe}-\text{N}_p$. The transition state between both minima corresponds to the situation where O–O overlaps $\text{Fe}-\text{N}_p$, which is found to be only 1.8 kcal mol $^{-1}$ higher in energy than the first minimum (Scheme 3.3b). Given this small energy difference, rotation of the O_2 ligand around the $\text{Fe}-\text{O}$ bond

at room temperature probably occurs frequently. Indeed, we will see in Section 3.3.3 that at room temperature the O₂ rotates around the Fe–O bond on the picosecond timescale.

The possibility of the O₂ being hydrogen bonded to the amino group of T_{piv}P is supported by several experimental studies. In particular, the NMR chemical shift of the amide hydrogen has been taken as an indication of a hydrogen bond [29d]. The computed H···O distance (3.09 Å and 2.93 Å for the first and second minima respectively) is quite large to be classified as a hydrogen bond. Significant electrostatic stabilization of the ligand occurs, however, because the strength of the Fe–O₂ bond is significantly enhanced (32 kcal mol⁻¹). The energy increase is much larger than that observed on addition of an imidazole axial ligand (6 kcal mol⁻¹) and contrasts with the results obtained for protoheme complexes (Fig. 3.5), for which the binding energy was found to be insensitive to the presence of the porphyrin substituents.

The origin of the different behavior of *picket-fence* and heme is probably related to the different polarity of the porphyrin substituents. For *picket-fence* the dipole moment of the amide groups of each T_{piv}P is oriented toward the diatomic ligand, which results in a stabilizing interaction with the dipole of the FeO₂ or FeCO fragments. For heme the polar acidic groups (Fig. 3.5) are far from the ligand position (5–7 Å from the terminal atom of the diatomic ligand). Because of their relative position (on opposite sides of the porphyrin), moreover, the total electrostatic interaction with the diatomic would vanish. An estimate of the electrostatic interaction between the diatomic molecule and the porphyrin substituents can be obtained from the computed charges on the atoms. By use of the simple formula $U = \sum_{i,j} \frac{q_i q_j}{R_{ij}}$ we obtain a large stabilizing interaction of the ligand in the *picket-fence* complexes (21 kcal mol⁻¹ for O₂ and 15 kcal mol⁻¹ for CO) but the amount is almost negligible for the protoheme complexes. Because we do not observe changes in the Fe–O and Fe–C bond orders on attaching the picket substituents, we conclude that electrostatic interactions are responsible for the large increase in binding energy.

That the environment could play a major role in determining ligand binding to iron in heme models and, probably, in myoglobin also, is in agreement with the conclusions of site-directed

mutagenesis experiments [2b] and recent evidence for the high oxygen affinity of *Ascaris* hemoglobin [54]. The stabilizing effect of a polar environment is also consistent with measurements of ligand binding affinities in heme models [8a, 26, 28, 31, 48]. Although a wide range of measurements has been reported in the literature, low affinities are found, in general, for binding pockets of low polarity [8a, 28, 31]. An example of the significant influence of the binding pocket polarity and stereochemistry is provided by the high O₂ affinity recently found in dendrite porphyrins [28d].

In summary, the computed structure of the FeT_{piv}PP(2-meIm)(O₂) biomimetic is in good agreement with the X-ray structure, and complements it in the determination of several details, e.g. the subtle structural deformations induced by O₂ motion, the structure of the Fe–O₂ and Fe–Im bonding, and the position of hydrogen atoms. In addition, our calculations show that, whereas the structure of the Fe–ligand bonds does not change in the presence of porphyrin substituents, the ligand-binding energy is very sensitive to the polarity of these substituents.

3.3.3

Heme-Ligand Dynamics

As a way of sketching the flexibility of the Fe–ligand bonds at room temperature we performed molecular dynamics simulations of models of the active center [35e]. First, a simulation for the FeP(Im)CO model was performed, for a total period of 18 ps and an average temperature of 300 K. This simulation provided evidence that the CO ligand moves rapidly around its linear equilibrium structure and, simultaneously, the imidazole ligand undergoes relatively slow rotations around the Fe–N_ε bond. Figure 3.9 (top) shows the molecular structure of FeP(Im)CO with the axis definition. As a way of sketching the motion of the CO ligand we monitored the projection of the C and O atoms on the average plane defined by the four porphyrin nitrogens. Figure 3.9 (bottom) shows the trajectory of both C and O projections on this plane. The trajectories seem rather complex and concentrated around the iron atom, with that for the oxygen being more spread (~ 0.4 Å from the center). Nevertheless, relative to

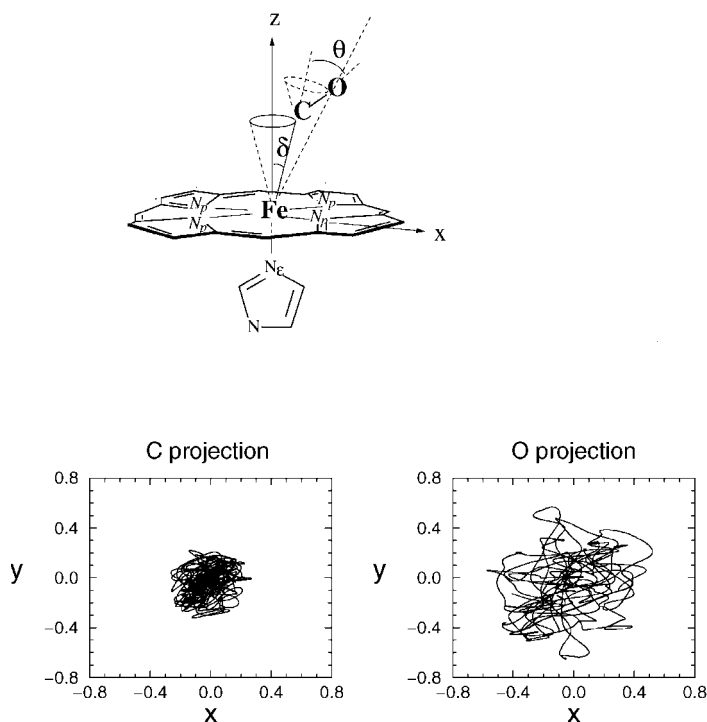


Fig. 3.9 Top: Tilt (δ) and bend angles (θ) used to define the structure of the Fe–CO unit in FeP(Im)(CO). Bottom: Configurational space sampled by the projection of the

C–O axis on the porphyrin plane. The Fe atom is located at the center of the plot, with the x and y axes aligned with the Fe–N_p bonds.

the size of the porphyrin (Fe–N_p = 2.02 Å), the whole spread of values shown in Fig. 3.9 (bottom) corresponds to just a very small area over the iron atom. Further analysis of the time evolution of the CO orientation (data not shown here) reveals that the projection of the C–O axis on the porphyrin plane visits all the porphyrin quadrants in a very short time (~ 0.5 ps). The global picture that can be inferred from our simulation is, therefore, that of an essentially upright FeCO unit, with the CO ligand undergoing rapid complex motion within a very small region around its equilibrium position.

An interesting property that can be extracted from the simulation is the allowed distortion of the Fe–C–O fragment. We quantified this distortion by using the tilt (δ) and bend (θ) angles, which have been often related to the protein discrimination

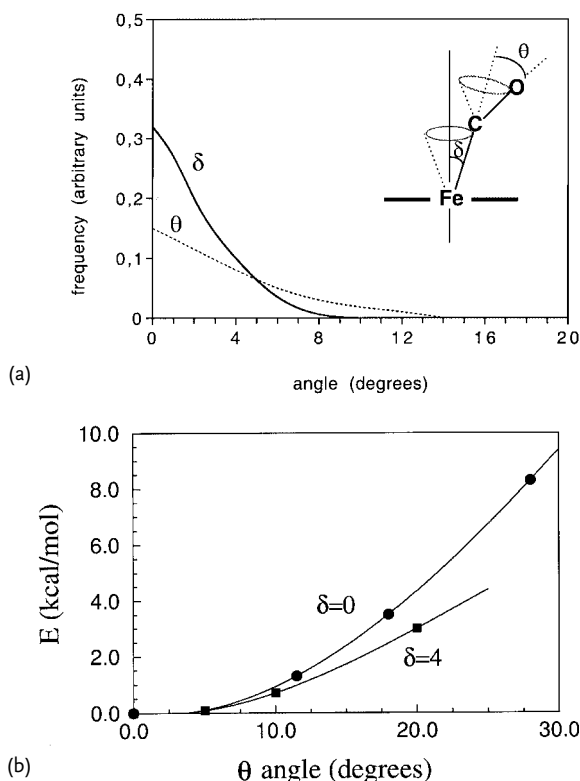


Fig. 3.10 (a) Frequency distribution corresponding to the tilt (δ) and bend angles (θ) commonly used to describe the distortion of the FeCO fragment. (b) Energy (E) required to bend and tilt the Fe-C-O angle.

against CO. Figure 3.10 a shows the probability distribution of the δ and θ angles obtained from our trajectory. It is apparent there is a sizable probability of small fluctuations ($\delta < 8^\circ$, $\theta < 13^\circ$) occurring, but larger deformations do not occur. We therefore conclude that for FeCO not perturbed by the environment, small δ and θ deviations (similar to those reported after spectroscopic studies [5]) can occur solely as a result of thermal motion and not as a consequence of steric hindrance by the protein. This is consistent with results from static calculations [6], which have predicted that small variations of δ and θ do not require significant amounts of energy. As is apparent from Fig. 3.10 b, bending of the CO by up to 10° (combined with tilting by up to 4°) involves only $0.6 \text{ kcal mol}^{-1}$, which is approximately the thermal

energy of the FeCO unit. Our results thus confirm that the small deformations of the FeCO bond predicted by spectroscopic studies are within the range of FeCO conformations sampled at room temperature. Thus these small deformations should not be relevant to protein discrimination against CO. It should also be noted that, given the complex motion of the ligand, the instantaneous structure of the FeCO unit cannot be easily defined in terms of the δ and θ angles alone; the problem should best be regarded as that of a highly dynamic FeCO moiety, sampling many different conformations with different probability in a short time.

An analogous simulation was performed for the heme-O₂ model, FeP(Im)(O₂), for a total time of 15.5 ps. Figure 3.11 shows selected snapshots of this simulation. The position of the O₂ molecule relative to the porphyrin will be described by use of the torsion angle defined in Fig. 3.12. Initially, the O–O axis projection lies on one of the porphyrin quadrants, although it undergoes large oscillations between the two closest Fe–N_p bonds. After 2.2 ps, however, the O–O axis projection jumps over one Fe–N_p bond toward the next porphyrin quadrant. Five of these transitions occurred during the whole simulation, with an average time interval of 4–6 ps. Every transition occurs as a result of the rotation of the O₂ around the Fe–O axis and involves a conformation with a more open Fe–O–O angle (124–129°) and a slightly tilted Fe–O bond (3–5°) relative to the heme perpendicular (i.e. the perpendicular direction relative to the average plane defined by the four N_p atoms). This confirms that the OO/Fe–N overlapping configuration is the transition state for the dynamic motion of O₂ between the porphyrin quadrants observed experimentally [20].

The average structure of the FeO₂ fragment that we obtain from the simulation is very similar to the equilibrium structure of the FeP(Im)–O₂ system – Fe–O=1.75 Å, \sphericalangle Fe–O–O=122°, O–O=1.30 Å. The Fe–O–O angle is, however, slightly more open (124°) and the Fe–O distance is larger (1.86 Å), because of the anharmonicity of the corresponding vibrational modes. This structure is, however, very different from that reported after X-ray study of their closely related biomimetics system, the *picket-fence* molecule, for which Fe–O=1.75 Å, \sphericalangle Fe–O–O=129°, 133°, and O–O=1.15 Å, 1.17 Å. We believe that this is because X-ray techniques do not measure distances and angles directly, but rather positions of maximum

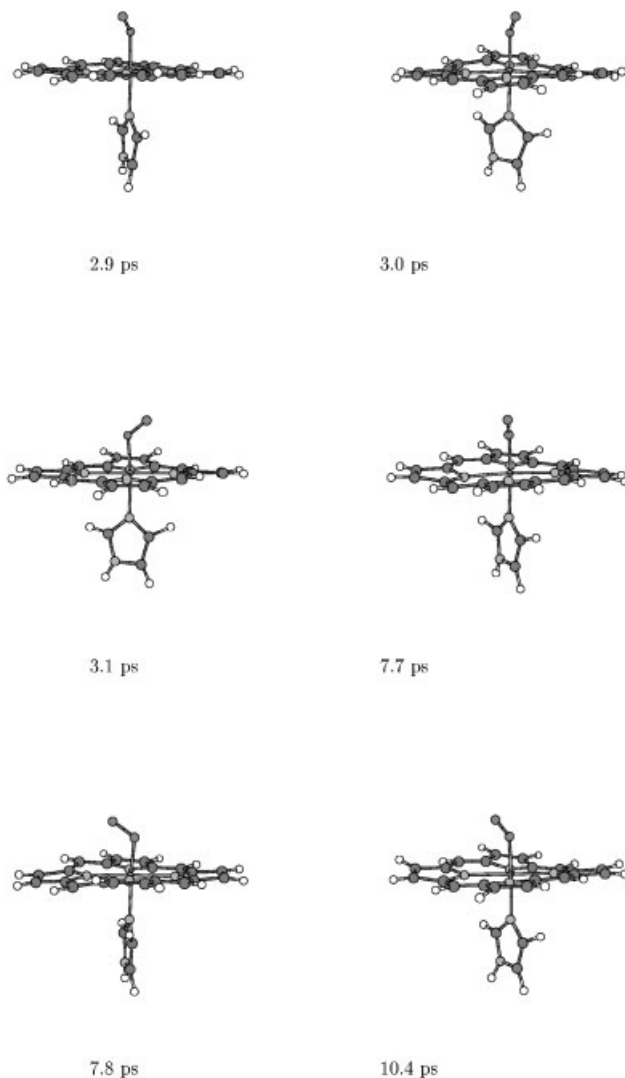


Fig. 3.11 Snapshots of the dynamics of the oxyheme FeP(Im)O_2 model.

probability, from which the other structural properties are deduced. We can, in fact, justify this argument from our MD trajectory. If we compute the position of maximum probability for the Fe and O atoms and use these positions to obtain the structural data, we find: $\text{Fe-O} = 1.72 \text{ \AA}$, $\angle \text{Fe-O-O} = 139^\circ$, and $\text{O-O} = 1.19 \text{ \AA}$. Although these values are in better agreement with experiment, they are not

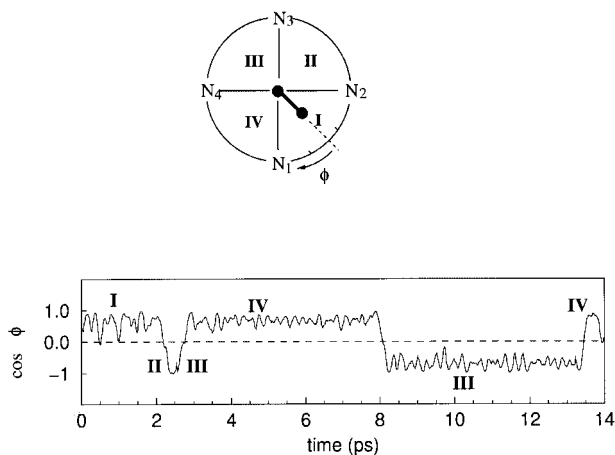


Fig. 3.12 Rotation of the O_2 ligand around the Fe–O bond in $FeP(Im)O_2$ as a function of time. The Φ angle corresponds to the N_p –Fe–O–O torsion.

realistic (the O–O distance, for instance, is shorter than the gas-phase value of 1.23 Å). Therefore, although this reconciles theory and experiment, it also illustrates the risk of assigning a static structure to a highly dynamic moiety such as the FeO_2 unit. Similar considerations are valid for data reported for the proteins. In particular, the Fe–O–O angles obtained from neutron and X-ray studies of oxy-myoglobin [14] and oxyhemoglobin [17] are very discrepant (115° for Mb O_2 , 153° for α -Hb and 159° for β -Hb) and not even sampled in our simulation. An Fe–O–O angle very similar to our computed value (122°) was, on the other hand, reported after a recent X-ray study [3f]. It might be argued that the Fe–O–O angle in the protein is affected by hydrogen-bonding to the His64 residue. It has, however, recently been shown [8b] that such a hydrogen bond does not significantly modify the structure of the FeO_2 unit.

Overall, our simulation reveals highly anharmonic dynamics for the O_2 ligand. It undergoes large-amplitude oscillations within one porphyrin quadrant and jumps from one to the other within 4–6 ps. This is consistent with the highly dynamic nature of O_2 bound to heme proposed by several experiments with proteins and biomimetics [18h, 19, 20] especially those which lack a hydrogen bond at the terminal oxygen. NMR experiments have shown that ligand rotation occurs in these models [20b, c], on the basis of the equivalence of the pyrrole proton resonances. Our results

suggest that, for nonhydrogen-bonded O₂, precise determination of the rate of rotation would require picosecond time resolution.

3.4

Interaction of the Heme with the Protein

The results described in the sections above demonstrate that several properties of heme–ligand bonds – their spin state, structure, and dynamics – are well reproduced by use of models of the active center. This leads to an obvious question – what is the effect of the rest of the protein on the ligand binding properties? It is well known that the protein has a key role, for instance, in controlling the entry and release, and the binding affinity, of different ligands [2d]. The values we compute for the binding energies have the right trends (NO > CO > O₂) but not the right absolute values. The relative binding energy for CO (20 kcal mol⁻¹ larger than that of O₂) is far too high to justify the experimental CO/O₂ ratio in the equilibrium constants for the ligand binding reaction ($\sim 10^{-3}$). This is, however, not surprising, because it is well known that the relative binding CO/O₂ is controlled by the protein [1]. A sensitive proof of this influence is the different CO stretching frequencies that appear in the IR spectrum of MbCO [11]; the origin of these has not yet been clarified.

In an attempt to aid interpretation of the IR spectrum of MbCO we decided to model the full protein by use of a hybrid quantum mechanics/molecular mechanics approach (QM/MM), to evaluate changes in the CO stretching frequency for different protein conformations. The QM/MM method used [44] combines a first-principles description of the active center with a force-field treatment (using the CHARMM force field) of the rest of the protein. The QM-MM boundary is modeled by use of link atoms (four in the heme vinyl and propionate substituents and one on the His64 residue). Our QM region will include the CO ligand, the porphyrin, and the axial imidazole (Fig. 3.13). The vinyl and propionate porphyrin substituents were not included, because we had previously found they did not affect the properties of the Fe–ligand bonds (Section 3.3.1). It was, on the other hand, crucial to include the imidazole of the proximal His (directly bonded to the

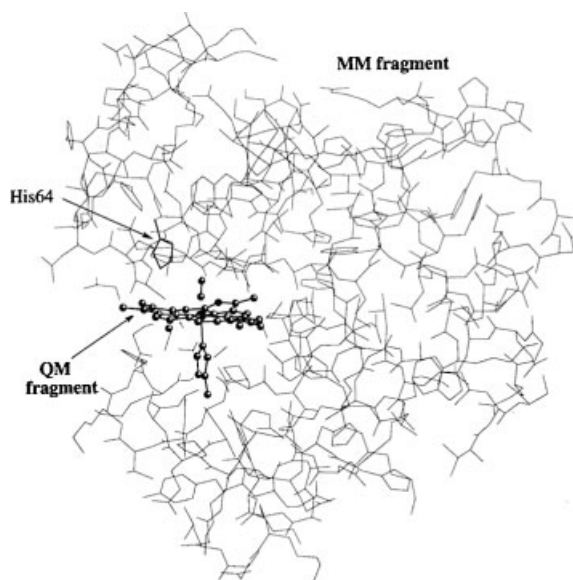


Fig. 3.13 QM/MM partition used in calculations for MbCO.

heme), because it strengthens the Fe–CO bond (Section 3.2.1). With this QM-MM partition we can be confident that the energy/spin/structure relationships of the heme are well described. The protein is, in addition, enveloped in a 37-Å sphere of equilibrated TIP3P water molecules (Fig. 3.14), to take solvation effects into account. The number of QM and MM atoms treated in the calculation are 63 and 20000, respectively.

Before starting the calculations, it is important to choose an appropriate initial structure for the protein. The X-ray structure is not a good starting point, because it corresponds to an average among the many different instantaneous protein conformations. It is physically more meaningful to take snapshots of previous MD simulations using the same force field as in the QM/MM calculations.

Classical simulations of MbCO using the CHARMM force field were performed for different tautomerization states of the distal histidine residue (His64) [33]. These simulations showed that when His64 is protonated at N_δ (denoted the N_δ tautomer) it often rotates such that it exposes either the N_δ –H bond or the unprotonated N_ϵ atom to the CO, as depicted in Scheme 3.4. We

Tab. 3.3 Main structural data defining the optimized heme–CO complex for each protein conformation I–V. The last row corresponds to the results obtained for the heme–CO isolated model (Section 3.3.1).

Structure	Type of interaction	O···X	C–O	Fe–C	Fe–N _p	∠ Fe–C–O
Expt.	CO···N _ε	4.0–2.60	1.09–1.21	1.73–2.21	2.01–2.06	120–172
I	CO···N _ε	3.39	1.16	1.76	2.00–2.02	177.3
II	CO···N _δ	3.47	1.16	1.75	1.98–2.03	179.3
III	CO···H–N _ε	2.69	1.17	1.74	1.99–2.02	176.1
IV	CO···N _δ	3.90	1.16	1.74	1.99–2.02	175.7
	CO···H–C	2.18				
V	CO···N _ε	6.03	1.16	1.75	1.99–2.03	177.6
Heme–CO	–	–	1.17	1.72	2.02	180.0

the heme–ligand structure is not very sensitive to protein conformation. We obtained a similar structure in all cases, and the data were not far from the gas-phase values (last row in Tab. 3.3). Most importantly, the Fe–CO angle is always essentially linear, even when the proton of the distal His is close to the CO (III). This shows that, in contrast with the results obtained from structural analyses [3], the FeCO fragment in MbCO is essentially linear.

The CO stretch frequencies change significantly depending on the protein conformation. As is apparent from Tab. 3.4, when the N–H proton of the distal His is close to the CO (II, III), the CO frequency decreases (the shift is negative). If, instead, the unprotonated nitrogen becomes close to CO (I) the shift is positive. The largest downshift (-23 cm^{-1}) is given by the CO···H–N interaction in the N_ε tautomer (III). This is expected, because the possible N–H···OC hydrogen bond becomes geometrically more favored in this configuration than in the others. The configuration with His64 far from the ligand (V) practically does not shift the CO frequency. This is in agreement with mutagenesis experiments showing enhancement of the highest CO frequency peak in the IR spectrum when His64 is replaced for an apolar residue [9a, 12]. Configuration V practically does not shift the CO frequency either. This could be because of the opposite effect of both the protonated and the unprotonated nitrogen at intermediate distances. Our calculations also revealed an inverse correlation between the $\Delta\nu_{\text{CO}}$ and $\Delta\nu_{\text{FeC}}$ values ($\Delta\nu_{\text{CO}}$ increases as $\Delta\nu_{\text{FeC}}$

Tab. 3.4 Shift of the C–O and Fe–C stretch frequencies relative to the isolated heme–CO system for each of the protein conformations I–V. Hydrogen bond energies are also listed. Distances are given in Å, frequencies in cm^{-1} and energies in kcal mol^{-1} .

Structure	Type of interaction	O···X	$\Delta\nu_{\text{CO}}$	$\Delta\nu_{\text{Fe-C}}$	$\Delta E_{\text{O···X}}$
I	CO···N _ε	3.39	+14	-62	+2.0
II	CO···N _δ	3.47	-14	18	-2.5
III	CO···H–N _ε	2.69	-23	61	-3.4
IV	CO···N _δ	3.90	-4	10	-0.9
	CO···H–C	2.18			
V	CO···N _ε	6.03	-1	10	-0.1
Heme–CO	–	–	0	0	–

decreases). This general trend has been observed across a wide range of heme proteins and biomimetic systems [28c, 9b].

The frequency changes in Tab. 3.4 can be rationalized in terms of variations in the Fe–CO back-bonding (i.e. the interaction of the Fe{d} levels with the empty π_{CO}^* orbitals). When a positive charge (e.g. the proton of His64) approaches the CO, the π_{CO}^* orbitals are energetically stabilized. As they become closer in energy to the Fe{d} orbitals, the back-bonding increases and, as a consequence, the CO frequency decreases. Thus, a downshifted ν_{CO} is obtained for the CO···H–N type interactions (II, III). In contrast, a negative charge (e.g. the nitrogen lone pair) approaching the CO reduces the back-bonding and increases ν_{CO} . In agreement with this argument, an upshift of ν_{CO} is obtained for arrangement I.

Similar changes in the CO stretch frequency have been observed experimentally. As mentioned early in this chapter the IR spectrum of carbon monoxymyoglobin contains three main CO stretching bands ($A_0 \approx 1965 \text{ cm}^{-1}$, $A_1 \approx 1949 \text{ cm}^{-1}$, and $A_3 \approx 1933 \text{ cm}^{-1}$). That we obtain different CO frequency shifts for different His64 conformations supports the interpretation that the distal His determines the A states [9b]. Our results can, moreover, aid interpretation of the peaks in the spectrum. First, it should be noted that our zero frequency (i.e. the conformation where the CO is not influenced by the protein environment) corresponds to the A_0 state. Thus, our computed frequency shifts are related to the frequency differences between each peak of the IR spectrum and the A_0

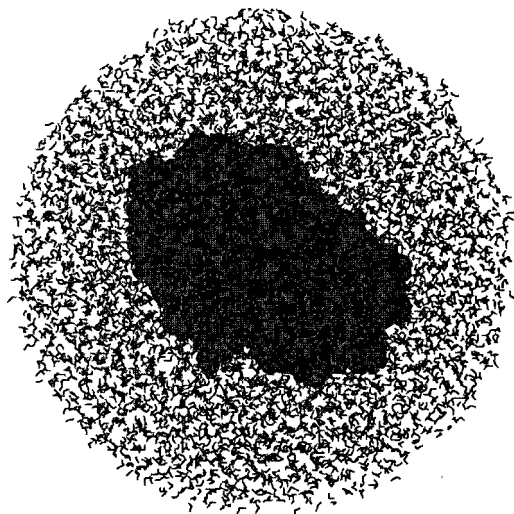


Fig. 3.14 Protein embedded in the water shell.

peak. Because both A_1 and A_3 are down-shifted relative to A_0 , only configurations II and III can possibly contribute to the IR spectrum; conformation I, leading to an up-shifted ν_{CO} , is excluded. This brought us to the conclusion that conformation I does not contribute to the dynamics of the heme pocket and, therefore, that the His64 residue is not protonated at N_δ . Conversely, the frequent rotation of His64 (as is apparent from classical MD simulations) would lead to the occurrence of conformation I and thus to the appearance of a peak up-shifted from A_0 . Taking into account that only one of the His64 tautomers can be responsible for the A states [11a], we propose that both A_1 and A_3 originate from type III conformations of the N_ϵ tautomer [8c].

As for vibrational frequencies, the interaction energy of the ligand with the distal residue (last column in Tab. 3.4) is highly dependent on the conformation and protonation state of the distal His. Configuration I leads to a repulsive interaction (2 kcal mol^{-1}), whereas the interaction is favorable when the protonated nitrogen is close to the CO. We find stabilization to be largest for the N_ϵ tautomer, because now the $\text{CO} \cdots \text{H}-\text{N}_\epsilon$ interaction becomes geometrically more favored than for the other configurations. This is at variance with the common assumption [8a] that only binding of O_2 could be stabilized by interaction with the distal histidine. Our calculations,

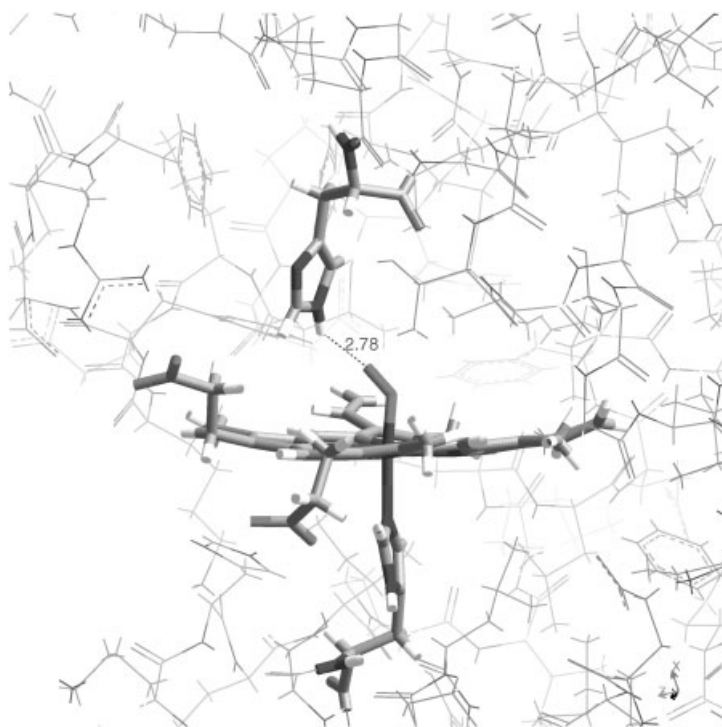


Fig. 3.15 Optimized structure of oxymyoglobin (MbO_2) in the region around the heme-active center.

however, support recent Resonance Raman measurements [13b] that provide spectroscopic evidence of a hydrogen bond between CO and His64.

For comparison purposes we performed a calculation replacing the CO by O_2 , to estimate the strength of the analogous $\text{O}_2 \cdots \text{His64}$ interaction. It is commonly accepted that His64 is protonated at N_ϵ in MbO_2 [14]. We thus performed the calculation on arrangement III in which the proton of N_ϵ is pointing towards the ligand (Fig. 3.15). In this instance the H-bond amounts to $-5.1 \text{ kcal mol}^{-1}$, a value $1.8 \text{ kcal mol}^{-1}$ larger than the maximum energy gain obtained for the CO ligand (Tab. 3.4). Our calculations therefore find the O_2 to be more stabilized than the CO, although the interaction is significant for both. The different amount of stabilization found for CO and O_2 (3 kcal mol^{-1} and 5 kcal mol^{-1} , respectively) might explain how the protein reduces the CO/O_2 affinity ratio relative to the synthetic analogs. In this

respect our study [55] supports several experimental and theoretical investigations that suggest that binding of O₂ is enhanced by hydrogen bonding with the His64 residue [2, 8b].

3.5

Conclusions

In this study we have quantified the structure, energetics, and electronic structure of the binding of CO, NO, and O₂ in myoglobin and in the *picket-fence* biomimetic complex. All three ligands induce significant curvature in the heme when binding to the iron atom, although the planarity of the porphyrin is restored by the *trans* binding of an imidazole ligand. The heme substituents are found not to influence the energy of the Fe-ligand bonds (Fe–CO, O₂, NO). Instead, the ligand-binding energy is enhanced by the presence of the polar porphyrin substituents of the *picket-fence* molecule. For both complexes (heme and *picket-fence*) the structure of the Fe–ligand bond is insensitive to the presence of the porphyrin substituents.

Significantly different binding properties are observed for the three ligands. The Fe–O₂ bond is much weaker than the Fe–CO and Fe–NO bonds, and the binding angle increases on going from O₂ to NO. Most of these changes can be traced back to electronic structure differences, and can be predicted by use of a simple picture of molecular orbital interaction. The role of the axial imidazole ligand is to reinforce the ligand-binding energy for O₂ and CO, whereas the opposite is found for NO. In this case, occupation of the Fe(d_{z²}) orbital results in a longer and weaker bond with the imidazole ligand. This explains the observed *trans* effect of NO on binding heme proteins such as guanylate cyclase.

Rotation of the oxygen around the Fe–O bond involves a small energy barrier (~ 2 kcal mol⁻¹), suggesting that several rotational conformations could be available at room temperature. Indeed, our molecular dynamics simulations show that the O₂ ligand undergoes large-amplitude oscillations within one porphyrin quadrant, jumping to another quadrant on the picosecond timescale. The dynamics of the FeCO unit are characterized by rapid mo-

tion of the ligand around its equilibrium position, with a maximum Δ FeCO distortion of 13° .

Hybrid QM/MM calculations based on density-functional theory combined with the CHARMM force field highlight the effect of the distal pocket conformation on the properties of the Fe–CO bond in MbCO. Our calculations show that, again, the local structure around the Fe atom is insensitive to the heme environment. An essentially linear FeCO bond is found for different distal pocket conformations; this leads to the conclusion that the heme–CO structure is quite robust and not influenced by the protein environment. Instead, both the CO stretch frequency and the strength of the CO \cdots His64 interaction seem to be very dependent on the conformation of the protein and, in particular, on the orientation and type of His64 tautomer. This can be rationalized in terms of the changes in the Fe–CO back-bonding when a positive/negative charged group approaches the CO ligand. In addition, our calculations reveal that the distal histidine residue (His64) is protonated at N_ϵ and suggest that the A_1 and A_3 peaks of the IR spectrum originate from protein conformations in which the N_ϵ –H bond of His64 points towards the CO ligand. We also find that the CO is substantially stabilized by interaction with the distal histidine residue, in contrast with the common assumption that such stabilization occurs for O_2 only. The strength of the CO \cdots His64 interaction is, nevertheless, smaller than that of the O_2 \cdots His64 interaction by at least 2 kcal mol $^{-1}$.

In summary, our calculations have quantified the interplay between the structure, energy, and dynamics of the heme-active center and its interaction with the protein. This helps us to understand several unclear aspects such as the precise structure of the Fe–ligand bonds, their intrinsic dynamics, the role of the proximal and distal histidines, and the origin of the CO stretch bands of the IR spectrum of MbCO. Of course, many relevant biological processes, e.g. ligand migration into the solvent and the folding of the protein, occur on long timescales and thus cannot currently be treated by use of this approach. This study illustrates, however, how first-principles molecular-dynamics simulations can provide useful hints enabling understanding of the mechanism of short-timescale processes in proteins.

3.6

Acknowledgments

Computing support from the Max-Planck Institute (Garching, Germany) and the CEPBA-IBM Research Institute of Barcelona (Spain) is acknowledged. I thank Jürg Hutter, Pietro Ballone, Enric Canadell, Mauro Boero, and Roger Rousseau for many useful discussions, and especially, Professor Michele Parrinello for his support and continuous interest throughout this work. Financial support from the “Ramón y Cajal” program of the Spanish Ministry of Science and Technology is also acknowledged.

3.7

References

- 1 L. STRYER, *Biochemistry*, 4th edn, Freeman, New York, 1995.
- 2 a) J.T. SAGE, P.M. CHAMPION, *Small Substrate Recognition in Heme Proteins*, in: *Comprehensive Supramolecular Chemistry*, 1996, 5, 171–213; b) B.A. SPRINGER, S. G. SLIGAR, J.S. OLSON, G.N. PHILLIPS, *Chem. Rev.* 1994, 94, 699; c) J.S. OLSON, G.N. PHILLIPS, *J. Biol. Chem.* 1996, 271, 17593; d) NIENHAUS et al. *Nature* 2000, 404, 205.
- 3 a) I. SCHLICHTING, J. BERENDZEN, G.N. PHILLIPS, R.M. SWEET, *Nature* 1994, 371, 808; b) F. YANG, G.N. PHILLIPS, *J. Mol. Biol.* 1996, 256, 762; c) J. KURIYAN, S. WILZ, M. KARPLUS, G. PETSKO, *J. Mol. Biol.* 1986, 192, 133; d) F. YANG, G.N. PHILLIPS, *J. Mol. Biol.* 1996, 256, 762; e) A.N. KACHALOVA, H. POPOV, D. BARTUNIK, *Science* 1999, 284, 473; f) J. VOJTECHOVSKY, K. CHU, J. BERENDZEN, R.M. SWEET, I. SCHLICHTING, *Biophys. J.* 1999, 77, 2153.
- 4 X. CHENG, B.P. SCHOENBORN, *J. Mol. Biol.* 1991, 220, 381.
- 5 a) M. LIM, T.A. JACKSON, P.A. ANFINRUD, *Science* 1995, 269, 962; b) J. T. SAGE, W. JEE. *J. Mol. Biol.* 1997, 274, 21.
- 6 a) A. GHOSH, D.F. BOCIAN, *J. Phys. Chem.* 1996, 100, 6363; b) C. ROVIRA, K. KUNC, J. HUTTER, P. BALLONE, M. PARRINELLO, *J. Phys. Chem. A* 1997, 101, 8914; c) R.H. HAVLIN, N. GODBOUT, R. SALZMANN, M. WOJDELSKI, W. ARNOLD, C.E. SCHULZ, E. OLDFIELD, *J. Am. Chem. Soc.* 1998, 120, 3144; d) T.G. SPIRO, P.M. KOZLOWSKI, *J. Am. Chem. Soc.* 1998, 120, 4524.
- 7 a) C. SLEBODNICK, J.A. IBERS, *J. Biol. Inorg. Chem.* 1997, 2, 521; b) T.G. SPIRO, P.M. KOZLOWSKI, *Acc. Chem. Res.* 2001, 34, 137.
- 8 a) G.B. RAY, X.-Y. LI, J.A. IBERS, J.L. SESSLER, T.G. SPIRO, *J. Am. Chem. Soc.* 1994, 116, 162; b) E. SIGFRIDSON, U. RYDE, *J. Biol. Inorg. Chem.* 1999, 4, 99; c) C. ROVIRA, B. SCHULZE, M. EICHINGER, J.

- EVANSECK, M. PARRINELLO, *Biophys. J.* **2001**, *81*, 435; d) D. A. SCHERLIS, D. A. ESTRIN, *J. Am. Chem. Soc.* **2001**, *123*, 8436.
- 9 a) T. LI, M. L. QUILLIN, G. N. PHILLIPS, J. S. OLSON, *Biochemistry* **1994**, *33*, 1446; b) G. N. PHILLIPS, M. L. TEODORO, T. LI, B. SMITH, J. S. OLSON, *J. Phys. Chem. B* **1999**, *103*, 8817.
- 10 a) W. H. FUCHSMAN, C. A. APPLEBY, *Biochemistry* **1979**, *18*, 1309; b) H. SHIMADA, W. S. CAUGHEY, *J. Biol. Chem.* **1982**, *257*, 11893; c) A. ANSARI, J. BERENDZEN, D. BRAUNSTEIN, B. R. COWEN, H. FRAUENFELDER, M. K. HONG, E. T. IBEN, J. B. JOHNSON, P. ORMOS, T. B. SAUKE, R. SCHOLL, A. SCHULTE, P. J. STEINBACH, J. VITTITOW, R. D. YOUNG, *Biophys. Chem.* **1987**, *26*, 337; d) M. K. HONG, D. BRAUNSTEIN, B. R. COWEN, H. FRAUENFELDER, I. E. T. IBEN, J. R. MOURANT, P. ORMOS, R. SCHOLL, A. SCHULTE, P. J. STEINBACH, A. H. XIE, R. D. YOUNG, *Biophys. J.* **1990**, *58*, 429; e) I. E. T. IBEN, D. BRAUNSTEIN, W. DOSTER, H. FRAUENFELDER, M. K. HONG, J. B. JOHNSON, S. LUCK, P. ORMOS, A. SCHULTE, P. J. STEINBACH, A. H. XIE, R. D. YOUNG, *Phys. Rev. Lett.* **1989**, *62*, 1916; f) D. MORIKIS, P. M. CHAMPION, B. A. SPRINGER, S. G. SLIGAR, *Biochemistry* **1989**, *28*, 4791; g) H. FRAUENFELDER, N. A. ALBERDING, A. ANSARI, D. BRAUNSTEIN, B. R. COWEN, M. K. HONG, I. E. T. IBEN, J. B. JOHNSON, S. LUCK, M. C. MARDEN, J. R. MOURANT, P. ORMOS, L. REINISCH, R. SCHOLL, A. SCHULTE, E. SHYAMSUNDER, L. B. SORENSON, P. J. STEINBACH, A. XIE, R. D. YOUNG, K. T. YUE, *J. Phys. Chem.* **1991**, *94*, 1024; h) L. ZHU, J. T. SAGE, A. A. RIGOS, D. MORIKIS, P. M. CHAMPION, *J. Mol. Biol.* **1992**, *224*, 207; i) J. R. MOURANT, D. BRAUNSTEIN, K. CHU, H. FRAUENFELDER, G. U. NIENHAUS, P. ORMOS, R. D. YOUNG, *Biophys. J.* **1993**, *65*, 1496; j) J. D. MÜLLER, B. H. MCMAHON, E. Y. T. CHIEN, S. G. SLIGAR, G. U. NIENHAUS, *Biophys. J.* **1999**, *77*, 1036.
- 11 a) J. B. JOHNSON, D. C. LAMB, H. FRAUENFELDER, J. D. MÜLLER, B. MCMAHON, G. U. NIENHAUS, R. D. YOUNG, *Biophys. J.* **1996**, *71*, 1563; b) G. U. NIENHAUS, R. D. YOUNG, *Protein Dynamics in: Encyclopedia of Applied Physics*, Vol. 15. G. L. TRIGG (ed.), VCH, New York, **1996**, pp. 163–184.
- 12 D. P. BRAUNSTEIN, K. CHU, K. D. EGEGERG, H. FRAUENFELDER, J. R. MOURANT, G. U. NIENHAUS, P. ORMOS, S. G. SLIGAR, B. A. SPRINGER, R. D. YOUNG, *Biophys. J.* **1993**, *65*, 2447.
- 13 a) P. ORMOS, D. BRAUNSTEIN, H. FRAUENFELDER, M. K. HONG, S.-L. LIN, T. B. SAUKE, R. D. YOUNG, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 8492; b) M. UNNO, J. F. CHRISTIAN, J. S. OLSON, J. T. SAGE, P. M. CHAMPION, *J. Am. Chem. Soc.* **1998**, *120*, 2670.
- 14 S. E. V. PHILLIPS, B. P. SCHOENBORN, *Nature* **1981**, *292*, 81.
- 15 S. BHATTACHARYA, S. F. SUKITS, K. L. MACLAUGHLIN, J. T. J. LECOMTE, *Biophys. J.* **1997**, *73*, 3230.
- 16 S. E. V. PHILLIPS, *J. Mol. Biol.* **1980**, *142*, 531.
- 17 a) B. SHAANAN, *Nature* **1982**, *296*, 683; b) B. SHAANAN, *J. Mol. Biol.* **1983**, *171*, 31.

- 18 P. COLLMAN, R. R. GAGNE, T. R. HALBERT, J. C. MARCHON, C. A. REED, *J. Am. Chem. Soc.* **1973**, 95, 7868; b) J. P. COLLMAN, J. L. HOARD, N. KIM, G. LANG, C. A. REED, *J. Am. Chem. Soc.* **1975**, 97, 2676; c) J. P. COLLMAN, R. R. GAGNE, C. A. REED, T. R. HALBERT, G. LANG, W. T. ROBINSON, *J. Am. Chem. Soc.* **1975**, 97, 1427; d) G. B. JAMESON, F. MOLINARO, J. A. IBERS, J. P. COLLMAN, J. I. BRAUMAN, E. ROSE, K. S. SUSLICK, *J. Am. Chem. Soc.* **1980**, 102, 3224; e) S.-M. PENG, J. A. IBERS, *J. Am. Chem. Soc.* **1976**, 98, 8032; f) W. R. SCHEIDT, M. E. FRISSE, *J. Am. Chem. Soc.* **1975**, 97, 17; g) W. R. SCHEIDT, P. L. PICIUOLO, *J. Am. Chem. Soc.* **1976**, 98, 1913; h) G. B. JAMESON, G. A. RODLEY, W. T. ROBINSON, R. R. GAGNE, C. A. REED, J. A. COLLMAN, *Inorg. Chem.* **1978**, 17, 850.
- 19 J. H. BOWEN, N. V. SHOKHIREV, A. M. RAITSIMRING, D. H. BUTTLAIRE, F. A. WALKER, *J. Am. Chem. Soc.* **1997**, 101, 8683.
- 20 a) K. SPARTALIAN, G. LANG, J. P. COLLMAN, R. R. GAGNE, C. A. REED, *J. Chem. Phys.* **1975**, 63, 5375; b) J. MISPELTER, M. MOMENTEAU, D. LAVALETTE, J.-M. LHOSTE, *J. Am. Chem. Soc.* **1983**, 105, 5165; c) E. OLDFIELD, H. C. LEE, C. CORETPOULOS, F. ADEBODUM, K. D. PARK, S. YANG, J. CHUNG, B. PHILLIPS, *J. Am. Chem. Soc.* **1991**, 113, 8680.
- 21 E. A. BRUCKER, J. S. OLSON, M. IKEDA-SAITO, G. N. PHILLIPS, *Prot. Str. Funct. Gen.* **1998**, 30, 352.
- 22 a) T. G. TRAYLOR, V. S. SHARMA, *Biochemistry* **1992**, 31, 2847; b) E. J. ROSE, B. M. HOFFMAN, *J. Am. Chem. Soc.* **1983**, 105, 2866.
- 23 M. A. MARLETTA, *Biochemistry* **1998**, 37, 12458.
- 24 M. F. PERUTZ, S. S. HASNAIN, P. J. DUKE, J. L. SESSLER, J. E. HAHN, *Nature* **1982**, 295, 535.
- 25 M. MOMENTEAU, C. A. REED, *Chem. Rev.* **1994**, 94, 659.
- 26 J. P. COLLMAN, *Inorg. Chem.* **1997**, 36, 5145.
- 27 T. G. TRAYLOR, *Acc. Chem. Res.* **1981**, 102.
- 28 a) K. S. SUSLICK, M. M. FOX, *J. Am. Chem. Soc.* **1983**, 105, 3507; b) J. P. COLLMAN, J. I. BRAUMAN, B. L. IVERSON, J. L. SESSLER, R. M. MORRIS, Q. H. GIBSON, *J. Am. Chem. Soc.* **1983**, 105, 3052; c) D. LEXA, M. MOMENTEAU, J. M. SAVEANT, F. XU, *J. Am. Chem. Soc.* **1986**, 108, 6937; d) J. P. COLLMAN, L. FU, A. ZINGG, F. DIEDERICH, *Chem. Commun.* **1997**, 2, 193; e) J. P. COLLMAN, X. ZHANG, K. WONG, J. I. BRAUMAN, *J. Am. Chem. Soc.* **1994**, 116, 6245.
- 29 a) G. B. JAMESON, R. S. DRAGO, *J. Am. Chem. Soc.* **1985**, 107, 3017; b) M. MOMENTEAU, B. LOOCK, D. LAVALETTE, C. TETREAU, J. MISPELTER, *J. Chem. Soc., Chem. Commun.* **1983**, 962; c) G. E. WUENSCHELL, C. TETREAU, D. LAVALETTE, C. A. REED, *J. Am. Chem. Soc.* **1992**, 114, 3346; d) I. P. GEROTHANASSIS, M. MOMENTEAU, B. LOOCK, *J. Am. Chem. Soc.* **1989**, 111, 7006.
- 30 a) J. P. COLLMAN, P. C. HERRMANN, L. FU, T. A. EBERSPACHER, M. EUBANKS, B. BOITREL, P. HAYOZ, X. ZHANG, J. I. BRAUMAN, V. W. DAY, *J. Am. Chem. Soc.* **1997**, 119, 3481; b) A. DESBOIS, M. MOMENTEAU, M. LUTZ, *Inorg. Chem.* **1989**, 28, 825.
- 31 J. A. IBERS, *Comments Inorg. Chem.* **1983**, 2, 97.

- 32 a) P. JEWSEBURY, T. KITAGAWA, *Biophys. J.* **1994**, *67*, 2236; b) J. STRAUB, M. KARPLUS, *Chem. Phys.* **1991**, *158*, 221.
- 33 B. SCHULZE, J. D. EVANSECK, *J. Am. Chem. Soc.* **1999**, *121*, 6444.
- 34 See, for instance: a) A. DEDIEU, M.-M. ROHMER, M. BENARD, A. VEILLARD, *J. Am. Chem. Soc.* **1976**, *98*, 3717; b) R. F. KIRCHNER, G. H. LOEW, *J. Am. Chem. Soc.* **1977**, *99*, 4639; c) D. A. CASE, B. H. HUYNH, M. KARPLUS, *J. Am. Chem. Soc.* **1979**, *101*, 4433; d) Z. S. HERMAN, G. H. LOEW, *J. Am. Chem. Soc.* **1980**, *102*, 1815; e) W. D. EDWARDS, B. WEINER, M. C. ZERNER, *J. Am. Chem. Soc.* **1986**, *108*, 2196; f) A. WALEH, N. HO, L. CHANTRANUPONG, G. H. LOEW, *J. Am. Chem. Soc.* **1989**, *111*, 2767; g) J. E. NEWTON, M. B. HALL, *Inorg. Chem.* **1984**, *23*, 4627; h) H. NAKATSUJI, J. HASEGAWA, H. UEDA, M. H. HADA, *Chem. Phys. Lett.* **1996**, *250*, 379; i) P. JEWSEBURY, S. YAMAMOTO, T. MINATO, M. SAITO, T. KITAGAWA, *J. Phys. Chem.* **1995**, *99*, 12677.
- 35 a) C. ROVIRA, M. PARRINELLO, *Chem. Eur. J.* **1999**, *5*, 250; b) C. ROVIRA, P. CARLONI, M. PARRINELLO, *J. Phys. Chem. B* **1999**, *103*, 7031; c) M. KAUPP, C. ROVIRA, M. PARRINELLO, *J. Phys. Chem. B* **2000**, *104*, 5200; d) C. ROVIRA, M. PARRINELLO, *Int. J. Quant. Chem.* **2000**, *80*, 1172; e) C. ROVIRA, M. PARRINELLO, *Biophys. J.* **2000**, *78*, 93; f) C. ROVIRA, K. KUNC, M. PARRINELLO, *Inorg. Chem.* **2001**, *40*, 11; g) D. HARRIS, G. H. LOEW, A. KOMORNICKI, *J. Phys. Chem. A* **1997**, *101*, 3959.
- 36 See, for instance: a) U. RÖTHLISBERGER, P. CARLONI, K. DOCCIO, M. PARRINELLO, *J. Biol. Inorg. Chem.* **2000**, *5*, 236; b) C. MOLteni, I. FRANK, M. PARRINELLO, *J. Am. Chem. Soc.* **1999**, *121*, 12177; c) F. ALBER, O. KUONEN, L. SCAPOZZA, G. FOLKERS, P. CARLONI, *Proteins: Struct. Func. Gen.* **1998**, *31*, 453; d) P. CARLONI, M. SPRIK, W. ANDREONI, *J. Phys. Chem. B* **2000**, *104*, 823; e) S. PIANA, D. SEBASTIANI, P. CARLONI, M. PARRINELLO, *J. Am. Chem. Soc.* **2001**, *123*, 8730; f) D. E. SAGNELLA, K. LAASONEN, M. L. KLEIN, *Biophys. J.* **1996**, *71*, 1172; g) J. HUTTER, P. CARLONI, M. PARRINELLO, *J. Am. Chem. Soc.* **1996**, *118*, 7847.
- 37 P. HOHENBERG, W. KOHN, *Phys. Rev. B* **1964**, *136*, 864.
- 38 W. KOHN, L. J. SHAM, *Phys. Rev. A* **1965**, *140*, 1133.
- 39 A. D. BECKE, *J. Chem. Phys.* **1986**, *84*, 4524; b) J. P. PERDEW, *Phys. Rev. B* **1986**, *33*, 8822.
- 40 N. TROULLIER, J. L. MARTINS, *Phys. Rev. B* **1991**, *43*, 1993.
- 41 S. G. LOUIE, S. FROYEN, M. L. COHEN, *Phys. Rev. B* **1982**, *26*, 1738.
- 42 a) R. CAR, M. PARRINELLO, *Phys. Rev. Lett.* **1985**, *55*, 2471; b) G. GALLI, M. PARRINELLO, *Computer Simulation in Materials Science*. V. PONTIKIS, M. MEYER, (eds), Kluwer, Dordrecht, **1991**, and references cited therein.
- 43 CPMD 3.0h program, written by J. HUTTER (Max-Planck-Institut für Festkörperforschung, Stuttgart, **1998**).
- 44 M. EICHINGER, P. TAVAN, J. HUTTER, M. PARRINELLO, *J. Chem. Phys.* **1999**, *110*, 10452.

- 45 M. EICHINGER, H. GRUBMÜLLER, H. HELLER, P. TAVAN, *J. Comput Chem.* **1997**, *118*, 1729.
- 46 B. R. BROOKS et al., *J. Comput. Chem.* **1983**, *4*, 187.
- 47 T. A. ALBRIGHT, J. K. BURDETT, M.-H. WHANGBO, in: *Orbital Interactions in Chemistry*, Wiley, New York, **1985**.
- 48 O. N. CHEN, S. GROH, A. LIECHTY, D. P. RIDGE, *J. Am. Chem. Soc.* **1999**, *121*, 11910.
- 49 a) The preferred orientation of the 2-meIm ligand is a compromise between the maximum N_p -Fe{d} bonding, achieved with an eclipsed conformation ($\Phi=0^\circ$) and the minimum steric interaction with the Fe- N_p bonds, which is achieved in a bisecting conformation ($\Phi=45^\circ$); b) W. R. SCHEIDT, D. M. CHIPMAN, *J. Am. Chem. Soc.* **1986**, *108*, 1163.
- 50 R. HOFFMANN, M.-L. CHEN, D. THORN, *Inorg. Chem.* **1977**, *16*, 503.
- 51 For a discussion of electronic rearrangements when binding O_2 to the iron-porphyrin see Refs. [6b, 50]. The same basic considerations are valid here, with only slight differences in the d-Fe character of the orbitals, because of the imidazole axial ligand. The notation d_{π_1} and d_{π_2} refer to the $d_{xz}+d_{yz}$ and $d_{xz}-d_{yz}$ combinations, respectively. $\pi_{g,s}^*$ and $\pi_{g,a}^*$ refer to the two antibonding oxygen orbitals, symmetric and antisymmetric, respectively, relative to the (x+y, z) plane.
- 52 The electron distribution can be shown schematically, in a more general case, as:
- $$d_{xy}(\uparrow\downarrow)(\mu_1 d_{\pi_1} + \mu_2 \pi_{g,s}^*)(\uparrow\downarrow)$$
- $$(\delta_1 \pi_{g,s}^* + \delta_2 d_{z^2} + \delta_3 n_{O_2})(\uparrow\downarrow)$$
- $$(\lambda_1 d_{\pi_2} + \lambda_2 \pi_{g,a}^*)(\uparrow)$$
- $$(\lambda_2 d_{\pi_2} + \lambda_1 \pi_{g,a}^*)(\downarrow)$$
- For the bent structure, our calculations give μ_1 1, μ_2 0; δ_1 1, δ_2 0, δ_3 0; λ_1 1, λ_2 0.
- 53 Note that the orbitals d_{xz} , d_{yz} , $\pi_{g,1}^*$ and $\pi_{g,2}^*$ of the linear conformation may turn into d_{p1} , d_{p2} , $\pi_{g,s}^*$ and $\pi_{g,a}^*$ in the bent conformation.
- 54 S. HUANG, J. HUANG, A. P. KLOEK, D. E. GOLDBERG, J. M. FRIEDMAN, *J. Biol. Chem.* **1996**, *271*, 958.

4

Density-functional Theory in Drug Design – the Chemistry of the Anti-tumor Drug Cisplatin and Photoactive Psoralen Compounds

JOHAN RABER, JORGE LLANO, and LEIF A. ERIKSSON

4.1

Introduction

When addressing problems in computational chemistry, the choice of computational scheme depends on the applicability of the method (i.e. the types of atoms and/or molecules, and the type of property, that can be treated satisfactorily) and the size of the system to be investigated. In biochemical applications the method of choice – if we are interested in the dynamics and effects of temperature on an entire protein with, say, 10,000 atoms – will be to run a classical molecular dynamics (MD) simulation. The key problem then becomes that of choosing a relevant force field in which the different atomic interactions are described. If, on the other hand, we are interested in electronic and/or spectroscopic properties or explicit bond breaking and bond formation in an enzymatic active site, we must resort to a quantum chemical methodology in which electrons are treated explicitly. These phenomena are usually highly localized, and thus only involve a small number of chemical groups compared with the complete macromolecule.

Including electronic interactions increases the complexity of the problem enormously, such that the number of atoms to be included typically ranges from 10–100, depending on the methodology chosen. The more accurate the treatment, the more computationally costly the calculation will be. This implies that quantum chemical calculations are generally performed on small *model systems*, usually in a vacuum or surrounded by a dielectric continuum that mimics bulk solution effects. Based on studies of, e.g., several alternative reaction mechanisms or properties of different

fragmentation products, and comparison with experimental observations, the models that best match known data can be selected, and provide further insight into the problem investigated. The roles of explicit orbital interactions, binding modes, charge distributions, intermediate conformations, effects of different substituents on reaction barriers and pathways, etc., can thereby be determined – information that normally cannot be extracted from experimental measurements.

Of the many quantum chemical approaches available, density-functional theory (DFT) has over the past decade become a key method, with applications ranging from interstellar space, to the atmosphere, the biosphere and the solid state. The strength of the method is that whereas conventional *ab initio* theory includes electron correlation by use of a perturbation series expansion, or increasing orders of excited state configurations added to “zero-order” Hartree-Fock solutions, DFT methods inherently contain a large fraction of the electron correlation already from the start, via the so-called *exchange-correlation functional*.

In this review, we present an introduction to the theory, and exemplify the wide range of problems that can be addressed with some illustrative results from our work in the field of *ab initio* drug design. The problems addressed are those of activation and DNA binding of the antitumor drug *cis*-platin ($\text{PtCl}_2(\text{NH}_3)_2$), and basic spectrometric data from a family of drugs known as psoralens.

4.2 Density-functional Theory

We begin our review with a brief description of the historic background and basic equations of DFT. For the interested reader there are many excellent textbooks on the subject [1–3] and review books with compilations of results from studies in, e.g., chemistry, biochemistry, or materials science [4–7].

The basic idea in DFT is to express the total energy as a functional of the electron density, i.e., $E_{\text{tot}} = E[\rho(\vec{r})]$. We are thereby able to (formally) reduce the many-body electronic problem to a dependence on three coordinates only (the value of the density at

point \vec{r}), as opposed to wave-function theory, in which the energy depends on the explicit coordinates of all electrons (i.e. $3N$ coordinates, where N is the number of electrons).

The development of the method started in the mid 1920's with the work of Thomas and Fermi [8, 9]. The aim was to formulate an electronic structure theory for the solid state, based on the properties of a homogeneous electron gas, to which we introduce a set of external potentials (i.e. the atomic nuclei). The original formulation, with later additions by Dirac [10] and Slater [11], was, however, inadequate for accurate description of atomic and molecular properties, and it was not until the ground-breaking work of Kohn and coworkers in the mid 1960's that the theory was put in a form more suited to computational chemistry [12, 13].

Starting from the relatively crude, or oversimplified, Thomas-Fermi-Dirac expression, Hohenberg and Kohn were able to state two very important theorems that came to form the basis on which to develop the theory further [12]. The first HK theorem states that there is a one-to-one relationship between the density and the external potential. That is, each external potential will modify the density in its own unique way. In addition, we know that the density determines the number of electrons ($N = \int \rho(\vec{r}) d\vec{r}$) and the occupied wavefunctions ($\rho(\vec{r}) = \sum_i^{\text{occ}} |\psi_i(\vec{r})|^2$). This means that if we know the exact density (for a non-degenerate ground state), we can determine the potential, the number of particles, the ground state wave function, and thus the total energy and all the properties of the system. The second theorem, the proof of which is straightforward, shows that there is a variational principle for the density, and that any trial density will generate an energy that lies above that corresponding to the exact density.

4.2.1

Basic Equations

Starting from a homogeneous electron gas and the above theorems, Kohn and Sham in 1965 proposed a solution to the problem of electronic interaction in many-electron systems based on defining and iteratively solving a set of coupled one-electron equations [13]. With this development DFT was put on similar

footing as ordinary HF theory, thus enabling the same type of computational schemes to be used, and also enabling comparative studies and stricter development of methodology.

We begin by separating the kinetic energy functional into a functional of noninteracting electrons, $T_s[\rho(\vec{r})]$, and a remainder that is included in the exchange-correlation functional $E_{xc}[\rho(\vec{r})]$. Next, we decompose the density into a set of single-particle wave functions, the so-called Kohn-Sham orbitals, $\rho(\vec{r}) = \sum_i^n \sum_s |\psi_i(\vec{r}, s)|^2$. The occupancy of the orbitals is chosen to be one for the n lowest and zero for the rest. The actual choice of orbitals is, however, arbitrary, if they constitute an orthonormal set and yield the correct density. These orbitals exactly describe a system of noninteracting electrons, enabling us to treat the remaining, lesser part of the kinetic energy indirectly. The corresponding expression for the kinetic energy is thus $T_s[\rho(\vec{r})] = \sum_i^n \langle \psi_i | -\frac{1}{2} \nabla^2 | \psi_i \rangle$, where $T_s[\rho(\vec{r})] < T_{\text{exact}}[\rho(\vec{r})]$. The exchange-correlation energy functional thereby becomes a combination of the interaction part of the kinetic energy and all electron–electron interaction that is not covered by the classical Coulomb term:

$$E_{xc}[\rho(\vec{r})] = (T_{\text{exact}}[\rho(\vec{r})] - T_s[\rho(\vec{r})]) + (V_{ee}[\rho(\vec{r})] - J[\rho(\vec{r})])$$

The total energy can thus be written as:

$$\begin{aligned} E[\rho(\vec{r})] &= T_s[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{xc}[\rho(\vec{r})] + \int v(\vec{r})\rho(\vec{r})d\vec{r} = \\ &= \sum_i^n \int \psi_i^*(\vec{r})(-\frac{1}{2}\nabla^2)\psi_i(\vec{r}) + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 \\ &\quad + E_{xc}[\rho(\vec{r})] + \int v(\vec{r})\rho(\vec{r})d(\vec{r}) \end{aligned}$$

Next we impose the orthonormality constraint on the wave functions by means of Lagrange multipliers, ε_{ij} , and obtain the n one-electron Euler-Lagrange equations:

$$\hat{H}_i \psi_i = (-\frac{1}{2} \nabla^2 + v_{\text{eff}}) \psi_i = \sum_j^n \varepsilon_{ij} \psi_j$$

where the effective potential, v_{eff} , is obtained from the functional derivatives of all but the kinetic energy terms:

$$v_{\text{eff}}(\vec{r}) = v(\vec{r}) + \frac{\delta J[\rho(\vec{r})]}{\delta \rho(\vec{r})} + \frac{\delta E_{\text{xc}}[\rho(\vec{r})]}{\delta \rho(\vec{r})} = v(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + v_{\text{xc}}(\vec{r})$$

Diagonalization of the matrix formed by the multipliers $\{\varepsilon_{ij}\}$ yields the Kohn-Sham orbitals and their eigen-energies:

$$\hat{H}_i \psi_i = \left(-\frac{1}{2}\nabla^2 + v_{\text{eff}}\right)\psi_i = \varepsilon_i \psi_i$$

These constitute the DFT equivalence of the HF equations and, like those, must be solved iteratively in a self-consistent (SCF) procedure. The total energy can then be obtained via the density by use of the equations above, or – again by analogy with the HF equations – by use of:

$$E_{\text{tot}} = \sum_i^n \varepsilon_i - \frac{1}{2} \int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r} d\vec{r}' + E_{\text{xc}}[\rho(\vec{r})] - \int v_{\text{xc}}(\vec{r})\rho(\vec{r})d\vec{r}$$

The advantage over the HF scheme is that whereas in conventional *ab initio* theory we must resort to costly perturbation theory or configuration interaction expansions, in DFT electron correlation is already included explicitly in the exchange-correlation functional. The key problem is instead to find an appropriate expression for E_{xc} . As stated above, when we have the correct functional we should be able to extract the exact energy, the exact ground state density, and all properties for our system.

4.2.2

Gradient Corrections and Hybrid Functionals

The initial implementation of DFT employed the so-called local density approximation, LDA (or, if we have separate α and β spin, the local spin density approximation, LSDA). The basic assumption is that the density varies only slowly with distance – which it is locally constant. Another way of visualizing the concept of LDA is that we start with a homogeneous electron gas and subsequently localize the density around each external potential – each nucleus in a molecule or a solid. That the density is locally constant is indeed true for the intermediate densities, but not necessarily so in the high- and low-density regions. To correct for this, it was rec-

ognized in the early 1980's that not only the density but also the gradient of the density should be included in the functional expressions, i.e. $E_{xc}[\rho(\vec{r})] = \int \rho(\vec{r}) \varepsilon_{xc}(\rho^\alpha, \rho^\beta, \nabla\rho^\alpha, \nabla\rho^\beta) d\vec{r}$. This gave rise to several different gradient expansion approximations (GEA) later refined in the form of generalized gradient approximations (GGA). With the advent of the gradient-corrected density functionals, DFT started evolving to a main-stream method in all areas of computational chemistry. It is currently the most widely applied quantum chemical approach, and has been used for almost a decade in theoretical studies of systems of biochemical interest.

In this section we will describe only the most commonly applied gradient-corrected exchange and correlation functionals and hybrid schemes in DFT; many more are available. The LDA expression by Dirac and Slater ("S") for exchange ($E_x[\rho(\vec{r})] = \frac{3}{4} (\frac{3}{\pi})^{1/3} \int \rho(\vec{r})^{4/3} d\vec{r}$) [10, 11], and Ceperly and Alder's Monte Carlo data for the local correlation energy of a homogeneous electron gas [14] as parameterized by Vosko, Wilk, and Nusair [15] ("VWN") form the basis of essentially all subsequent gradient-corrected functionals. LDA (or LSDA) and "S-VWN" are hence analogous acronyms for performing a local density calculation using the above functional forms.

Most gradient-corrected functionals thus use the LDA functionals as a basis, and then add corrections to these. The first generation of gradient corrections, developed during the mid 1980's, were derived with special attention given to the properties of the exchange-correlation hole – the "vacuum" surrounding each electron. To make the expressions more readily computable, this was often followed by polynomial fitting of the analytical expressions and/or parameterization to properties of noble gases.

In 1988, Becke proposed a gradient-corrected scheme for the exchange functional ("B88"), ensuring that this should have the correct asymptotic limit ($1/r$) as $r \rightarrow \infty$ [16]:

$$E_x^{B88}[\rho(\vec{r})] = E_x^{LDA}[\rho(\vec{r})] - \beta \sum_{\sigma} \rho_{\sigma}^{4/3}(\vec{r}) \frac{x_{\sigma}^2}{1 + 6\beta x_{\sigma} \sinh^{-1} x_{\sigma}} d\vec{r}$$

where σ denotes the spin, $x_{\sigma} = |\nabla\rho|/\rho_{\sigma}^{4/3}$, and β is a semiempirical term (0.0042 a.u.) determined from least square fitting to exact HF data for the noble gases He–Rn. Other popular exchange functionals are those by Perdew and Wang from 1986 [17] and from 1991 [18] (PW86 and PW91, respectively). The PW91 ex-

change functional was derived without use of parameters, starting from the B88 expressions. In 1991 Perdew and Wang also derived a correlation functional based on corrections (scaled density gradients and local screening vectors) to the VWN expression, without empirical parameters but including parameters from numerical fitting to the analytical expressions.

An alternative approach was offered by Lee, Yang, and Parr [19], who derived a gradient-corrected correlation functional (“LYP”) from the second-order density matrix in HF theory. Together with PW91, this functional is currently the most widely used correlation functional for molecular calculations.

It was, however, not until 1993, when Becke’s hybrid functionals appeared [20, 21], that the real breakthrough in molecular calculations occurred. The idea behind these is to combine functionals for noninteracting electrons with those in which the electron–electron interaction is fully switched on. The mixing between the two “extremes” is generally obtained from least squares fitting to a large set of empirical data. The most popular, and accurate, form was launched under the acronym “B3LYP” [21, 22], and is a mixture of exact exchange, and the S-VWN, B88, and LYP functionals:

$$E_{xc}^{\text{B3LYP}} = E_{xc}^{\text{SVWN}} + 0.20(E_x^{\text{KS}} - E_x^{\text{S}}) + 0.72E_x^{\text{B88}} + 0.81E_c^{\text{LYP}}$$

For the G2 set of compounds (a standardized test set of small molecules) the mean error to the atomization energy is approximately 2.5 kcal mol⁻¹ at the B3LYP level, compared with 78 kcal mol⁻¹ for HF theory, and in the range of 1 kcal mol⁻¹ for the most accurate correlated *ab initio* methods. For most cases in which a moderately sized systems (10–50 atoms) is to be investigated, the B3LYP functional is currently the method of choice.

Subsequent to these developments much work has been devoted to improving the older functionals by means of some form of parameterization and fitting, or to find new and better hybrid combinations. Among this second generation functionals we find PBE [23], PBE0 [24], HCTH [25], and similar. The focus is this time more on the *chemistry* (energies, bond strengths, etc.) rather than on the properties of the exchange–correlation hole. Another line of improvement is that of the so-called “meta-GGAs”, in which one also includes the kinetic energy gradient (PKZB, [26]).

That is, we now aim to describe in a more appropriate way the interaction part of the kinetic energy that is introduced to the exchange-correlation functional in the Kohn-Sham scheme. Including the kinetic energy corrections increases the computational requirements substantially, but the accuracy is also much improved compared with conventional gradient-corrected functionals.

4.2.3

Time-dependent Density-functional Response Theory (TD-DFRT)

TD-DFRT (usually known as TD-DFT) is applied to the calculation of dynamic polarizabilities, hyperpolarizabilities, and electronic excitation spectra [27]. An electronic absorption spectrum is a record of spectral intensity as a function of the frequency (ν) of the radiation absorbed by the molecule. The frequency assigned to a band is related to the electron excitation energy according to the Bohr frequency condition ($\Delta E = h\nu$), and the intensity of the band is directly proportional to the oscillator strength, f . The permanent molecular dipole moment ($\bar{\mu}_0$) and the dipole-polarizability tensor (a) at zero field describe the response of the molecular system to the external electric perturbation generated by neighboring molecules or an external apparatus. Accordingly, the interaction of a molecule with light can be modeled as the interaction with a time-dependent, sinusoidally-varying electric field of angular frequency ω . It can be shown that the vertical excitation energies ($\omega_I = E_I - E_0$) from the molecular electronic ground state to the excited state I ($I \leftarrow 0$) can be obtained from the poles of the mean dynamic polarizability $\bar{a}(\omega)$:

$$\bar{a}(\omega) = \frac{1}{3} \text{tra}(\omega) = \sum_I \frac{f_I}{\omega_I^2 - \omega^2}$$

Whereas the classic Kohn-Sham (KS) formulation of DFT is restricted to the time-independent case, the formalism of TD-DFT generalizes KS theory to include the case of a time-dependent, local external potential $w(t)$ [27].

The time-dependent KS equations:

$$F^{\text{KS}}(\vec{r}, t)\psi_i(\vec{r}, t) = i\frac{\partial}{\partial t}\psi_i(\vec{r}, t)$$

where

$$F^{\text{KS}}(\vec{r}, t) = -\frac{1}{2}\nabla^2 + v_{\text{eff}}(\vec{r}, t)$$

are derived under the assumption that there is an effective potential $v_{\text{eff}}(\vec{r}, t)$, for an independent (noninteracting) particle system, the orbitals $\psi_i(\vec{r}, t)$ of which yield the same charge density $\rho(\vec{r}, t)$ as the interacting system. The potential $v_{\text{eff}}(\vec{r}, t)$ can be split up into the contributions of the local external potential $w(t)$ and the SCF potential. The latter consists of the Coulombic electron–electron potential $v_{\text{C}}(\vec{r}, t)$ and the exchange–correlation potential $v_{\text{xc}}(\vec{r}, t)$:

$$v_{\text{eff}}(\vec{r}, t) = w(t) + v_{\text{SCF}}(\vec{r}, t) = w(t) + \{v_{\text{C}}(\vec{r}, t) + v_{\text{xc}}(\vec{r}, t)\}$$

To evaluate $v_{\text{eff}}(\vec{r}, t)$ at a particular time τ , the adiabatic approximation is introduced. This approximation is local in time, and thus the Coulomb and exchange–correlation potentials are just those of time-independent DFT, evaluated using the density determined at time τ .

Time-dependent response theory concerns the response of a system initially in a stationary state, generally taken to be the ground state, to a perturbation turned on slowly, beginning some time in the distant past. The assumption that the perturbation is turned on slowly, i.e. the adiabatic approximation, enables us to consider the perturbation to be of first order. In TD-DFT the density response $\delta\rho$, i.e. the density change which results from the perturbation δv_{eff} , enables direct determination of the excitation energies as the poles of the response function $\delta\mathbf{P}$ (the linear response of the KS density matrix in the basis of the unperturbed molecular orbitals) without formally having to calculate $\bar{a}(\omega)$.

The quality of the TD-DFT results is determined by the quality of the KS molecular orbitals and the orbital energies for the occupied and virtual states. These in turn depend on the exchange–correlation potential. In particular, excitations to Rydberg and valence states are sensitive to the behavior of the exchange–correlation potential in the asymptotic region. If the exchange–correla-

tion potential is corrected for the asymptotic region, then high quality orbitals and energies for the excited states are obtained. The typical error of TD-DFT in estimating excitation energies is in the range 0.2–0.3 eV [28].

4.2.4

Applicability and Applications

Although there is no strict relationship between the basis sets developed for, and used in, conventional *ab initio* calculations and those applicable in DFT, the basis sets employed in molecular DFT calculations are usually the same or highly similar to those. For most practical purposes, a standard valence double-zeta plus polarization basis set (e.g. the Pople basis set 6-31G(d,p) [29] and similar) provides sufficiently accurate geometries and energetics when employed in combination with one of the more accurate functionals (B3LYP, PBE0, PW91). A somewhat sweeping statement is that the accuracy usually lies mid-way between that of MP2 and that of the CCSD(T) or G2 conventional wave-function methods.

Because of favorable scaling of the DFT equations with increasing number of basis functions, and because a variety of computational “short-cuts” make it possible to avoid the worst bottlenecks, DFT calculations are applicable to considerably larger systems than correlated *ab initio* methods in general. In addition, the cost of increasing the basis set is not too high, meaning that it is relatively straightforward to perform single point energy or property calculations with a significantly larger basis set (e.g. including multiple polarization or diffuse functions) when a stationary point has been optimized. Calculations with up to 50 atoms are currently more or less routine within DFT, such that we can start to construct increasingly realistic models, for instance by including most of the chemically important amino acids at the active site of an enzyme. The method has been applied to a wide variety of problems, e.g. metallo-enzymes, DNA damage and repair, drug design, or specific magnetic or spectroscopic properties of biochemically relevant molecules. For more details of the systems so far explored in biochemistry we may refer the reader to two recent compilation volumes [6, 7].

Of course, we always aim to investigate even more realistic systems and find better descriptions of a variety of properties. One approach which enables explicit inclusion of a larger fraction of the surroundings is the so-called QM/MM methodology [30]. In this, we mix an accurately described quantum chemical “core” of the system (e.g. the active site plus substrate) with a classical molecular mechanics embedding (the rest of the enzyme plus part of the solvation shell). This methodology, albeit very promising, has thus far mainly been applied using semiempirical approaches (AM1 or PM3) for the QM part, and the real breakthrough in terms of DFT/MM applications is still to come.

Another approach is that of including dynamics in the calculations. A dynamical formalism of DFT was first developed by Car and Parrinello [31], and has been employed in a wide range of areas, e.g. solvation problems, reactions on surfaces, solid-state interactions, and a variety of biochemical applications. In CP-MD one normally uses a plane wave basis to reduce the computational requirements and enable easy implementation of periodic boundary conditions. Nonetheless, CP-MD simulations are rather costly, and are normally not applied to systems larger than, say, 1–200 atoms, and over relatively short time frames.

We should conclude this section with a few words about the drawbacks of DFT. To begin with, unlike conventional *ab initio* methods, we cannot in DFT systematically include an increasing amount of the electron correlation. That is, we have (as yet) no means of systematically improving an exchange-correlation functional to become increasingly “correct”. The choice of functional must, therefore, inherently involve much pragmatism. Secondly, DFT is a single-reference method and, like all single-reference methods (including most HF- based methods), it fails to describe accurately problems that require a multi-state approach. In addition there are other, more formalistic, aspects referred to as N- and ν -representability, self-interaction corrections, and similar; these however, go beyond the scope of this review.

4.3

Modes of Action of Anti-tumor Drug Cisplatin

Cisplatin, or *cis*-diaminedichloroplatinum(II) (*cis*-DDP), is a potent anti-cancer drug especially effective against tumors in the sex glands, head, and neck [32]. The most likely target of the drug, correlating with anti-tumor activity, has in several investigations been shown to be cellular DNA [33]; it distorts the tertiary structure of DNA and thereby inhibits the replication and transcription machinery of the cell [34–36]. Recently the persistent nature of the inhibitory effect has been shown, in all likelihood, to be a masking of the cisplatin-induced damage by means of indigenous proteins of the cell nucleus [37]. The drug was discovered by Rosenberg et al. in 1965 [38] and is, despite its simple structure, one of the most potent anti-cancer drugs known. Even though it has been used for over thirty years, very little is known about the reasons for its efficacy, and virtually none of the modifications of cisplatin has improved its performance against cancer [39]. The structural isomer *trans*-platin also attacks DNA, bonding to the same bases as cisplatin, but has no clinical activity against cancer cells [40].

Several adducts can be formed by binding of cisplatin to DNA [41]; the most important are the intrastrand adjacent 5'-GG adduct, intrastrand adjacent 5'-AG, and nonadjacent intrastrand didentate adduct GXG (where X=any base), at 65%, 25% and 6%, respectively (Fig. 4.1). The remaining part consists of cisplatin monofunctionally bound to G, and interstrand bifunctional G-G adducts at ~3%. The adducts are, according to experimental evidence, formed exclusively at the N7 position of the purine bases A and G exposed in the major groove of the DNA helix; the other possible sites of complexation are either exposed in the too narrow minor groove or involved in the base-pairing of the two strands.

Another important feature is the twist of the DNA helix. DNA in its native B-form twists around its axis 360° in ten base pairs (~34 Å in length), with the result that bases with the same neighbors have different local chemical environments. For instance, in the sequence 5'-AGGA-3', the two G's will experience different chemical environments even though they have the same neighboring bases. As a consequence the AG adduct is di-

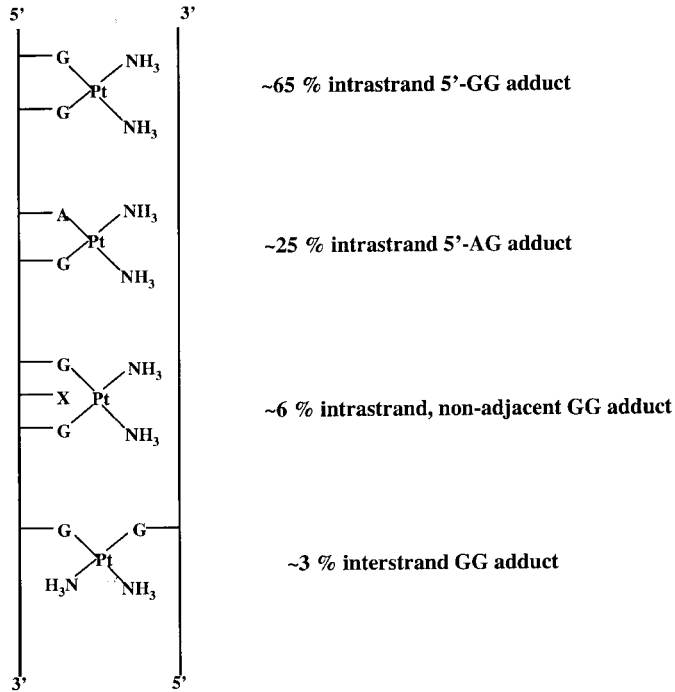


Fig. 4.1 Schematic representation of the different DNA adducts formed by cisplatin.

reaction specific, 5'-AG-3'. No exception to this directionality has been found. In addition, no monofunctional adducts to A have been detected [41].

The distortion of the tertiary structure of DNA induced by cisplatin depends on which type of adduct is formed. Only two types of structures have been determined, either by means of X-ray crystallography or NMR – the GG intrastrand and interstrand adducts [42–44]. There is, however, good reason to believe that 5'-AG intrastrand adduct is structurally very similar to the GG counterpart.

The formation of the intrastrand 5'-GG adduct disrupts the helical structure by de-stacking the two adjacent base pairs and locally unwinding DNA at the site of the lesion, thus creating a hydrophobic pocket facing the minor groove, which is widened and flattened (Fig. 4.2). As a consequence, a kink in the helical axis is introduced towards the major groove, with an average value of $\sim 50^\circ$ in crystal structure and $\sim 70\text{--}80^\circ$ in NMR studies. Larger

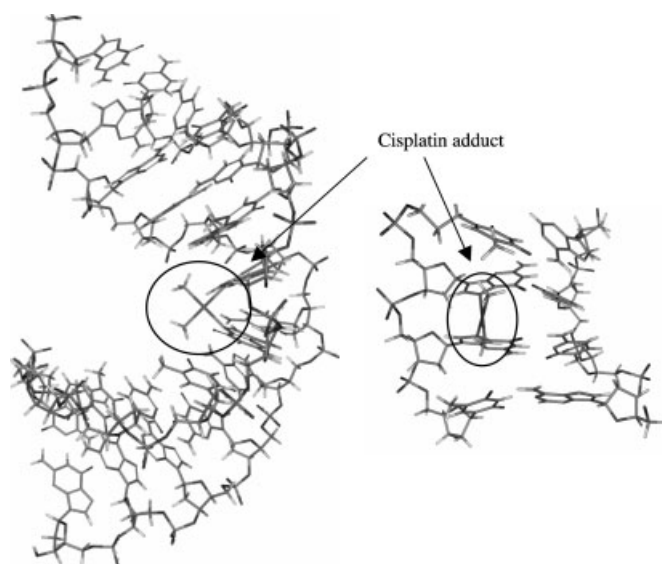


Fig. 4.2 NMR-determined structure of a dodecamer duplex DNA complexed with cisplatin (left) and a close up of the induced damage (right). The kink is clearly visible and centred around the site of complexation. PDB entry 1A84.

distortions are observed if the adduct is flanked by less rigid (A–T) base pairs. The distortion displaces the centrally coordinated platinum atom out of the plane of the bases by 0.8–1.0 Å, and places the top ammine group within hydrogen bonding distance of one of the oxygen atoms of the backbone phosphate group (Fig. 4.2).

Although relatively few structural studies of the interstrand GG adduct [42, 45, 46] have been reported, the data presented reveal the structural distortion to be significantly different from that of the intrastrand adduct. The prime feature of this adduct is the cross-linking between the two strands at GC sequences, thereby causing a kink in the double helix. In this instance, however, the kink is towards the minor groove, with a value of $\sim 47^\circ$ (Fig. 4.3). Another feature of this adduct not present in the other intrastrand adducts is the complementary cytosine bases extruding from the lesion site.

As for the intrastrand adduct, however, the platinum atom is forced out of the plane of the bases by 0.3–0.6 Å. One study revealed a very well ordered water structure around the site of the

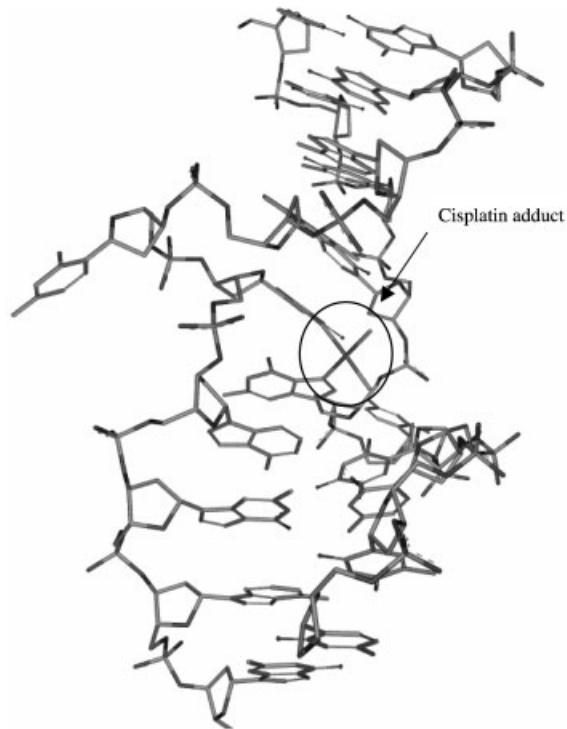


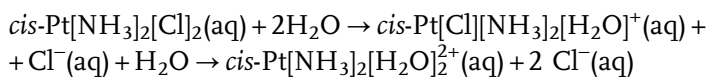
Fig. 4.3 View from the major groove of the inter-strand lesion caused by cisplatin. Note the complementary cytosine bases extruding from the duplex. PDB entry 1A2E.

platinum lesion [42]. In particular, two water molecules were found on the quaternary axis of the platinum square, well positioned for nucleophilic attack on platinum. This might account for the relative instability (compared to the intrastrand adduct) of this lesion and hence the low ratio of this adduct [47].

4.3.1

Activation Reactions

Despite the importance of the drug, many questions about the mechanism of cisplatin activation, and its binding to DNA, still remain unresolved. Activation of cisplatin is believed to be a two-step process:



Although the specific mechanisms of action of cisplatin and its derivatives against different tumors are beginning to unfold [48–52], it is, for example, still under dispute in which form the drug reaches the cellular DNA and binds to its targets, i.e. as $\text{Pt}[\text{NH}_3]_2[\text{H}_2\text{O}][\text{Cl}]^+$ or as $\text{Pt}[\text{NH}_3]_2[\text{H}_2\text{O}]_2^{2+}$ [53–55].

Experimental data on cisplatin published thus far include X-ray structure determination of the parent compound [56], the barrier height of the first aquation step, $cis\text{-Pt}[\text{NH}_3]_2[\text{Cl}]_2 + \text{H}_2\text{O} \rightarrow cis\text{-Pt}[\text{NH}_3]_2[\text{H}_2\text{O}][\text{Cl}]^+ + \text{Cl}^-$ [57–59], barriers for the first chloride anation of the fully aquated cisplatin, $cis\text{-Pt}[\text{NH}_3]_2[\text{H}_2\text{O}]_2^{2+} + \text{Cl}^- \rightarrow cis\text{-Pt}[\text{NH}_3]_2[\text{H}_2\text{O}][\text{Cl}]^+ + \text{H}_2\text{O}$ [60, 61], and kinetic data [53, 54, 62]. The most accurate previous theoretical predictions stem from Car-Parrinello (DFT-MD) simulations of the first substitution step, by Carloni et al. [63], and a recent DFT study (mPW1PW91/SDD), by Zhang et al. [64], in which both aquation steps were followed. In the latter study the estimated standard Gibbs energies of activation involved calculations with and without a solvent model, and gave $24.1 \text{ kcal mol}^{-1}$ (Onsager solvation model) and $23.6 \text{ kcal mol}^{-1}$ (gas phase) for the first aquation, and $32.3 \text{ kcal mol}^{-1}$ and $36.6 \text{ kcal mol}^{-1}$, respectively, for the second aquation.

In this work we employed the B3LYP hybrid functional [21, 22] in combination with the Lanl2DZ basis set [65] for geometry optimization and frequency calculations *in vacuo*; solvation energies were obtained from single-point calculations using the D-PCM method [66], with water as solvent. All calculations were performed with the Gaussian 98 software package [67].

The water-substitution reactions of cisplatin proceed by way of trigonal pyramidal transition states (TS); in Fig. 4.4 we show a schematic representation of the first aquation reaction. In the present work we have chosen to maintain the stoichiometry through the entire reaction sequence (i.e., both the attacking water molecules, L_{ent} , and the two leaving chloride anions, L_{leav} are present throughout). In Tab. 4.1 we present relevant geometric data at the various stationary points. It is worth noting that the Pt complexes in the minima (reactant, intermediate and product) are almost perfectly square planar, and that the two transition state structures are perfectly trigonally bipyramidal ex-

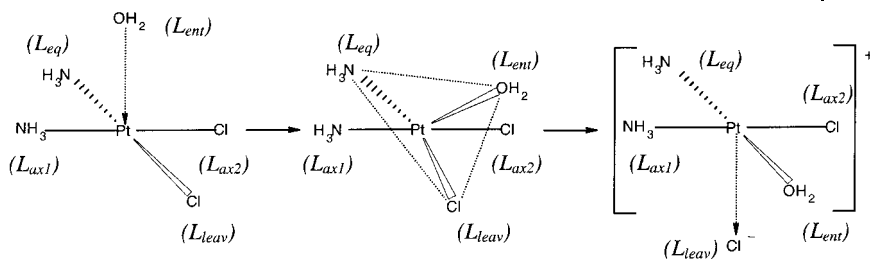


Fig. 4.4 Idealized reaction mechanism of the first step in the aquation of cisplatin, showing the labeling of the different ligands.

Tab. 4.1 B3LYP/LanI2DZ-optimized geometric data for all stationary points in the double aquation reaction of cisplatin. The experimental values [56] follow the convention used in the text although no nonbonded ligands are present.

	Exp.	RC1	RC1 _{alt}	TS1	TS1 _{alt}	PC1	RC2	TS2	PC2
<i>Distance (Å)</i>									
L _{ent} -Pt	—	3.67	3.66	2.51	2.48	2.06	3.68	2.40	2.07
L _{leav} -Pt	2.33	2.44	2.44	2.79	2.81	4.01	2.44	2.80	3.98
L _{eq} -Pt	2.01	2.09	2.09	2.09	2.09	2.07	2.09	2.09	2.07
L _{ax1} -Pt	2.01	2.09	2.09	2.08	2.08	2.09	2.07	2.06	2.07
L _{ax2} -Pt	2.33	2.44	2.44	2.44	2.44	2.44	2.05	2.06	2.07
<i>Angle (°)</i>									
L _{ent} -Pt-L _{leav}	—	56.6	56.8	67.1	67.2	42.9	56.6	67.0	43.5
L _{ent} -Pt-L _{eq}	—	123.9	124.5	148.5	148.1	178.0	125.8	152.3	176.9
L _{ent} -Pt-L _{ax1}	—	46.9	47.1	88.8	88.9	89.1	46.2	82.2	91.1
L _{ent} -Pt-L _{ax2}	—	132.2	131.6	85.3	86.8	87.2	132.4	91.6	86.0
L _{leav} -Pt-L _{eq}	—	179.0	178.7	144.1	144.6	139.1	176.6	140.3	139.0
L _{leav} -Pt-L _{ax1}	90.2	85.5	85.4	74.8	74.8	49.0	90.2	79.3	49.5
L _{leav} -Pt-L _{ax2}	91.9	93.6	93.6	102.1	102.5	127.4	87.5	94.5	127.9
L _{eq} -Pt-L _{ax1}	87.0	95.5	95.6	96.1	98.1	92.8	93.2	97.1	91.8
L _{eq} -Pt-L _{ax2}	90.2	85.5	85.4	89.3	86.2	90.9	89.2	90.2	91.1
L _{ax1} -Pt-L _{ax2}	—	179.0	178.7	174.1	175.6	176.3	177.6	172.6	176.9

cept that the equatorial planes are slightly skewed with respect to the central axes. The bond lengths of ligands not directly involved in the substitution at hand are only marginally affected between the different states (i.e. within 0.01–0.02 Å).

The optimized structure of cisplatin in this study, the reactant complex, has a Pt–Cl bond distance of 2.44 Å compared with

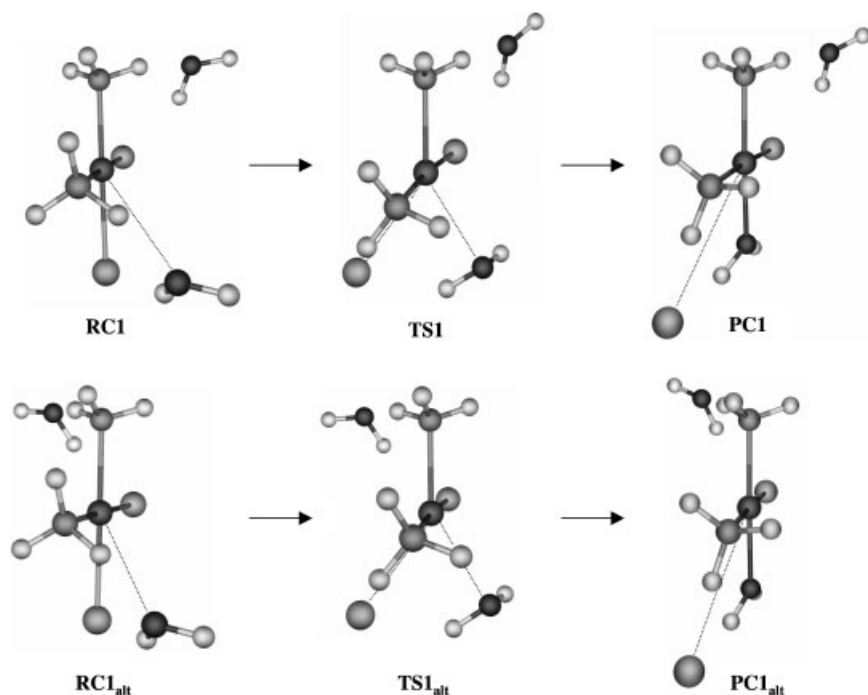


Fig. 4.5 Optimized structures of the first aquation reaction of cisplatin. The top row shows the reaction starting from a *syn* arrangement of the nonbonded water molecules.

Bottom row, as above but starting from the alternative, *anti* arrangement of the nonbonded water molecules.

2.33 Å determined by X-ray crystallography [56], whereas the distance to the amine groups deviates less, 2.09 Å in this work and 2.01 Å experimentally. Previous theoretical studies by Carloni et al. [63] and Zhang et al. [64] yielded Pt–Cl distances of 2.34 and 2.37 Å, respectively. The difference from the more recently obtained value is related to the choice of basis set and exchange-correlation functional.

The optimized structures of the first water substitution are shown in Fig. 4.5. The initial reactant complex (RC1) is symmetric with the two water molecules located on the same side of the plane formed by the nonaquated cisplatin, the *syn* arrangement. The oxygen atoms of each water molecule are hydrogen bonded to one of the amine groups, at a distance of 1.7 Å. The lowest vibrational level (31.9 cm^{-1}) corresponds to the reaction coordinate for water substitution. An alternative complex with

the two water molecules on opposite sides of the cisplatin plane, the *anti* arrangement (RC_{alt}), was also found. The geometric data and energetics for this system are almost identical with those of the *syn* form.

The transition structures have single imaginary frequencies of -170.2 cm^{-1} and -172.3 cm^{-1} , respectively, characterized by a bond angle vibration of L_{eq} around the bipyramidal axis, a decrease in the Pt- L_{ent} distance, and an increase in the Pt- L_{leav} distance. The low value of the imaginary frequencies indicates a very flat potential-energy surface around the transition state geometry. In the intermediate complexes the nonbonded Cl^- interacts with the acidic hydrogens on L_{ent} and L_{ax1} (H_2O and NH_3 , respectively). The relative position of the water molecule that is not participating in the substitution reaction remains unchanged along the reaction path.

The characteristics of the second aquation follow the first very closely. The vibrational mode characterizing the TS is -165.7 cm^{-1} , again indicating a very flat potential energy surface around the saddle point. The product complex reached from the initial *syn* arrangement of the water molecules (PC1) passes to the second transition state via the stationary point of PC1_{alt} during the course of the optimization. The final, disubstituted, product complex (PC2) is, as was the initial reactant complex, a symmetrical ternary complex with the two nonbonded chlorides now residing on opposing sides of the diaquated cisplatin plane, i.e. in an *anti* conformation (Fig. 4.6). The lowest vibrational eigenmode consists of variations of the Pt- Cl^- distances, and constitutes the reaction coordinate for the reverse reaction.

As mentioned above, results from kinetic experiments [53, 54, 62] are inconclusive about the state of activated cisplatin, and have under some experimental conditions (chloride-depleted environment) revealed the most probable state of activated cisplatin to be the mono-aquated form, $\text{cis-Pt}[\text{Cl}][\text{NH}_3]_2[\text{H}_2\text{O}]^+$. The first aquation reaction has been determined to be approximately two orders of magnitude faster than the second. This opinion is far from undisputed, however, and results from other experiments, based on the ratio of the amounts of DNA adducts formed by the different aquated cisplatin moieties [54], strongly indicates that the likely state of cisplatin binding to DNA is in fact the diaquated state.

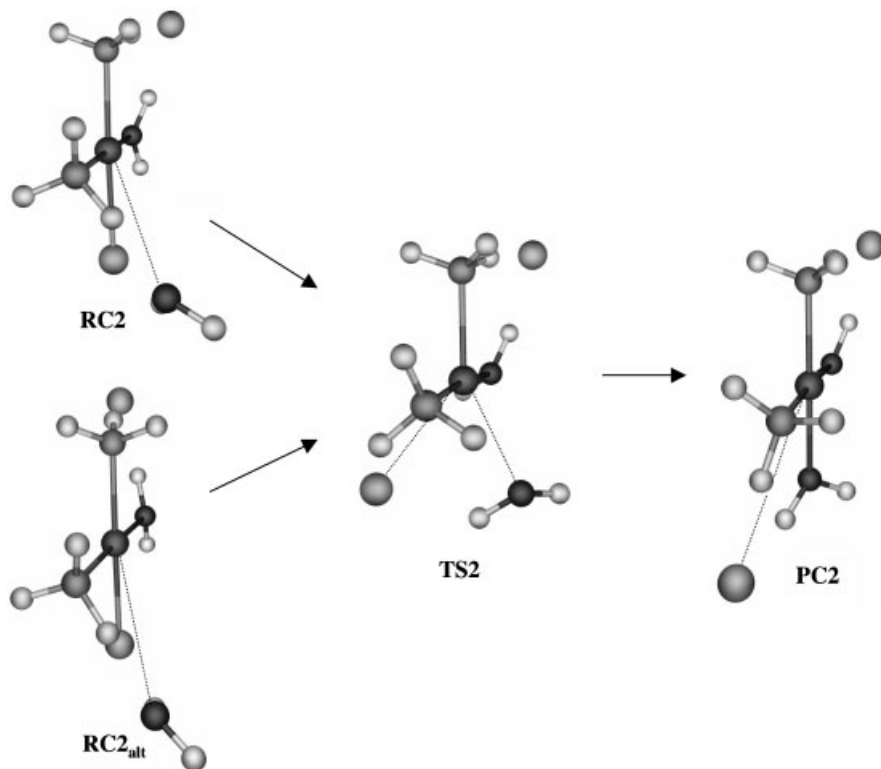


Fig. 4.6 Optimised structures of the second aquation reaction of cisplatin. Both RC2 and RC2_{alt} converge to the same transition structure (TS2) and consequently the same PC2, but RC2 does so by passing through RC2_{alt} geometry.

In a study supporting mono-aquated cisplatin as the activated state [53], detectable levels of diaquated cisplatin appeared approximately 3 h after mono-aquated cisplatin, indicating a difference in barrier height between the two hydrolytic substitutions of ca $1.5 \text{ kcal mol}^{-1}$. As is apparent from our computed energy surfaces (Fig. 4.7), the experimental data are supported by calculations using PCM energy corrections, which give the first barrier height as $20.7 \text{ kcal mol}^{-1}$ and the second as $21.7 \text{ kcal mol}^{-1}$. By use of Car-Parrinello simulations employing the BLYP exchange-correlation functional and a customized pseudopotential basis set Carloni et al. [63] estimated the (nonoptimized) barrier height for the first hydrolytic substitution to be $\sim 21 \text{ kcal mol}^{-1}$. This value is in good agreement with experimental data and our calculations.

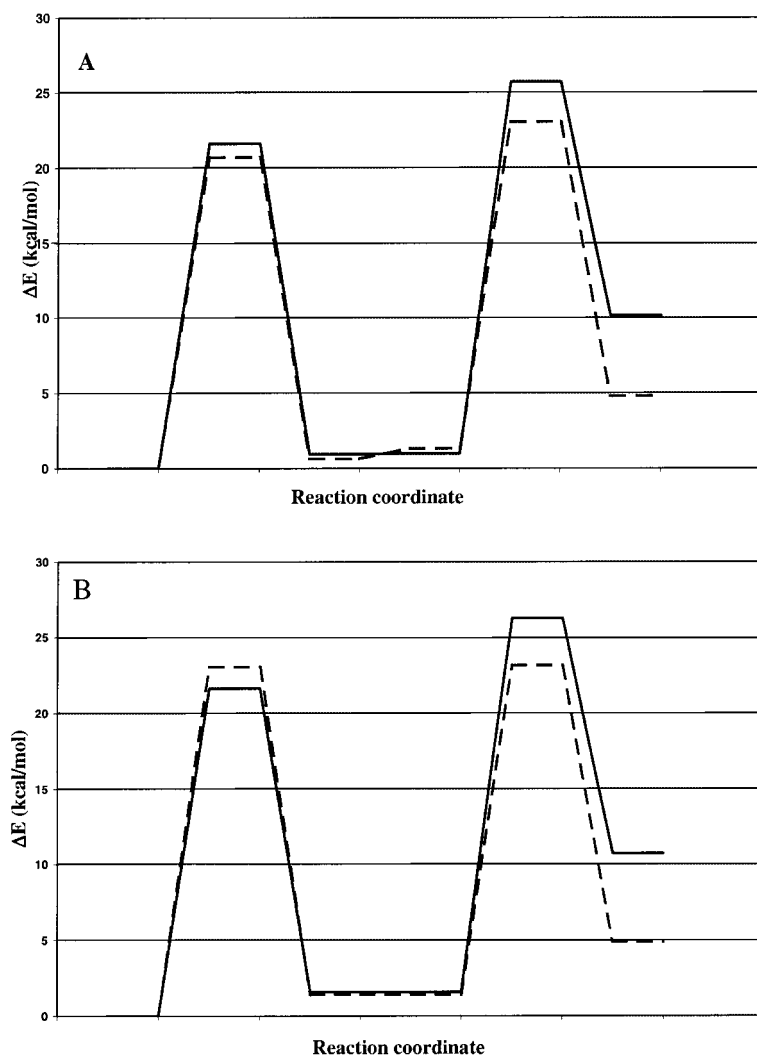


Fig. 4.7 Reaction energy surfaces for the first and second aquation reactions of cisplatin starting from (a) the *syn* arrangement of the initial reactant complex and (b) the alternative *anti* arrangement. Solid lines indicate gas phase; dashed lines include the PCM solvation model.

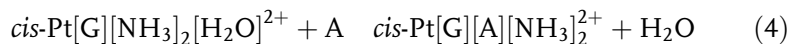
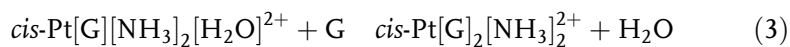
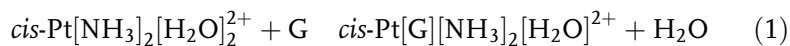
The barrier height for chloride anation of diaquated cisplatin has been determined experimentally to be $16.6 \text{ kcal mol}^{-1}$ [60, 61], in good agreement with current computational results of $18.2 \text{ kcal mol}^{-1}$. The reaction is endothermic by ca $4.5 \text{ kcal mol}^{-1}$.

On the basis of the calculations presented here, especially considering the relative stability of the reactant and product complexes, it is clear that the form of activated cisplatin that partakes in subsequent reactions will be highly dependent on the composition of the environment in which it acts; this explains, to some extent, the disparate conclusions of the experimental work [53, 54].

4.3.2

Interactions Between DNA and Cisplatin

Clarification of interactions between activated cisplatin and DNA has primarily been limited to experimental determination of the structures of the final didentate adducts [42–44]. Carloni et al. have explored, theoretically, the geometry of the intrastrand didentate GG adduct [63]. As always when investigating reactions involving large molecules such as DNA the choice of model system is crucial. In this instance it would be preferential to include a large portion of DNA, because hydrogen bond donors and acceptors are likely to influence the course of adduct formation. Because of computational limitations, however, here we have restricted our treatment to a small model system comprising the activated (doubly substituted) cisplatin and the nucleobase in question. This reduction makes the investigation mainly qualitative, and the reported values for barrier heights, etc., should be considered with this in mind. The following reactions are investigated:



In this set of computations we employed the pure DFT functional BLYP [16, 19] in combination with the Lanl2DZ basis set [65], as implemented in the Gaussian 98 program package [67]. The geometrical data presented follow the same notation as in the previous section, i.e. L_{ent} is the entering ligand (N7 of a purine base) and L_{leav} is the leaving ligand (always H_2O).

Tab. 4.2 Geometric data for the first nucleobase substitutions by doubly aquated cisplatin for adenine (top) and guanine (bottom). The terms are defined in Fig. 4.4.

	Configuration 1			Configuration 2		
	RC	TS	PC	RC	TS	PC
<i>Angle (°)</i>						
L _{ent} -Pt-L _{leav}	35.0	68.0	80.6	24.8	69.5	68.2
L _{ent} -Pt-L _{eq}	141.4	138.8	173.1	155.2	139.8	178.5
L _{leav} -Pt-L _{eq}	176.4	151.6	97.4	178.0	150.6	113.1
L _{ent} -Pt-L _{ax1}	80.2	90.1	91.0	70.1	89.7	88.1
L _{ent} -Pt-L _{ax2}	97.5	90.9	92.5	107.8	89.7	92.8
L _{leav} -Pt-L _{ax1}	93.5	86.2	66.4	84.9	80.8	37.2
L _{leav} -Pt-L _{ax2}	83.7	93.6	110.0	92.9	99.5	144.5
L _{eq} -Pt-L _{ax1}	85.3	85.0	82.2	96.7	95.1	93.4
L _{eq} -Pt-L _{ax2}	97.5	94.67	94.4	85.4	85.0	85.7
<i>Distance (Å)</i>						
L _{ent} -Pt	3.95	2.60	2.06	4.34	2.68	2.06
L _{leav} -Pt	2.07	2.53	4.84	2.06	2.38	4.06
L _{eq} -Pt	2.12	2.10	2.11	2.12	2.10	2.13
L _{ax1} -Pt	2.10	2.07	2.06	2.07	2.06	2.07
L _{ax2} -Pt	2.08	2.13	2.14	2.14	2.15	2.16
	Configuration 1			Configuration 2		
	RC	TS	PC	RC	TS	PC
<i>Angle (°)</i>						
L _{ent} -Pt-L _{leav}	31.2	68.1	72.3	27.1	68.6	75.0
L _{ent} -Pt-L _{eq}	145.7	141.7	174.5	152.2	140.5	174.1
L _{leav} -Pt-L _{eq}	176.9	150.1	112.8	175.9	150.8	101.3
L _{ent} -Pt-L _{ax1}	77.3	88.2	92.5	78.8	91.6	91.9
L _{ent} -Pt-L _{ax2}	99.6	92.3	90.7	100.1	89.1	91.6
L _{leav} -Pt-L _{ax1}	93.5	101.5	143.9	88.2	79.8	37.8
L _{leav} -Pt-L _{ax2}	83.4	79.1	40.2	90.6	100.8	147.3
L _{eq} -Pt-L _{ax1}	84.4	81.9	82.6	95.4	93.8	82.5
L _{eq} -Pt-L _{ax2}	98.7	97.5	94.9	85.7	85.4	93.9
<i>Distance (Å)</i>						
L _{ent} -Pt	4.10	2.62	2.07	4.23	2.62	2.08
L _{leav} -Pt	2.07	2.47	4.07	2.06	2.41	3.97
L _{eq} -Pt	2.11	2.09	2.11	2.12	2.10	2.11
L _{ax1} -Pt	2.10	2.08	2.09	2.06	2.05	2.08
L _{ax2} -Pt	2.08	2.09	2.10	2.15	2.16	2.11

Tab. 4.3 Geometric data for adenine and guanine attack on the guanine-cisplatin mono adduct. The terms are defined in Fig. 4.4.

	GA			GG		
	RC	TS	PC	RC	TS	PC
<i>Angle (°)</i>						
L _{ent} -Pt-L _{leav}	29.1	67.8	106.8	28.5	74.3	127.6
L _{ent} -Pt-L _{eq}	152.2	146.0	176.6	153.2	142.6	178.2
L _{leav} -Pt-L _{eq}	176.4	145.9	69.8	175.3	142.5	50.8
L _{ent} -Pt-L _{ax1}	77.3	84.1	86.5	77.0	90.3	90.4
L _{ent} -Pt-L _{ax2}	100.8	97.3	93.7	99.5	90.6	91.6
L _{leav} -Pt-L _{ax1}	85.0	90.4	73.0	83.1	78.2	72.1
L _{leav} -Pt-L _{ax2}	92.3	90.4	106.3	92.6	98.7	105.8
L _{eq} -Pt-L _{ax1}	92.3	89.9	92.5	93.2	91.8	88.3
L _{eq} -Pt-L _{ax2}	90.3	88.7	87.3	91.1	89.6	89.8
<i>Distance (Å)</i>						
L _{ent} -Pt	4.13	2.61	2.07	4.10	2.54	2.08
L _{leav} -Pt	2.06	2.42	3.43	2.05	2.39	3.44
L _{eq} -Pt	2.14	2.12	2.12	2.12	2.12	2.10
L _{ax1} -Pt	2.11	2.11	2.11	2.10	2.09	2.10
L _{ax2} -Pt	2.07	2.08	2.09	2.07	2.08	2.07

The reactions involving the activated cisplatin and the purine bases are very similar to the hydrolytic substitutions presented in Section 4.3.1, with the main difference being the introduction of a hydrogen bond acceptor/donor on the purine base (O6 in guanine, N6 or its adherent hydrogens in adenine). The main geometric features are covered in Tabs. 4.2 and 4.3. The adduct bond lengths (i.e. the bond between N7 and Pt) derived experimentally are ~ 1.99 Å from X-ray crystallography and ~ 2.05 Å from NMR determination [42–44], whereas we obtain a value of 2.07 Å.

For the formation of the first adduct both adenine and guanine yield two different reactant complexes with corresponding transition states and product complexes, as shown in Figs. 4.8 (adenine) and 4.9 (guanine). The main difference between reactant complexes of types 1 and 2 is a 90° rotation of the cisplatin moiety. Reactant complexes of type 1 have hydrogen bonds involving the two water ligands of cisplatin and N7 and O6/N6, whereas in type 2 the hydrogen bonding groups of cisplatin are one water and one amine group. For all reactant complexes there is partial proton transfer

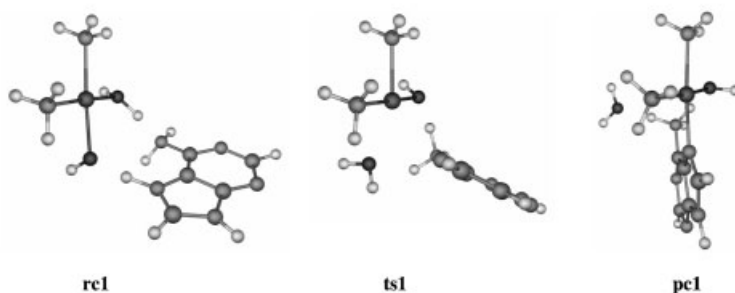
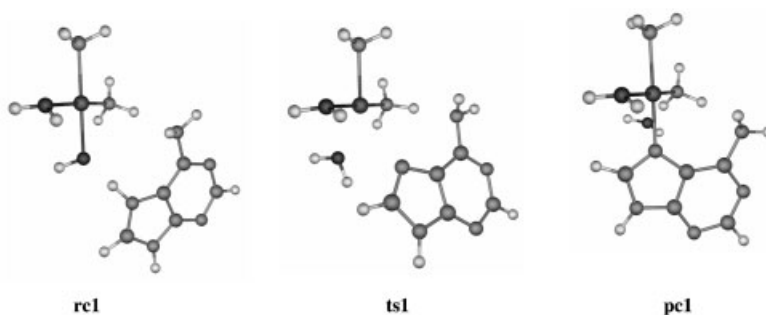
Conformation 1**Conformation 2**

Fig. 4.8 The first substitution reactions between activated cisplatin and adenine.

from a water ligand to the N7 position of the purine base. The acidity of the water ligands can, at least in part, be attributed to the electron-withdrawing properties of the highly charged platinum.

Geometrically, the main differences between the different complexes, when following the reactions towards the product, is the rotation and tilt of the attacking nucleobase; all other distances and angles are highly similar, as is apparent from Tabs. 4.2 and 4.3. Similarly to the aquation reactions, the imaginary frequencies of the transition states are very low ($\sim -150 \text{ cm}^{-1}$), indicating flat potential surfaces with large degrees of freedom when passing over the barrier.

The energy surfaces for formation of the first adducts are given in Fig. 4.10. The barriers to addition are all very high, 35–40 kcal mol⁻¹, a value much too high for any significant amount of adduct formation under physiological conditions. The most probable expla-

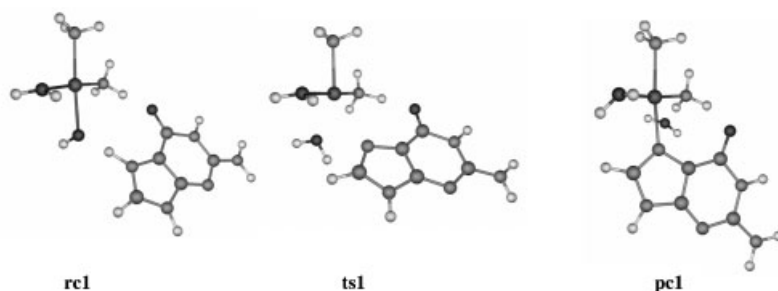
Conformation 1**Conformation 2**

Fig. 4.9 The first substitution reactions between activated cisplatin and guanine.

nation is the use of a small model system in vacuum. Protonation of the reactant complexes reduces the energy of these to a much greater degree than if the optimizations included a solvation model, or a larger fraction of the surroundings. We note, however, a systematic difference between the conformers amounting to $\sim 7\text{--}10\text{ kcal mol}^{-1}$. This is related to the difference between the interaction energy of the O6/N6-H₂O hydrogen bond and that of the O6/N6-H₃N hydrogen bond, and could possibly constitute a selection criterion for the preferred route to a bifunctional adduct in DNA (see below). We also note that the reaction energetics alone do not explain why only monofunctional adducts to G are found (or the 5'-AG directionality), instead the explanation must be sought in steric and hydrogen bonding interactions with the surroundings.

For the second substitution reaction, we have chosen to start from the guanine complex of conformer 1 only. The optimized geometric structures for the reactant complexes, transition states,

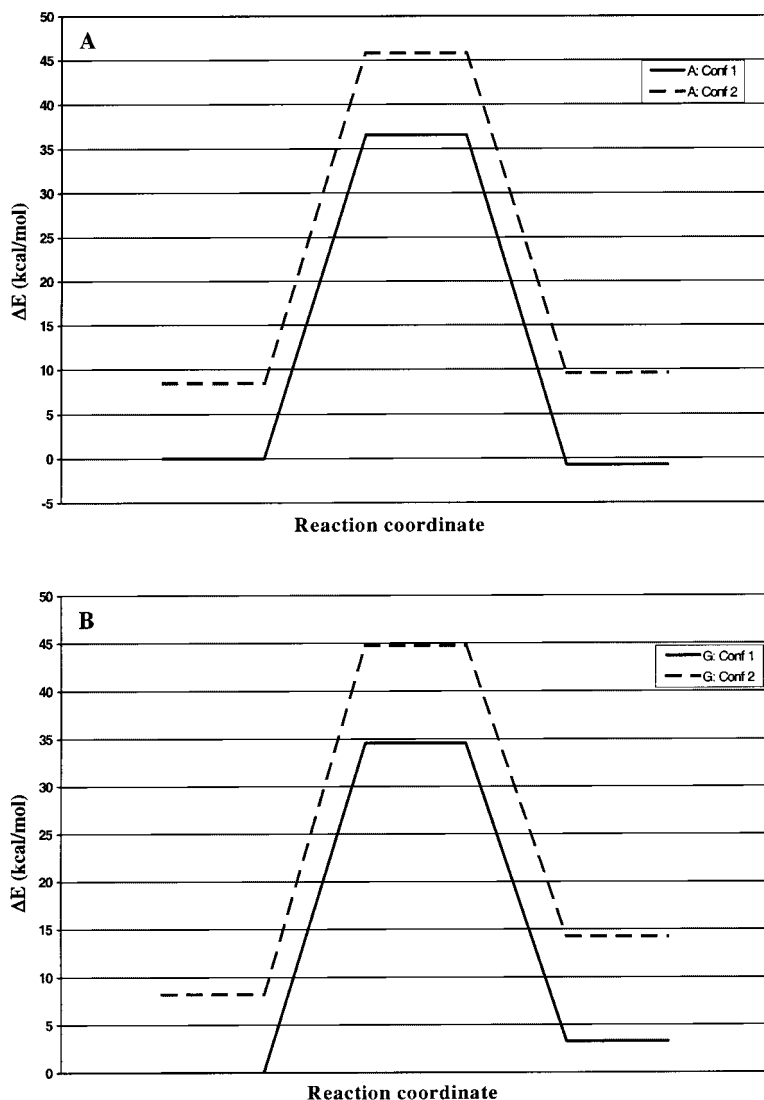


Fig. 4.10 Reaction energies for cisplatin-nucleobase monoadduct formation with (A) adenine and (B) guanine. The solid line denotes conformation 1 and the dashed line conformation 2.

and products for adenine and guanine attack on $\text{cis-Pt[G]([NH}_3)_2(\text{H}_2\text{O})^{2+}$ are displayed in Fig. 4.11; relevant geometric data are listed in Tab. 4.3.

Although the basic geometrical features are the same as in the first substitution, an interesting difference between the two substi-

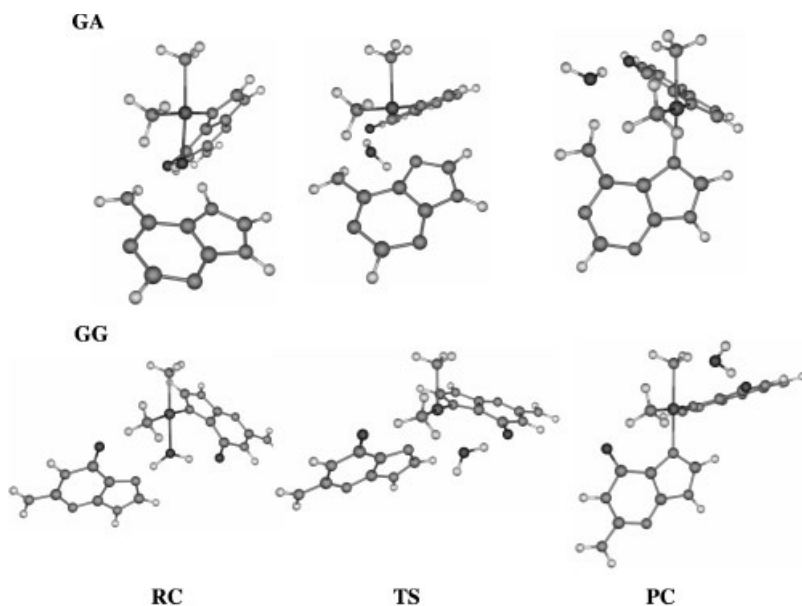


Fig. 4.11 The second nucleobase substitution of cisplatin. Top row: formation of GA product complex; bottom row: formation of GG product complex. All reactions employ guanine-cisplatin product conformer 1 as reactant.

tutions is the geometry of the product complexes. Whereas the GA complex reproduces the overall structure of an intrastrand bifunctional adduct fairly well, the GG complex resembles more closely that of an interstrand adduct. This is either because of the extended degrees of freedom of this model or because the final structures are very dependent on the initial conformations. For the GA reaction complex we observe the same type of partial proton transfer from water ligand to nucleobase N7, as noted above.

Energetically the two reactions again have very high barriers (Fig. 4.12), with the GG system having a slightly lower barrier and being somewhat more thermodynamically favored. Again, the energetics alone do not explain the observed differences between product distribution. They do, however, present a schematic view of the reactions occurring, the role of the different ligands, and the importance of stabilizing hydrogen bonding and possible steric hindrance in the full DNA system.

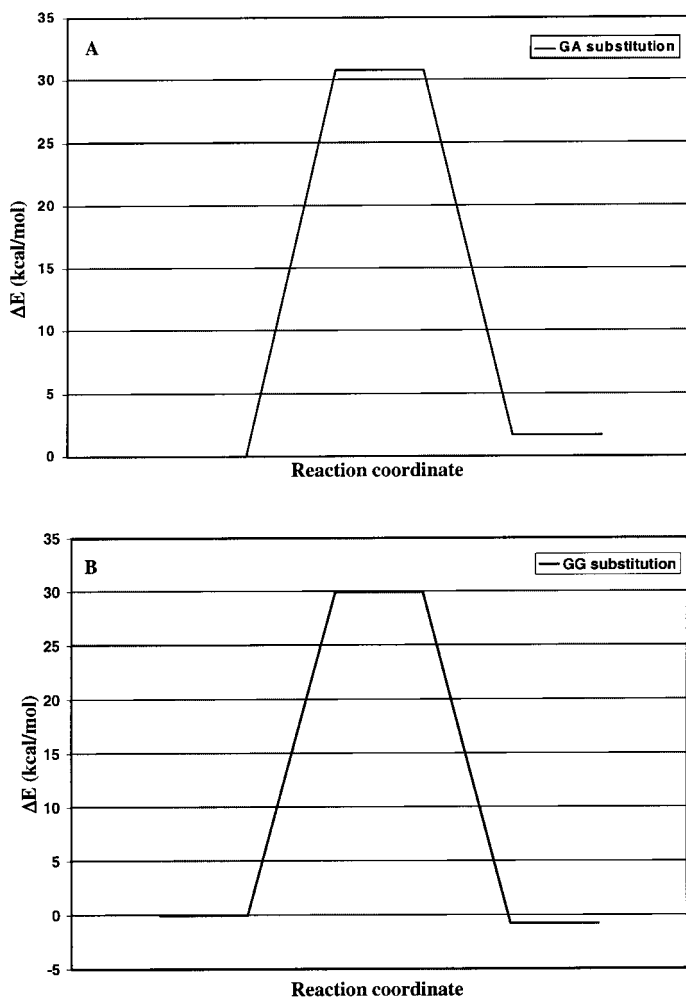


Fig. 4.12 Reaction energies for the second nucleobase-cisplatin reaction (Fig. 4.11). (A) Formation of GA complex and (B) formation of GG complex.

4.4 Photochemistry of Psoralen Compounds

Psoralens, or furocoumarins, are a class of heterocyclic aromatic compounds used in photochemotherapy treatment of a variety of skin diseases such as psoriasis, vitiligo, mycosis fungoides, polymorphous light eruption, and more [68–71]. The compounds are present in numerous plants throughout the world. In photoche-

motherapy the drug is either applied topically or administered orally, after which the patient is irradiated with UV-A light (320–400 nm). Occasionally visible light can also be used [72]. Increased aromaticity of the systems, compared with smaller heterocycles, also enables transitions with strong bands in the 250–300 nm range. These wavelengths are however too energetic for photoactivation, and instead lead to photodegradation reactions.

On absorption of the UV-A/visible radiation the psoralen can undergo several different reactions [73]. In oxygen-dependent type I reactions the compound is raised to the first excited singlet state and might, via intersystem crossing, come to reside in the relatively long-lived first excited triplet state. From there the photosensitized compound might readily perform redox chemistry, typically resulting in formation of reactive superoxide anion radicals and fragmenting substrate cations.

In oxygen-dependent type II reactions the excitation energy of the first excited triplet state of the drug is transferred to molecular oxygen. The excited singlet molecular oxygen in turn reacts rapidly and essentially without discrimination with a wide variety of biomaterials, and thus causes severe damage. Type II reactions do, however, impose some constraints on the sensitizer, such that the triplet state must be very long-lived, its triplet energy must lie above that of oxygen, and the drug itself should not be susceptible to attack. It is, however, not yet satisfactorily proven that *in-situ* formation of singlet molecular oxygen actually occurs inside biological cells.

In the oxygen-independent Type III reactions the excited/sensitized psoralen donates its excitation energy directly to, or reacts with, the target compound. This occurs if the substrate and the target compound (e.g., DNA) are already in close proximity or intercalated. The reactions will proceed very rapidly via the excited singlet state, and are, typically, cyclization reactions or electron-transfer between the sensitizer and the target. In addition, the psoralen can be ionized, either directly or via the excited state, and react with the target compound in the form of a radical cation. Furocoumarins are also employed in treatment of cutaneous T-cell lymphoma and some infections connected with AIDS, by so-called photopheresis processes [71, 74–76]. In this case, peripheral blood is exposed to, e.g., photoactivated (sensitized) 8-methoxypsoralen (8-MOP) in an extracorporeal flow system. This

is also the idea behind the recently developed pathogen eradication technology (PET), in which viruses, bacteria, and parasites are removed from blood products by addition of riboflavin (vitamin B₂) and exposing the mixture to visible light. The process is believed to proceed through UV-induced electron ejection from the flavin ring [77].

Exactly which biochemical moieties the activated psoralen compounds (or the thereby activated molecular oxygen) bind to or interact with to function as specific drugs in the wide variety of diseases listed above is, however, not yet fully understood. It is, for example, known that photoactivated psoralens can interact with membrane and cytoplasmic receptors [78, 79], that they form C₄-cycloadducts with unsaturated fatty acids and lecithins in lipid membranes [80–82], that they form cross-links with DNA (primarily by binding to the two thymine residues of opposite strands in an AT sequence) [83–85], and that they inactivate enzymes and ribosomes [86]. In particular, some psoralens are known to inhibit the epidermal growth factor (EGF) receptor on exposure of the receptor-psoralen complex to UV-A light [79, 87]. The biological effects the drugs have on skin disorders has been attributed to this particular property [88]. Psoralens are also used in nucleic acid research aimed at better understanding of DNA damage and repair processes [89, 90]. That psoralen compounds bind to DNA is also a complicating factor in the treatment of skin disorders, because long-term exposure can cause mutations and thus the development of skin cancer [91–93].

Here we will present some basic theoretical photochemistry of a set of furocoumarin compounds, starting from the basic building blocks furan and pyrone, as shown in Fig. 4.13. Comparison is also made with the photochemistry of flavin, the active component of vitamin B₂.

4.4.1

Ionization Potentials

As mentioned above, one type of mechanism proposed for photoactivation is that the drug is ionized by the radiation and the electron is taken up by the target compound (e.g. a nucleobase), with subsequent rearrangements, fragmentations, dimerizations,

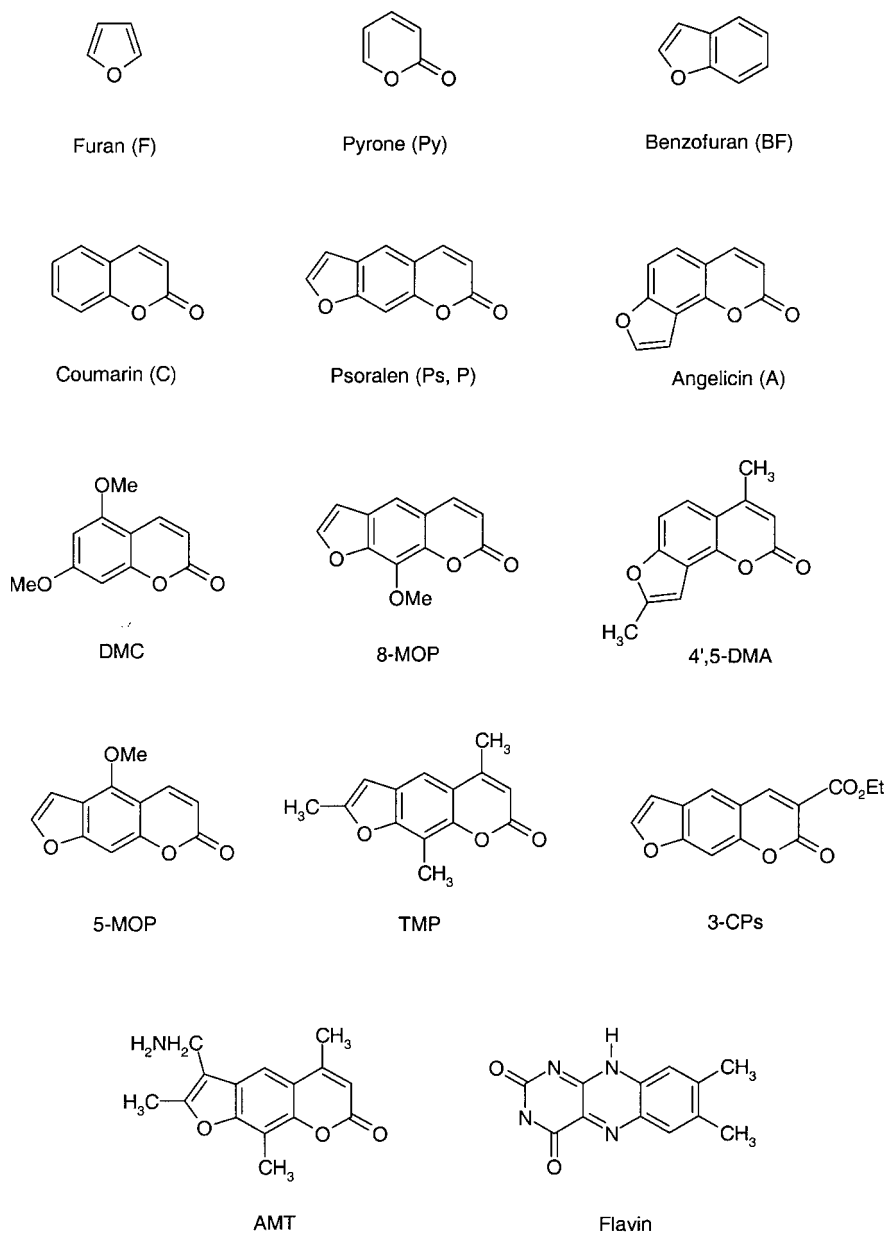


Fig. 4.13 Furocoumarins and related compounds investigated in this work.

and other reactions. It might thus be of interest to compare the vertical and adiabatic ionization potentials (VIP and AIP, respectively) of a set of compounds, and the effects of the surroundings on these. The difference between the VIP and AIP indicates the extent of relaxation of the cation, i.e. how much the structure of the compound is effected by the ionization. All structures and energies described in this section were computed at the B3LYP/6-31G(d,p) level, and are reported without inclusion of zero-point vibrational effects (ZPE). The effect of bulk water as solvent is included by use of single-point calculations using the polarized continuum model (PCM) of Tomasi and coworkers [66], and all calculations are performed with the Gaussian 98 program package [67].

Starting with the small “building blocks” furan (F) and pyrone (Py), we see from Tab. 4.4 that furan is easier to ionize than the larger pyrone ring by between 4 and 8 kcal mol⁻¹, depending on vacuum or bulk water surroundings. The ionization potentials of the small heterocycles are, however, significantly larger than if we fuse the system to a benzene ring to form benzofuran (BF) and coumarin (C), respectively. The effect of the increased aromaticity is smaller in aqueous solution. Finally, going to the full parent compounds psoralen (P) and angelicin (A), the ionization potentials of these lie approximately 10 kcal mol⁻¹ lower than for coumarin, but are still higher than those for the benzofuran system. We also note that for all systems investigated, the inclusion of the solvent reduces the IP by approximately 40 kcal mol⁻¹ (1.7 eV).

Several modifications of the basic parent compounds have also been suggested, primarily by methyl or methoxy substituents (cf. Fig. 4.13). Two of the most active and widely used psoralens are 8-methoxypsoralen (8-MOP) and trimethylpsoralen (TMP). With the exception of 3-CP, the substituents reduce the ionization energies by 5–10 kcal mol⁻¹ compared with the unsubstituted parent compounds (Tab. 4.4). In bulk water the effects are somewhat smaller. The largest effects are observed for TMP and AMT. The IP of flavin are very similar to those of the unsubstituted psoralen.

The relaxation effects of the cation are throughout rather small, of the order of 2–10 kcal mol⁻¹. The reason for this is that most aromatic compounds retain their planarity on ionization, and thus structural reorganizations are small. Similar observa-

Tab. 4.4 Vertical and adiabatic ionization energies (kcal mol⁻¹) for a set of furocoumarins and related compounds, in vacuum and in bulk water solvent. 1 eV = 23.06 kcal mol⁻¹.

Compound	Vac, VIP	Vac, AIP	Water, VIP	Water, AIP
Furan	200.2	196.2	144.2	135.3
Pyrone	203.6	200.2	152.0	142.5
Benzofuran (BF)	185.4	181.4	139.0	131.9
Coumarin (C)	194.0	191.3	148.7	146.1
Psoralen (P)	182.8	178.5	141.1	135.9
Angelicin (A)	183.5	180.0	140.2	136.1
DMC	171.7	171.3	135.1	132.9
4,5-DMA	176.4	172.2	136.5	130.3
5-MOP	180.5	169.6	137.8	127.1
8-MOP	177.8	168.8	135.3	129.0
3-CP	183.7	178.9	141.7	138.1
TMP	171.3	166.7	132.9	125.2
AMT	167.1	161.5	130.0	124.7
Flavin	180.6	178.1	139.7	138.3

tions have been made for, e.g., the various nucleobases [94]. The largest effects are again noted for the more substituted compounds 5-MOP, 8-MOP, TMP, and AMT.

4.4.2

Excitation Spectra

Oxygen-free reactions of psoralens, when in close proximity to the target, proceed via the first excited states in which the 3,4- and the 4',5' π -bonds of the pyrone and furan moieties, respectively, can undergo C₄-cyclization reactions with, e.g., unsaturated bonds of lipids, or the C5=C6 double bonds of thymine in DNA. In reactions with DNA the psoralen is believed to intercalate with DNA in the dark. Subsequent irradiation at 400 nm usually leads to furan-side 4',5'-monoadduct formation, whereas irradiation at 350 nm increases the formation of crosslinks in which the furan and pyrone rings form C₄ cycloadducts to thymines on opposite strands [95]. Subsequent irradiation of the 4',5'-monoadducts at 350 nm leads to formation of crosslinks and conversion into pyrone-side 3,4-monoadducts. Shorter wave-

Tab. 4.5 Singlet excitation energies of a small subset of psoralen compounds. Experimental data in aqueous solution given in parentheses [99]. Experimental absorption coefficients are denoted w=weak (<8000), i=intermediate (8000–15 000), and s=strong (>15 000).

Furan	E_{exc} (eV)	5.87	6.05	6.31	6.51	6.67	6.94
	λ (nm)	211	205	196	190	186	179
	ϵ	0	0.169	0.032	0	0	0
BF	E_{exc} (eV)	4.96	5.11	5.73	5.92	5.96	6.13
	λ (nm)	250	243	217	210	208	202
	ϵ	0.055	0.132	0.009	0.002	0.258	0.008
Pyrone	E_{exc} (eV)	4.43	4.57	5.85	6.01	6.11	6.42
	λ (nm)	280	271	212	206	203	193
	ϵ	0.128	0	0.002	0.086	0	0.014
C	E_{exc} (eV)	4.13 (3.96)	4.45	4.59	5.39	5.76	5.96
	λ (nm)	300 (313)	279	270	230	215	208
	ϵ	0.120 (w)	0	0.167	0.017	0.117	0.006
P	E_{exc} (eV)	3.76 (3.76)	4.34	4.48	4.90	5.14 (5.08)	5.51
	λ (nm)	330 (330)	286	277	253	241 (244)	225
	ϵ	0.076 (–)	0.221	0	0.094	0.401	0.121
A	E_{exc} (eV)	3.86	4.30 (4.13)	4.58	4.84	5.09 (5.04)	5.71
	λ (nm)	321	289 (300)	271	256	243 (246)	217
	ϵ	0.042	0.156 (i/w)	0	0.112	0.426	0.010

lengths (<320 nm) can lead to photoreversal of formed adducts and degradation of non-intercalated psoralens. For the frequently utilized 8-MOP compound this is particularly efficient at $\lambda = 300$ nm [76, 96].

In Tab. 4.5 we report excitation energies and probability coefficients for the six lowest singlet excitations of the smaller set of compounds F, Py, BF, C, P, and A, and compare with experimental data as available. All excitation energies are computed at the B3LYP/6-31+G(d,p) level in vacuum on the B3LYP/6-31G(d,p) optimized structures. This has been found to yield good accuracy for nucleobase excitations, and is generally accurate to within 5 kcal mol⁻¹ (ca 0.2 eV). The effects of including conventional solvent effects (PCM) at this level of approximation are usually very small.

From the table it is apparent that the excitation spectra of furan and pyrone both lie above the UV-A region (320–400 nm) in energy. Fusing the systems to a benzene ring reduces the excita-

tion energies considerably, especially for furan, and we are now starting to approach the UV-A-active region. The calculated lowest singlet excitation for coumarin, 4.13 eV, is close to that observed experimentally (3.96 eV). The deviation between the computed and experimental values (+0.15 eV, or ca 3 kcal mol⁻¹) is indicative of the accuracy of the method.

For psoralen and angelicin, the lowest singlet excitation is somewhat lower than for coumarin, and occurs at 320–330 nm. This excitation does however have a rather low probability compared with the second lowest excited state, which lies ca 0.5 eV higher in energy. A transition with high probability is also found at 5.1 eV, which is very well reproduced by theory. It should also be noted that the excitation spectra of the two isomers P and A are very similar, both in wavelength and in transition probabilities. Differences between their photochemical behavior can thus primarily be attributed to the geometry differences. In angelicin the photoactive double bond in furan lies on the opposite side to that of the pyrone moiety. This explains why most psoralen derivatives, e.g. 5-MOP, 8-MOP, and TMP, are known to form diadducts with adjacent thymines in an AT sequence of DNA, whereas coumarin and angelicin derivatives are usually monofunctional [97, 98]. The same is true for 3-CP, in which the pyrone double bond is sterically crowded by the substituent. Monofunctional adducts are usually less genotoxic, but are also less efficient.

Turning to the more substituted compounds (Tab. 4.6), the lowest singlet excitation lies in the same range as for C, P, and A – between 3.4 and 4.1 eV. The lowest excitation energy is found for 3-CP – 3.43 eV. Most of the UV-A-active excitations (below 4 eV, or > 305 nm) have very low probability, whereas the photodegrading second set of excitations, in the energy region 4.0–4.3 eV (305–285 nm) all are of intermediate strength. A band of strong transitions also occurs at 4.9–5.2 eV (255–240 nm). Overall, there is a very good agreement between theory and experimentally measured excitations in aqueous solution.

The main effects of substitution compared with the parent compounds is reduction of the various excitation energies by ~0.1 eV and (usually) increased transition probabilities. We note, however, that there is a significant difference between the energies needed for excitation of the free psoralens in aqueous solu-

Tab. 4.6 Singlet excitation energies of substituted furocoumarins and related compounds. Experimental data in aqueous solution given in parentheses [99]. Experimental absorption coefficients are denoted w=weak (<8000), i=intermediate (8000–15 000), and s=strong (>15 000).

DMC	E_{exc} (eV)	4.01 (3.79)	4.17	4.59	5.10	5.42	5.52
	λ (nm)	309 (327)	298	270	243	229	225
	ϵ	0.270	0.084	0	0.074	0.003	0.014
5-MOP	E_{exc} (eV)	3.70	4.23 (3.97)	4.49	4.93	4.99	5.34
	λ (nm)	335	293 (312)	276	252	249	232
	ϵ	0.032	0.272 (i/s)	0	0.042	0.448	0.015
8-MOP	E_{exc} (eV)	3.57 (3.54)	4.16 (4.09)	4.51	4.93	4.98 (5.04)	5.05
	λ (nm)	348 (350)	298 (303)	275	252	249 (246)	246
	ϵ	0.022 (w)	0.233 (i)	0.002	0.194	0.254 (s)	0.001
3-CP	E_{exc} (eV)	3.43 (3.40)	4.04 (3.90)	4.17	4.27	4.89	5.07 (5.02)
	λ (nm)	362 (365)	307 (318)	297	291	254	244 (247)
	ϵ	0.075 (-)	0.389 (i)	0	0	0.199	0.292 (-)
TMP	E_{exc} (eV)	3.66 (3.70)	4.28 (4.16)	4.55	4.96	5.12 (4.96)	5.20
	λ (nm)	339 (335)	290 (298)	273	250	242 (250)	239
	ϵ	0.088 (-)	0.198 (i/w)	0	0.178	0.520 (i)	0
AMT	E_{exc} (eV)	3.66 (3.69)	4.26 (4.09)	4.57	4.78	4.89	4.98 (4.84)
	λ (nm)	339 (336)	291 (303)	272	260	253	249 (256)
	ϵ	0.071	0.225	0	0	0.038	0.170
Flavin	E_{exc} (eV)	3.00	3.17	3.32	3.80	3.92	4.04
	λ (nm)	413	391	374	326	316	307
	ϵ	0.202	0	0	0.150	0	0.012

tion, compared with the energetics used for adduct formation to DNA (350–400 nm, or even visible light). The reason for this difference lies in the surroundings – the π -cloud of the psoralen interacts with the π -systems of the neighboring nucleotides, which further reduces the excitation energies.

Interestingly, the flavin molecule is significantly easier to excite than the other heterocycles investigated. The first singlet excitation energy is only 3.0 eV (413 nm), and the probability of transition is intermediate. The second band with possible singlet excitation lies at 3.8 eV. Hence, if one intends to construct systems that are more readily excited, substituted flavins seem to be a more appropriate route than the furocoumarins, whereas the latter are easier to ionize.

4.5

Acknowledgments

The Swedish Science Research Council (VR) is gratefully acknowledged for financial support. We also acknowledge grants of computer time at the national supercomputing facility (NSC) in Linköping, and equipment grants from the Göran Gustafsson foundation and the Royal Physiographical Society in Lund.

4.6

References

- 1 PARR, R. G.; YANG, W., *Density Functional Theory of Atoms and Molecules*, Oxford University Press, NY, 1989.
- 2 GROSS, E. K. U.; DREIZLER, R. M., *Density Functional Theory*, Springer, 1992.
- 3 KOCH, W.; HOLTHAUSEN, M. C., *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, 2001.
- 4 *Theoretical and Computational Chemistry*, Vol. 2, *Modern Density Functional Theory – A Tool for Chemistry*, SEMINARIO, J. M.; POLITZER, P., (eds), Elsevier, Amsterdam, 1995.
- 5 CHONG, D. P. (ed.), *Recent Advances in Density Functional Methods*, Vols 1 and 2, World Scientific, Singapore, 1995.
- 6 LESZCZYNSKI, J. (ed.), *Theoretical and Computational Chemistry*, Vol. 8, *Computational Molecular Biology*, Elsevier, Amsterdam, 1999.
- 7 ERIKSSON, L. A. (ed.), *Theoretical and Computational Chemistry*, Vol. 9, *Theoretical Biochemistry – Processes and Properties of*

- Biological Systems*, Elsevier, Amsterdam, 2001.
- 8 THOMAS, L. H., *Proc. Camb. Phil. Soc.* **1927**, 23, 542.
 - 9 FERMI, E., *Z. Phys.* **1932**, 48, 73.
 - 10 DIRAC, P. A. M., *Proc. Camb. Phil. Soc.* **1930**, 26, 376.
 - 11 SLATER, J. C., *Phys. Rev.* **1951**, 81, 385.
 - 12 HOHENBERG, P.; KOHN, W., *Phys. Rev. B* **1964**, 136, 864.
 - 13 KOHN, W.; SHAM, L. J., *Phys. Rev. A* **1965**, 140, 1133.
 - 14 CEPERLEY, D. M.; ALDER, B. J., *Phys. Rev. Lett.* **1980**, 45, 566.
 - 15 VOSKO, S. J.; WILK, L.; NUSAIR, M., *Can. J. Phys.* **1980**, 58, 1200.
 - 16 BECKE, A. D., *Phys. Rev. A* **1988**, 38, 3098.
 - 17 PERDEW, J. P.; WANG, Y., *Phys. Rev. B* **1986**, 33, 8800.
 - 18 PERDEW, J. P.; WANG, Y., *Phys. Rev. B* **1991**, 44, 13298; *Phys. Rev. B* **1992**, 45, 13244.
 - 19 LEE, C.; YANG, W.; PARR, R. G., *Phys. Rev. B* **1988**, 37, 785.
 - 20 BECKE, A. D., *J. Chem. Phys.* **1993**, 98, 1372.
 - 21 BECKE, A. D., *J. Chem. Phys.* **1993**, 98, 5648.
 - 22 STEVENS, P. J.; DEVLIN, F. J.; CHABALOWSKI, C. F.; FRISCH, M. J., *J. Phys. Chem.* **1994**, 98, 11623.
 - 23 PERDEW, J. P.; BURKE, K.; ERNZERHOF, M., *Phys. Rev. Lett.* **1996**, 77, 3865.
 - 24 ADAMO, C.; BARONE, V., *J. Chem. Phys.* **1999**, 110, 6158.
 - 25 HAMPRECHT, F. A.; COHEN, A. J.; TOZER, D. J.; HANDY, N. C., *J. Chem. Phys.* **1998**, 109, 6264.
 - 26 PERDEW, J. P.; KURTH, S.; ZUPAN, A.; BLAHA, P., *Phys. Rev. Lett.* **1999**, 82, 2544.
 - 27 CASIDA, M. E., in Ref. [5].
 - 28 STRATMANN, R. E.; SCUSERIA, G. E.; FRISCH, M. J., *J. Chem. Phys.* **1998**, 109, 8218.
 - 29 KRISHNAN, R.; BINKLEY, J. S.; POPLE, J. A., *J. Chem. Phys.* **1980**, 72, 650; MCLEAN, A. D.; CHANDLER, G. S., *J. Chem. Phys.* **1980**, 72, 5639; FRISCH, M. J.; BINKLEY, J. S.; POPLE, J. A., *J. Chem. Phys.* **1984**, 80, 3265.
 - 30 MULHOLLAND, A. J., in Ref. [7].
 - 31 CAR, R.; PARRINELLO, M., *Phys. Rev. Lett.* **1985**, 55, 2471.
 - 32 COMESS, K. M.; LIPPARD, S. J., in: *Molecular Aspects of Anticancer Drug-DNA Interactions*; NEEDLE, S.; WARING, M., (eds), MacMillan, London, **1993**.
 - 33 LIPPARD, S. J., *Science* **1982**, 218, 1075.
 - 34 ANDERSON, K. S., *Mutation Res.* **1979**, 67, 209.
 - 35 CORDA, Y.; ANIN, M.-F.; LENG, M.; JOB, D., *Biochemistry* **1992**, 31, 1904.
 - 36 BRADLEY, L. J. N. et al., *Biochemistry* **1993**, 32, 982.
 - 37 LIPPARD, S. J. et al., *Nature* **1999**, 399, 708.
 - 38 ROSENBERG, B.; VAN CAMP, L.; KIRGAN, T., *Nature* **1965**, 205, 698.
 - 39 WANG, A. H. J.; YANG, D.; *Prog. Biophys. Mol. Biol.* **1996**, 66, 81, and references cited therein.
 - 40 BRABEC, V.; SIP, M.; LENG, M., *Biochemistry* **1993**, 32, 11676.
 - 41 FICHTINGER-SCHPEMAN, A. M. J. et al., *Biochemistry* **1985**, 24, 707.
 - 42 COSTE, F. et al., *Nucl. Acids Res.* **1999**, 27, 1837.
 - 43 LIPPARD, S. J.; GELASCO, A., *Biochemistry* **1998**, 37, 9230, and references cited therein.
 - 44 TAKAHARA, P. M. et al., *Nature* **1995**, 377, 649.
 - 45 HUANG, H. et al., *Science* **1995**, 270, 1842.
 - 46 PAQUET, F.; PEREZ, C.; LENG, M., *J. Biomol. Struct. Dyn.* **1996**, 14, 67.

- 47 WANG, A. H. J.; YANG, D., *Prog. Biophys. Mol. Biol.* **1996**, *66*, 104.
- 48 OHNDORF, U.-M.; ROULD, M. A.; HE, Q.; PABO, C. O.; LIPPARD, S. J., *Nature* **1999**, *399*, 708.
- 49 GELASCO, A.; LIPPARD, S. J., *Biochemistry* **1998**, *37*, 9230.
- 50 TAKAHARA, P. M.; FREDERICK, C. A.; LIPPARD, S. J., *J. Am. Chem. Soc.* **1996**, *118*, 12309.
- 51 KARTALOU, M.; SANSON, L. D.; ESSIGMAN, J. M., *Biochemistry* **2000**, *39*, 8032.
- 52 HE, Q.; OHNDORF, U.-M.; LIPPARD, S. J., *Biochemistry* **2000**, *39*, 14426.
- 53 DAVIES, M. S.; BERNERS-PRICE, S. J.; HAMBLEY, T. W., *Inorg. Chem.* **2000**, *39*, 5603.
- 54 LEGENDRE, F.; BAS, V.; KOZELKA, J.; CHOTTARD, J.-C., *Chem. Eur. J.* **2000**, *6*, 2002.
- 55 DAVIES, M. S.; BERNERS-PRICE, S. J.; HAMBLEY, T. W., *J. Am. Chem. Soc.* **1998**, *120*, 11380.
- 56 MILBURN, G. H. W.; TRUTER, M. R., *J. Chem. Soc. A* **1966**, 1609.
- 57 PERUMAREDDI, J. R.; ADAMSON, A. W., *J. Phys. Chem.* **1978**, *72*, 414.
- 58 STETSENKO, A. I.; SCI'DERKHANOVA, L. B., *Zh. Neorg. Khim.* **1981**, *25*, 164.
- 59 BOSE, R. N.; CORNELIUS, R. D.; VIOLA, R. D., *J. Am. Chem. Soc.* **1984**, *106*, 3336.
- 60 HINDMARSH, K.; HOUSE, D. A.; TURNBULL, M. M., *Inorg. Chim. Acta* **1997**, *257*, 11.
- 61 MILLER, S. E.; HOUSE, D. A., *Inorg. Chim. Acta* **1989**, *166*, 189.
- 62 DE BOLSTER, M. W. G.; CAMMACK, R.; COUCOUVANIS, D. N.; REEDIJK, J.; VEEGER, C., *J. Biol. Inorg. Chem.* **1996**, *1*, G1.
- 63 CARLONI, P.; SPRIK, M.; ANDREONI, W., *J. Phys. Chem. B* **2000**, *104*, 823.
- 64 ZHANG, Y.; GUO, Z.; YOU, X.-Z., *J. Am. Chem. Soc.* **2001**, *123*, 9378.
- 65 DUNNING JR, T. H.; HAY, P. J., in: *Modern Theoretical Chemistry*, SCHAEFFER, H. F., (ED.), PLENUM, NEW YORK, **1976**; HAY, P. J.; WADT, W. R., *J. Chem. Phys.* **1985**, *82*, 270; WADT, W. R.; HAY, P. J., *J. Chem. Phys.* **1985**, *82*, 284; HAY, P. J.; WADT, W. R., *J. Chem. Phys.* **1985**, *82*, 299.
- 66 MIERTUS, S.; SCROCCO, E.; TOMASI, J., *Chem. Phys.* **1981**, *55*, 117; MIERTUS, S.; TOMASI, J., *Chem. Phys.* **1982**, *65*, 239; COSI, M.; BARONE, V.; CAMMI, R.; TOMASI, J., *Chem. Phys. Lett.* **1996**, *255*, 327.
- 67 Gaussian 98, Revisions A.7–A.11, FRISCH, M. J. et al., Gaussian, Inc., Pittsburgh, PA, **1998**.
- 68 PARRISH, J. A.; STERN, P. S.; PATHAK, M. A.; FITZPATRICK, T. B., in: *The Science of Photomedicine*, REGAN, J. D.; PARRISH, J. A., (eds), Plenum, NY, **1982**.
- 69 PARRISH, J. A.; FITZPATRICK, T. B.; TANENBAUM, L.; PATHAK, M. A., *New Engl. J. Med.* **1974**, *291*, 1207.
- 70 MUSAJO, L.; RODIGHIERO, G.; CARPORALE, G.; DALL'ACQUA, F.; MARCIANI, S.; BORDIN, F.; BACCICAHETTI, F.; BEVILAQUA, R., in: *Sunlight and Man*, PATHAK, M. A. (ed.), University of Tokyo Press, Tokyo, **1974**.
- 71 EDELSON, R. et al., *New Engl. J. Med.* **1987**, *316*, 297.
- 72 GASPARRO, F. P.; GATTOLIN, P.; OLACK, G.; SUMPPIO, B. E.; DECKELBAUM, L. I., *Photochem. Photobiol.* **1993**, *57*, 100.

- 73 HARRIMAN, A., in: *CRC Handbook of Photochemistry and Photobiology*, CRC Press, Boca Raton, 1995.
- 74 EDELSON, R., *Yale J. Biol. Med.* **1989**, 62, 565.
- 75 ROOK, A.H. et al., *Arch. Dermatol.* **1992**, 128, 337.
- 76 TAYLOR, A.; GASPARRO, F.P., *Sem. Hematol.* **1992**, 29, 132.
- 77 KIRSEBOM, L., *Kemivärlden med. Kemisk Tidsskrift* **2001**, 26 (12).
- 78 HORNICECK, F.J. et al., *Photochem. Photobiophys.* **1985**, 9, 263.
- 79 LASKIN, J.D. et al., *Proc. Natl. Acad. Sci. USA* **1985**, 82, 6158.
- 80 CAFFIERI, S.; VEDALDI, D.; DAGA, A.; DALL'ACQUA, F., in: *Psoralens in 1988, Past, Present and Future*, FITZPATRICK, T.B.; FORLOT, P.; PATHAK, M.A.; URBACH, F. (eds), John Libbey Eurotext, Montrouge, France, **1988**.
- 81 CAFFIERI, S.; ZABRESKA, Z.; DALL'ACQUA, F., in: *Frontiers in Photobiology*, SHIMA, A.; ICHAHASCI, M.; FUJIWARA, Y.; TAKEBE, H., (eds), Excerpta Medica, Amsterdam, **1995**.
- 82 ANTONY, F.A.; LABODA, M.M.; DOWDY, J.C.; COSTLOW, M.E., *Photochem. Photobiol.* **1993**, 57, 25S.
- 83 MUSAJO, L.; RODRIGHIERO, G., in: *Photobiochemistry, Vol. III*, GIESE, A.C. (ed.), Academic Press, NY, **1972**.
- 84 DALL'ACQUA, F., in: *Research in Photobiology*, CASTELLANI, A., (ed.), Plenum Press, NY, **1977**.
- 85 MUSAJO, L.; BORDIN, F.; CAPORALE, G.; MARCIANI, S.; RIGATTI, G., *Photochem. Photobiol.* **1967**, 6, 711.
- 86 YOSHIKAWA, K.; MORI, N.; SAKAKIBARA, S.; MIZUNO, N.; SONG, P.-S., *Photochem. Photobiol.* **1979**, 29, 1127.
- 87 LASKIN, J.D.; LASKIN, D.L., in: *New Psoralen DNA Photobiology, Vol. II*, GASPARRO, F.P., (ED.), CRC PRESS, BOCA RATON, **1988**.
- 88 LASKIN, J.D.; LEE, E., *Biochem. Pharmacol.* **1991**, 41, 125.
- 89 CIMINO, G.P.; GAMPER, H.B.; ISAACS, S.T.; HEARST, J.E., *Annu. Rev. Biochem.* **1985**, 54, 1151.
- 90 HEARST, J.E., *Annu. Rev. Biophys. Bioeng.* **1985**, 10, 69.
- 91 AVERBECK, D., *Photochem. Photobiol.* **1989**, 50, 859.
- 92 BRIDGES, B.; STRAUSS, G., *Nature* **1980**, 283, 523.
- 93 SONG, P.S.; TAPLEY, K.J., *Photochem. Photobiol.* **1979**, 29, 1177.
- 94 WETMORE, S.D.; ERIKSSON, L.A.; BOYD, R.J., in Ref. [7].
- 95 GASPARRO, F.P., in: *CRC Handbook of Photochemistry and Photobiology*, CRC Press, Boca Raton, **1995**.
- 96 TESSMAN, J.W.; ISAACS, S.I.; HEARST, J.E., *Biochemistry* **1985**, 24, 1669.
- 97 RODIGHIERO, G.; DALL'ACQUA, F.; AVERBECK, D., in: *New Psoralen DNA Photobiology, Vol. I*, GASPARRO, F.P. (ed.), CRC Press, Boca Raton, **1988**.
- 98 KANNE, D.; STRAUB, K.; HEARST, J.E.; RAPOPORT, H., *J. Am. Chem. Soc.* **1982**, 104, 6754.
- 99 SHIM, S.C., in: *CRC Handbook of Photochemistry and Photobiology*, CRC Press, Boca Raton, **1995**.

QM/MM Approaches

5

***Ab Initio* Methods in the Study of Reaction Mechanisms – Their Role and Perspectives in Medicinal Chemistry**

MIKAEL PERÄKYLÄ

5.1 Introduction

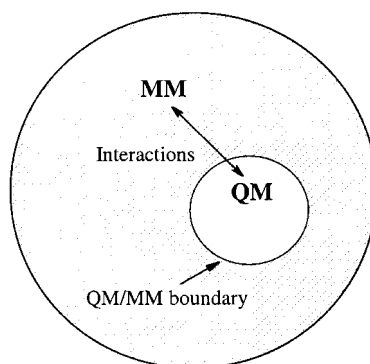
Ab initio quantum-mechanical (QM) methods can be used to predict with chemical accuracy, at least in principle, the structures, reactivities, and properties of almost all type of molecules. The applicability and accuracy of the QM methods is limited by the computing power available – high level *ab initio* QM calculations with the state-of-art electron correlation methods can be applied to systems with a relatively small number of atoms. This limitation has been relieved, and will be further relieved, by advances in computational methods, e.g. the advent in linear-scaling algorithms [1–5], local electron-correlation methods in molecular orbital-based methods [6, 7], and the introduction of methods based on density-functional theory to computational chemistry [8, 9]. Advances in computer technology, by providing more powerful processors and more efficient parallel computer architectures, also make accurate studies of larger molecular systems feasible. *Ab initio* QM methods are more or less routinely used to calculate optimized geometries and a variety of molecular properties of isolated molecules and molecular complexes. These data can be used to predict the physicochemical properties of molecules, explain the reactivities of the molecules in chemical reactions and to derive quantum-chemical descriptors [10] for quantitative structure-activity and structure-property relationship analysis. QM methods have also been widely used to investigate chemical reactions with isolated gas-phase systems. Many of the processes of interest of medicinal chemists, however, occur in the condensed phase or otherwise involve a large number species impor-

tant to the phenomenon studied, and so systems far too large to be studied rigorously with purely *ab initio* QM methods are needed to properly account for all the relevant interactions. Alternative approaches are needed for simulation of complex condensed-phase systems.

A number of models based on continuum representation of solvents have been developed to include the effects of solvation in QM calculations. These models reproduce the effects of solvation on molecular properties at least satisfactorily but, unfortunately, their use in the studies of chemical reactions is still limited. An interesting application of continuum solvation calculations relevant to medicinal chemistry is the estimation of pK_a values by *ab initio* QM methods. By combining accurate quantum mechanically calculated gas-phase basicities with solvation energies of the acid and the conjugate base a good correlation is obtained between calculated aqueous-phase proton affinities and experimental pK_a values [11–13]. With a precalculated correlation equation the pK_a values of the molecules of interest with unknown pK_a values can be estimated with good accuracy. Such calculations have been used to estimate the pK_a values of species during chemical reactions [14, 15] and to suggest an activation mechanism for the histamine H_3 -receptor [16].

To make QM studies of chemical reactions in the condensed phase computationally more feasible combined quantum mechanical/molecular mechanical (QM/MM) methods have been developed. The idea of combined QM/MM methods, introduced first by Levitt and Warshell [17] in 1976, is to divide the system into a part which is treated accurately by means of quantum mechanics and a part whose properties are approximated by use of QM methods (Fig. 5.1). Typically, QM methods are used to describe chemical processes in which bonds are broken and formed, or electron-transfer and excitation processes, which cannot be treated with MM methods. Combined QM and MM methods have been extensively used to study chemical reactions in solution and the mechanisms of enzyme-catalyzed reactions. When the system is partitioned into the QM and MM parts it is assumed that the process requiring QM treatment is localized in that region. The MM methods are then used to approximate the effects of the environment on the QM part of the system, which, via steric and electrostatic interactions, can be substantial. The

Fig. 5.1 Schematic representation of the division of a system into quantum mechanical (QM) and molecular mechanical (MM) parts.



noncovalent interactions between the QM and MM parts are treated by including the point charges of the MM atoms in the QM calculations and by employing molecular mechanical Lennard-Jones terms across the QM/MM boundary. When the interface of the QM and MM regions is between atoms connected by covalent bonds, special treatment of the QM/MM boundary is required.

The term QM/MM has been used in the literature to describe several methods using a combination of quantum mechanical and molecule mechanical methods; this has led to some confusion. Here the methods are classified, as suggested by Wang et al. [18], into two major subdomains – those that *structurally* couple the QM and MM calculations and those that *thermodynamically* couple them. In structurally coupled QM/MM the structures and energies of the QM and MM parts of the system are calculated simultaneously by use of a hybrid QM/MM potential of the whole system. In the thermodynamically coupled method the structures and the energies of the species on the reaction pathway are first calculated quantum mechanically. After that the energy difference between the species on the pathway, because of the environment, are determined in separate calculations using standard free-energy and MD simulation methods.

The empirical valence bond (EVB) method of Warshel [19] has features of both the structurally and thermodynamically coupled QM/MM method. In the EVB method the different states of the process studied are described in terms of relevant covalent and ionic resonance structures. The potential energy surface of the QM system is calibrated to reproduce the known experimental

data in solution. This calibrated EVB Hamiltonian is then used to simulate the reaction within a protein. More details about the EVB method can be found elsewhere [19]. The EVB method is computationally much less demanding than the structurally coupled QM/MM but it is less easy to use. The EVB potential must be parameterized for each different system and parameterization of complicated reaction mechanisms can be difficult. An attempt has recently been made to parameterize an EVB potential energy surface automatically by using gas-phase *ab initio* QM calculations to simulate solution and enzyme reactions [20]. The EVB method has been applied to an impressive number of enzyme reactions and has provided detailed information on the mechanisms of these enzymes and furthered our understanding of the nature of enzyme catalysis. Before structurally coupled QM/MM simulations became more popular, mainly as a result of the increase in computing power, the EVB method was almost the only way to perform realistic simulations of enzyme reactions. On the basis of on EVB calculations Warshel has suggested that electrostatic effects are the major factor in enzyme catalysis. These electrostatic effects are a consequence of the pre-organized polar environment of the enzyme active site and the relatively small reorganization free energy of this environment [21].

In this chapter the basic theory of the structurally coupled QM/MM is summarized. This is followed by some technical points important in the practical use of the method. In particular, details about the treatment of the QM/MM boundary are discussed. The thermodynamically coupled quantum mechanical/free energy (QM/FE) method is then introduced. Some representative applications of QM/MM methods are then described. The examples are selected to provide a representative picture of the potential applications of QM/MM methods on studies of reaction mechanisms. Here there is special emphasis on recent advances in the computational methodologies and in the future developments needed to improve the applicability of the methods.

5.2 Methods

5.2.1

Hybrid QM/MM Potential

The system studied by use of structurally coupled QM/MM is described with a Hamiltonian comprising of both the quantum mechanically and molecular mechanically treated part. The effective Hamiltonian, \hat{H}_{eff} , representing the interactions of the whole QM/MM system is:

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{QM/MM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{boundary}} \quad (1)$$

\hat{H}_{QM} is the Hamiltonian of the QM region and might be based on semiempirical, *ab initio* molecular orbital or density functional theory (DFT) methods. $\hat{H}_{\text{QM/MM}}$ represents the interactions between the QM and MM regions, \hat{H}_{MM} is the Hamiltonian of the purely MM region, and $\hat{H}_{\text{boundary}}$ is Hamiltonian for the boundary of the system, if this contribution is included. The corresponding total energy (E_{tot}) of the QM/MM system is:

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{QM/MM}} + E_{\text{MM}} + E_{\text{boundary}} \quad (2)$$

where E_{QM} is the quantum mechanical energy of the QM part, $E_{\text{QM/MM}}$ is the energy contribution arising from the interactions between the QM and MM parts of the system, E_{MM} is the molecular mechanical energy of the MM part of the system and E_{boundary} is the energy from the boundary taking into account the effects of bulk solvent.

$\hat{H}_{\text{QM/MM}}$ in Eq. (1) describes interactions between the QM and MM parts and has the general form:

$$\hat{H}_{\text{QM/MM}} = - \sum_{i,M} \frac{q_m}{r_{iM}} + \sum_{a,M} \frac{Z_a q_M}{R_{aM}} + \sum_{aM} \left\{ \frac{A_{aM}}{R_{aM}^{12}} - \frac{B_{aM}}{R_{aM}^6} \right\} + \hat{H}_{\text{QM/MM}} \quad (3)$$

(bonded int.)

where i represents the quantum mechanical electrons, a the quantum mechanical nuclei, and M the molecular mechanical atoms. The first term in Eq. (3) is the electrostatic interaction between an MM atom of charge q_M and electron i . This term is in-

incorporated into the QM Hamiltonian and, thus, the electronic structure of the QM part responds to the electrostatic field of the environment described with the MM force field. The second term is the electrostatic interaction between a charge on the MM atom and a charge on the nucleus of a QM atom. The third term describes the van der Waals interaction between a QM atom and an MM atom. The last term in Eq. (3) represents the bonded interactions (bond, angle, torsion angle, and improper torsion angle) across the QM/MM boundary. How bonded interactions are treated varies between the different QM/MM implementations. Below, the treatment of the QM/MM boundary with the link atom and hybrid orbital approaches is discussed.

It must be noted here that approaches have also been developed for study of molecular systems with QM methods of different accuracy for different parts of the system [22–25]. For example, a high-level QM method with electron correlation included might be used for the part of the system requiring highly accurate quantum mechanical description and lower level QM for atoms surrounding the high-level region. This type of method is particularly well suited for systems which can be treated completely with QM methods. In principle there are no limitations on which type of computational method – molecular mechanics, semiempirical QM, *ab initio* QM, or DFT methods – are mixed. The ONIOM-type methods [22, 23] developed by Morokuma and co-workers have been implemented in the recent version of the Gaussian 98 software [26]. In this implementation the system can be divided into two or three layers treated with different QM methods or molecular mechanically by use of AMBER [27], DREIDING [28], or UFF [29] force fields. This method has been developed further and applied to protein systems [30]. Woo et al. [24] have implemented this methodology for performing molecular dynamics (MD) simulations with hybrid QM/MM potential.

5.2.2

QM/MM Boundary – The Link Atom Approach

When QM/MM methods are used to study chemical reactions in solution and the reacting species are small enough to be treated completely with QM methods it is straightforward to separate

the system into a quantum mechanical solute and molecular mechanically treated solvent. Under these conditions there are only noncovalent interactions between the QM and MM parts. The noncovalent QM/MM boundary is simply treated by employing nonbonding electrostatic and Lennard-Jones interactions (Eq. 3). This means that molecular mechanical data for calculation of Lennard-Jones interactions must be assigned for the QM atoms. The parameters of the MM force fields are not, in general, transferable for QM atoms but must be reparameterized to reproduce correct interaction energies for noncovalent interactions across the QM/MM boundary. These parameters also depend on the force field and QM method in question [31–33]. It has been observed that it is necessary to increase the van der Waals radii of QM atoms by 5–10% [34]. This increase is necessary mainly to keep the QM region from being overly attracted by the MM charges, because the MM parameters result in insufficient Pauli repulsion of the QM region [34]. A related problem is the assignment of Lennard-Jones parameters to QM atoms, the chemical nature of which changes substantially during the reaction studied by QM/MM methods. In such circumstances, at least in principle, these parameters should also be changing during the reaction. The importance and consequences of this point have not so far been investigated in detail but it is likely it plays a minor role in determining the accuracy of QM/MM calculations. It is, however, advisable to construct the QM part in such a way that, if possible, the MM atoms are not in direct contact with reacting atoms [34].

When a biomolecular system is separated into QM and MM regions one must usually cut amino acid side chains or the protein backbone at covalent bonds (Fig. 5.2a). The construction of the covalent boundary between the QM and MM parts is key to accurate results from QM/MM calculations. Because there is no unique way to treat the covalent boundary, several different approaches have been described. In the first applications of coupled QM/MM simulations link atoms were used to create the covalent QM/MM boundary (Fig. 5.2b). Link atoms are atoms added to the QM part to fill the broken valences of the boundary QM atoms. These atoms are placed along the broken QM/MM bond at a distance appropriate for the QM bond added. The link atoms have usually been hydrogen atoms but methyl groups and pseudohalogen atoms have also been used [35].

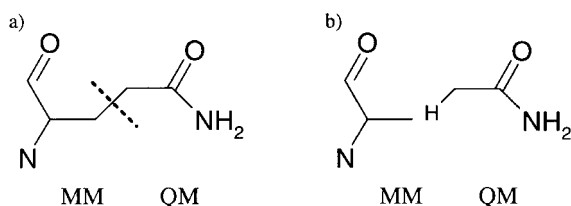


Fig. 5.2 (a) Division of an amino acid residue (glutamine) into a QM and MM part at a covalent bond. (b) Use of a hydrogen link atom to fill the empty valence of the boundary QM atom.

Although results from QM/MM studies have been observed to depend on the placement and treatment of the link atom [36, 37], it has been the most common way to treat the QM/MM boundary. The link atom approach seems to provide satisfactory results when semiempirical QM methods are applied. The link atoms are necessary for proper treatment of the QM part but, because they are not present in the original QM/MM system, their interactions with the MM part are not chemically defined. This requires a subjective decision of how the extra interactions between the atoms at the QM/MM boundary are treated. Several different ways of accounting for the interactions between the boundary atoms and the rest of the system have been suggested. In the first implementations of the QM/MM method the interactions between the link atom and the MM atoms were not included, because the link atoms do not exist in the original system [38, 39]. This approach has, however, been found to result in large polarization of the QM part and an unrealistically large partial charge on the hydrogen link atom. In some implementations the electrostatic interactions between the link atoms and first MM groups (i.e. the groups of MM atoms covalently bound to the QM atoms) or all MM atoms were removed or the QM part did not feel the influence of the first MM groups [36, 40]. In the charge shift method the link atom interacts with the MM atoms but the charge on the MM atom replaced by the link atom is shifted to the neighboring MM atoms directly bonded to this atom. In addition, dipoles are introduced to the neighboring atoms to compensate for this charge shift [41]. Specific parameterized scaling factors for close-lying electrostatic interactions around the boundary bonds have also been introduced [42]. The test calculations of Reuter et al. [35] showed that the

use of link atoms which interact with the MM charges yields QM/MM energies in good agreement with full QM calculations. Additional constraints were, however, required to maintain the proper geometry of the covalent bonds between the QM and MM regions.

5.2.3

QM/MM Boundary – The Hybrid Orbital Approach

To circumvent problems associated with the link atoms different approaches have been developed in which localized orbitals are added to model the bond between the QM and MM regions. Warshel and Levitt [17] were the first to suggest the use of localized orbitals in QM/MM studies. In the local self-consistent field (LSCF) method the QM/MM frontier bond is described with a strictly localized orbital, also called a frozen orbital [43]. These frozen orbitals are parameterized by use of small model molecules and are kept constant in the SCF calculation. The frozen orbitals, and the localized orbital methods in general, must be parameterized for each quantum mechanical model (i.e. energy-calculation method and basis set) to achieve reliable treatment of the boundary [34]. This restriction is partly circumvented in the generalized hybrid orbital (GHO) method [44]. In this method, which is an extension of the LSCF method, the boundary MM atom is described by four hybrid orbitals. The three hybrid orbitals that would be attached to other MM atoms are fixed. The remaining hybrid orbital, which represents the bond to a QM atom, participates in the SCF calculation of the QM part. In contrast with LSCF approach the added flexibility of the optimized hybrid orbital means that no specific parameterization of this orbital is needed for each new system.

Murphy et al. [34, 45] have parameterized and extensively tested a QM/MM approach utilizing the frozen orbital method at the HF/6-31G* and B3LYP/6-31G* levels for amino acid side chains. They parameterized the van der Waals parameters of the QM atoms and molecular mechanical bond, angle and torsion angle parameters (Eq. 3, $\hat{H}_{\text{QM/MM}}$ (bonded int.)) acting across the covalent QM/MM boundary. High-level QM calculations were used as a reference in the parameterization and the molecular mechanical calculations were performed with the OPLS-AA force

field of Jorgensen et al. [46]. This QM/MM method has been implemented in the commercial Jaguar program (Schrödinger). In this program the energy of the bulk solvent, E_{boundary} (Eq. 2), can be taken into account within the context of the continuum solvation method based on Poisson-Boltzmann equations. Single-point and geometry optimization calculations of the QM/MM system are possible with E_{boundary} included. This QM/MM method has been used to model cytochrome P-450 chemistry by use of a system with 125 quantum mechanical active-site atoms and 6950 molecular mechanical atoms. QM calculations were performed at the B3LYP/6-31G* level [34].

Reuter et al. [37] have conducted a detailed comparison of the accuracy of the LSCF and link hydrogen atom approaches using the CHARMM program [47]. In their work the QM calculations were performed with the semiempirical AM1 method. They found that the choice of electronic interactions included in the frontier region is of considerable importance in determining the electron distribution of the QM region and the overall energy. The link atom and LSCF methods were found to be of generally similar accuracy if care was taken in the choice of the QM/MM boundary. It was observed that charged atoms in proximity to QM atoms are a potential source of error and, therefore, it is advisable to select classical frontier atoms with small charges. Reuter et al. [37] also raised several practical points concerning the handling of the QM/MM frontier. Hall et al. [48] investigated the performance of a variety of link atom schemes and LSCF methods at the HF/6-31G* level. They found that LSCF method gives the best agreement with the fully QM-optimized structures and, if additional constraints are used, the link atom approach also works reliably.

5.3

Thermodynamically Coupled QM/MM

Jorgensen et al. [49] developed a combined quantum mechanical and molecular mechanical approach for study of organic reactions and applied it with success to many solution reactions. Inspired by this Kollman et al. [50, 51] developed the approach further for study of enzyme reactions. This quantum mechanical/free energy (QM/

FE) method is a thermodynamically coupled QM/MM approach in which the structures, energies and charge distributions of the species on the reaction pathway are first determined in the gas phase by means of QM calculations. The QM model is isolated from the enzyme by replacing the missing valences by link hydrogen atoms and optimized in the gas phase in the orientation the species of the QM part have in the enzyme. The gas-phase-optimized structures are then solvated, to calculate the reference reaction in water, or inserted into an enzyme. Free energy differences for the species on the reaction pathway as a result of the solvent or protein environment are then calculated by use of standard free-energy methods. The difference between the free energy of two structures on the reaction pathway (ΔG) is:

$$\Delta G = \Delta E_{\text{QM}} + \Delta G_{\text{MM}} - T\Delta S_{\text{QM}}$$

where ΔE_{QM} is the gas-phase energy difference of the QM part, ΔG_{MM} is the free energy difference of the QM species because of interactions with the environment and $T\Delta S_{\text{QM}}$ is the entropy change of the QM part. An obvious limitation of this method is that the QM part has a gas-phase structure and charge distribution. Despite these limitations, however, the QM/FE method has found interesting applications in the study of enzyme function [50, 52–54]. Because, in the QM/FE method, the role of the environment on the energetics of the reaction is calculated with standard free energy perturbation (FEP) and molecular dynamics methods, all MD simulation techniques (e.g. periodic box and the particle mesh Ewald for long-range electrostatic interactions) can be applied to sample the conformational space of the environment and to include the effects of long-range electrostatic interactions. When the thermodynamic integration method is used in FEP simulations, free energy component analysis enables qualitative estimation of the effects of protein and solvent groups on the reaction free energy [52]. QM/FE investigation of the enzymatic reaction catalyzed by citrate synthase showed that the energetics of the enzyme reaction are sensitive to proper treatment of long-range electrostatic interactions [53]. This emphasizes the importance of electrostatic interactions in the catalytic efficiency of enzymes, as suggested first by Warshel on the basis of EVB sim-

ulations [19]. It also shows the importance of proper treatment of electrostatic interactions in QM/MM simulations of biomolecules.

In the structurally coupled QM/MM implementation of Zhang et al. [55, 56], in which the QM/MM boundary was treated by use of the pseudobond approach [55, 57], the QM/MM minimization of the QM part is combined with FEP calculations. In this procedure the energy profile of the enzyme reaction is first determined by use of QM/MM energy minimizations. The structures and charges of the QM atoms are then used, in the same manner as in the QM/FE method, to determine the role of environment on the energy profile of the reaction. In this way the effects of a large number of MM conformations of protein and solvent environment can be included in the total energies.

5.4

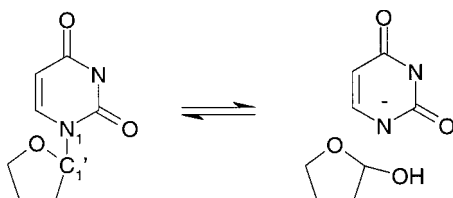
Selected Applications of QM/MM Methods

5.4.1

Uracil-DNA Glycosylase

The QM/MM study of Dinner et al. [58] on uracil-DNA glycosylase, UDG, is a good example of detailed work providing insight into the catalytic mechanism of an enzyme. Uracil-DNA glycosylase excises uracil from DNA, where they are either a result of misincorporation of deoxyuridine (dU) or deamination of cytosine. In this work the QM region was treated by use of the AM1 method. A limited number of explicit water molecules were included in the model to achieve a reasonable set of structures and the effect of bulk solvent was modeled implicitly by use of continuum electrostatic methods based on the Poisson equation. Link hydrogen atoms, which interacted with all MM atoms, were used to terminate the QM region. In the first step of the enzyme-catalyzed reaction studied in this work the glycosylic bond of dU in DNA is hydrolyzed (Fig. 5.3). In contrast with the concerted associative mechanism proposed initially, the calculations showed that UDG catalyzes base excision by a stepwise dissociative mechanism – in the reaction the C_1-N_1 bond is broken first followed by attack of a water molecule on C_1' and formation of

Fig. 5.3 Reaction catalyzed by uracil-DNA glycosylase.



the product. The calculated activation energy for the reaction was 62.3 kJ mol^{-1} . The corresponding solution reaction, which occurs by a concerted mechanism, has a calculated reaction barrier of $144.3 \text{ kJ mol}^{-1}$. These numbers are in good agreement with the corresponding measured activation enthalpies of 50.6 kJ mol^{-1} and $134.3 \text{ kJ mol}^{-1}$, respectively.

Note that in this reaction the mechanism is different in enzyme and in water. The role of electrostatic interactions of individual DNA and protein groups on the reaction transition-state energy was calculated by cumulatively setting to zero the charges of each MM group (amino acid residues, water molecules, and nucleotides) in order of decreasing distance from the QM region. This type of analysis had been used earlier to reveal the role of different parts of the environment on enzyme reaction energies [59–61]. It must be noted that although the contributions of enzyme groups calculated in this way are not quantitative predictions of site-directed mutations of the same groups, they provide interesting information about the origin of enzyme catalysis. For UDG most of the rate acceleration comes from stabilization of the oxocarbenium cation of the reaction transition state by four phosphate groups of the target DNA. The calculations in this work were performed with the CHARMM program [47]. In the QM/MM implementation of CHARMM, QM calculations can be performed at the semiempirical (the MOPAC program) or *ab initio* QM levels (the GAMESS and CADPAC programs) [62].

5.4.2

QM/MM Simulations of Quantum Effects

Experiments have implied that quantum mechanical effects can be involved in the proton and hydride transfer of several biological processes. Specifically, large kinetic isotope effects (KIE) observed for

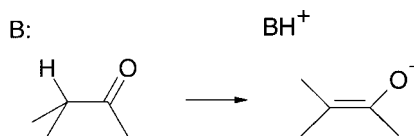
several proton and hydride transfers have suggested that tunneling is significant in the catalytic mechanisms of several enzymes [63]. Elucidation of these experimental observations by QM/MM simulations has quite recently stimulated the implementation of several new simulation algorithms including quantum effects in enzyme reaction studies. Hwang and Warshel [64] evaluated the deuterium isotope effect for the proton transfer step in the catalytic mechanism of carbonic anhydrase, using the EVB method and an approach based on path integral algorithm. The calculated isotope effect was in very good agreement with that observed. The path integral method has also been used to study proton transfer in the reaction catalyzed by flu virus neuraminidase [65]. Approaches based on the canonical variational transition-state theory (TST) combined with the hybrid QM/MM potential energy surface have been used quite recently to study proton tunneling in enzyme systems. Reaction rates and kinetic isotope effects of proton or hydride transfer reactions catalyzed by enolase [66], liver alcohol dehydrogenase [67], and methylamine dehydrogenase [68, 69] have also been simulated. Because the energy surface of the enzyme reaction must be mapped in detail in this kind of study, and a large number of second derivatives must be calculated, semiempirical QM Hamiltonians have been used. The calculations have been able to reproduce the KIE in good agreement with experimental results, indicating the usefulness of the simulations for studying quantum effects. These calculations have provided a detailed atomic-level picture of the factors playing a role in tunneling in biological systems. They have also highlighted the importance of thermal fluctuations and the flexibility of the whole protein in enhanced hydrogen tunneling and on the classical reaction free-energy barrier. This kind of calculation might further our understanding of the role of protein motion in enzyme catalysis – one of the most interesting problems in the nature of enzyme catalysis not yet fully understood.

5.4.3

Miscellaneous Applications

Enzymes which catalyze proton abstraction from a carbon activated by carbonyl or carboxyl group adjacent to the carbon have been of great interest to experimentalists and theoreticians

Fig. 5.4 Abstraction of an α -proton from a carbon atom activated by a carbonyl group.



(Fig. 5.4). In recent years the mechanisms of enzyme reactions catalyzed by citrate synthase [53, 70, 71], enolase [56], mandelate racemase [72], and triosephosphate isomerase [41, 73] have been determined by use of QM/MM methods. In these studies different QM/MM implementations and slightly different calculation procedures and QM Hamiltonians (semiempirical, *ab initio* MO and DFT methods) have been used. It is interesting to note that although a variety of methods has been used, the results reported have, in general, been in good agreement with each other. The nature of the reaction intermediate after abstraction of the proton and the possible existence of a transition-state-stabilizing, low-barrier hydrogen bond (LBHB) between substrate and hydrogen-bond donor of the enzyme active site, in particular, have been the subject of debate. It has been suggested that LBHBs are important because they reduce the barrier of enzyme reactions [74, 75]. In the QM/MM calculations performed so far no support has been found for the existence of LBHB in this type of reaction. Results from these studies have implied that the catalytic power of the enzymes catalyzing proton abstraction might be because of stabilization of the reaction transition state by the pre-organized electrostatic interactions of the enzyme. It is, however, not impossible that more detailed theoretical studies, which, for example, include quantum effects of the proton involved in the LBHB, could change the picture.

Chorismate mutase catalyzes the Claisen rearrangement of chorismate to prephenate at a rate 10^6 times greater than that in solution (Fig. 5.5). This enzyme reaction has attracted the attention of computational (bio)chemists, because it is a rare example of an enzyme-catalyzed pericyclic reaction. Several research groups have studied the mechanism of this enzyme by use of QM/MM methods [76–78]. It has also been studied with the effective fragment potential (EFP) method [79, 80]. In this method the chemically active part of an enzyme is treated by use of the *ab initio* QM method and the rest of the system (protein environment) by effective fragment potentials. These potentials account

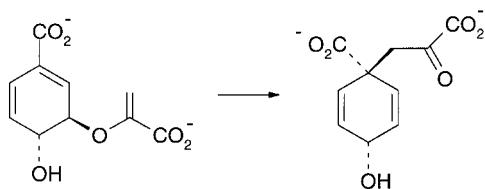


Fig. 5.5 Chorismate mutase catalyzes the rearrangement of chorismate to prephenate.

for the non-bonding interactions of the protein environment for the QM region. Guo et al. [81] have used QM/MM molecular dynamics simulations to account for the motion in the chorismate mutase-substrate complex. In this work a fast semiempirical self-consistent charge density functional tight-binding method [82] was used for QM description of the substrate. The main conclusion, that the catalytic power of chorismate mutase is largely a result of the binding of the reactant, chorismate, in a reactive pseudodiaxial conformation, was reached in all QM/MM studies. The pseudodiaxial conformation is structurally close to the reaction transition state. Guo et al. [81] further showed that chorismate mutase is capable of binding non-reactive conformations of chorismate, which in solution are considerably more stable than the reactive conformation, and transform the bound substrate into a reactive pseudodiaxial conformation. This work illustrates the potential of molecular dynamics in providing information on functional properties of enzymes that is beyond the scope of QM/MM minimization of stationary points on the reaction pathways and the role protein motion might have in enzyme catalysis.

A more complete list of early applications of QM/MM methods to enzymatic reactions can be found elsewhere [18, 35, 83, 84]. Gao [85] has reviewed QM/MM studies of a variety of solution phenomena. QM/MM methods have also been used to study the spectra of small molecules in different solvents [86] and electrochemical properties of photosynthetic reaction centers within a protein environment [87–89]. An approach has also been developed for calculation of NMR shielding tensors by use of a QM/MM method [90].

5.5

Conclusions

The applications of *ab initio* QM methods to complex chemical reactions presented in this chapter show that simulation techniques have developed to a level at which the accurate investigations in realistic condensed-phase models are possible. The studies have, however, revealed several caveats in simulation techniques, for example the importance of properly including long-range electrostatic interactions and structural changes of the protein environment. Because computational algorithms and computing hardware continues to improve, computer simulations will become integral part of medicinal chemistry research. Especially when theoretical studies are combined with experiments new insights are obtained from biological functions. It is likely that not in a too distant future we will be able to conduct genuine computer experiments. Thus, in the present era of genomics and proteomics, computational experiments on biological systems will be among the tools exploited to try to unravel the secret of life at the atomic and molecular level.

5.6

References

- 1 YANG, W., LEE, T.-S., *J. Chem. Phys.* **1995**, *103*, 5674–5678.
- 2 YORK, D.M., LEE, T.-S., YANG, W., *J. Am. Chem. Soc.* **1996**, *118*, 10940–10941.
- 3 NADIG, G., VAN ZANT, L. C., DIXON, S. L., MERZ, K. M., *J. Am. Chem. Soc.* **1998**, *120*, 5593–5594.
- 4 GOEDECKER, S., *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- 5 VAN DER VAART, A., GOGONEA, V., DIXON, S. L., MERZ, K. M., *J. Comput. Chem.* **2000**, *16*, 1494–1504.
- 6 MURPHY, R. B., BEACHY, M. D., FRIESNER, R. A., RINGNALDA, M. N., *J. Chem. Phys.* **1995**, *103*, 1481–1490.
- 7 SAEBO, S., PULAY, P., *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- 8 KOCH, W., HOLTHAUSEN, M. C., *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, New York, **2000**.
- 9 FRIESNER, R. A., DUNIETZ, B. D., *Acc. Chem. Res.* **2001**, *34*, 351–358.
- 10 KARELSON, M., LOBANOV, V. S., KATRITZKY, A. R., *Chem. Rev.* **1996**, *96*, 1027–1043.
- 11 PERÄKYLÄ, M., *J. Org. Chem.* **1996**, *61*, 7420–7425.
- 12 JANG, Y.H., SOWERS, L. C., CAGIN, T., GODDARD, W.A. III, *J. Phys. Chem. A* **2001**, *105*, 274–280.
- 13 LIPTAK, M. D., SHIELDS, G. C., *J. Am. Chem. Soc.* **2001**, *123*, 7314–7319.

- 14 PERÄKYLÄ, M., *J. Am. Chem. Soc.* **1998**, *120*, 12895–12902.
- 15 PERÄKYLÄ, M., *Phys. Chem. Chem. Phys.* **1999**, *1*, 5643–5647.
- 16 KOVALAINEN, J. T., CHRISTIAANS, J. A. M., ROPPONEN, R., POSO, A., PERÄKYLÄ, M., VEPSÄLÄINEN, J., LAATIKAINEN, R., GYNTHNER, J., *J. Am. Chem. Soc.* **2000**, *122*, 6989–6996.
- 17 WARSHHEL, A., LEVITT, M., *J. Mol. Biol.* **1976**, *103*, 227–249.
- 18 WANG, W., DONINI, O., REYES, C. M., KOLLMAN, P. A., *Annu. Rev. Biomol. Struct.* **2001**, *30*, 211–243.
- 19 WARSHHEL, A., *Computer Modeling of Chemical Reactions in Enzymes*, Wiley, New York, **1991**.
- 20 BENTZIE, J., MULLER, R. P., FLORIÁN, J., WARSHHEL, A., *J. Phys. Chem. B* **1998**, *102*, 2293–2301.
- 21 VILLÁ, J., ŠTRAJBL, M., GLENNO, T. M., SHAM, Y. Y., CHU, Z. T., WARSHHEL, A., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11899–11904.
- 22 HUMBEL, S., SIEBER, S., MOROKUMA, K., *J. Chem. Phys.* **1996**, *105*, 1959–1967.
- 23 SVENSSON, M., HUMBEL, S., FROESE, R. D. J., MATSUBARA, S., SIEBER, S., MOROKUMA, K., *J. Phys. Chem.* **1996**, *100*, 19357–19363.
- 24 WOO, T. K., CAVALLO, L., ZIEGLER, T., *Theor. Chim. Acc.* **1998**, *100*, 307–313.
- 25 GOGONEA, V., WESTERHOFF, L. M., MERZ, K. M., *J. Chem. Phys.* **2000**, *113*, 5604–5613.
- 26 *Gaussian* 98, FRISCH, M. J., TRUCKS, G. W., SCHLEGEL, H. B., SCUSERIA, G. E., ROBB, M. A., CHEESEMAN, J. R., ZAKRZEWSKI, V. G., MONTGOMERY, J. A., STRATMANN, R. E., BURANT, J. C., DAPPRICH, S., MILLAM, J. M., DANIELS, A. D., KUDIN, K. N., STRAIN, M. C., FARKAS, O., TOMASI, J., BARONE, V., COSSI, M., CAMMI, R., MENNUCCI, B., POMELLI, C., ADAMO, C., CLIFFORD, S., OCHTERSKI, J., PETERSON, G. A., AYALA, P. Y., CUI, G., MOROKUMA, K., MALICK, D. K., RABUCK, A. D., RAGHAVACHARI, K., FORESMAN, J. B., CIOŚLÓWSKI, J., ORTIZ, J. V., STEFANOV, B. B., LIU, G., LIASHENKO, A., PISKORZ, P., KOMAROMI, I., GOMPERS, R., MARTIN, R. L., FOX, D. J., KEITH, T., AL-LAHAM, M. A., PENG, C. Y., NANAYAKKARA, A., GONZALEZ, C., CHAL-LACOMBE, M., GILL, P. M. W., JOHNSON, B. G., CHEN, W., WONG, M. W., ANDRES, J. L., HEAD-GORDON, M., REPLOGLE, E. S., POPLE, J. A., Gaussian Inc., Pittsburgh, PA, **1998**.
- 27 CORNELL, W., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., SPELLMEYER, D. C., FOX, T., CALDWELL, J. W., KOLLMAN, P. A., *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- 28 MAYO, S. L., OLAFSON, B. D., GODDARD, W. A., *J. Phys. Chem.* **1990**, *94*, 8897–8906.
- 29 RAPPÉ, A. K., CASEWIT, C. J., COLWELL, K. S., GODDARD, W. A., SKIFF, W. M., *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- 30 TORRENT, M., VREVEN, T., MUSAEV, D. G., MOROKUMA, K., FARKAS, Ö., SCHLEGEL, H. B., *J. Am. Chem. Soc.* **2002**, *124*, 192–193.
- 31 STANTON, R. V., HARTSOUGH, D. S., MERZ, K. M., *J. Comput. Chem.* **1995**, *16*, 113–128.
- 32 FREINDORF, M., GAO, J., *J. Comput. Chem.* **1996**, *17*, 386–395.
- 33 LUQUE, F. J., REUTER, N., CARTIER, A., RUIZ-LÓPEZ, M. F., *J.*

- Phys. Chem. A* **2000**, *104*, 10923–10931.
- 34 MURPHY, R. B., PHILIPP, D. M., FRIESNER, R. A., *J. Comput. Chem.* **2000**, *21*, 1442–1457.
- 35 FIELD, M. J., *J. Comput. Chem.* **2002**, *23*, 48–58.
- 36 EURENIUS, K. P., CHATFIELD, D. C., BROOKS, B. R., HODOSCEK, M., *Int. J. Quantum Chem.* **1996**, *60*, 1189–1200.
- 37 REUTER, N., DEJAEGERE, A., MAIGRET, B., KARPLUS, M., *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- 38 SINGH, U. C., KOLLMAN, P. A., *J. Comput. Chem.* **1986**, *7*, 718–730; STANTON, R. V., HARTSOUGH, D. S., MERZ, K. M., *J. Comput. Chem.* **1995**, *16*, 113–128.
- 39 FIELD, M. J., BASH, P. A., KARPLUS, M., *J. Comput. Chem.* **1990**, *11*, 700–733.
- 40 VASILYEV, V. V., *J. Mol. Struct. (Theochem.)* **1994**, *304*, 129–141.
- 41 LENNARTZ, C., SCHÄFER, A., TERSTEGEN, F., THIEL, W., *J. Phys. Chem. B* **2002**, *106*, 1758–1767.
- 42 PHILIPP, D. M., FRIESNER, R. A., *J. Comput. Chem.* **1999**, *20*, 1468–1494.
- 43 THÉRY, V., RINALDI, D., RIVAIL, J.-L., MAIGRET, B., FERENCZY, G. G., *J. Comput. Chem.* **1994**, *15*, 269–282.
- 44 GAO, J. L., AMARA, P., ALHAMBRA, C., FIELD, M. J., *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- 45 MURPHY, R. B., PHILIPP, D. M., FRIESNER, R. A., *Chem. Phys. Lett.* **2000**, *321*, 113–120.
- 46 JORGENSEN, W. L., MAXWELL, D. S., TIRADO-RIVES, J., *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- 47 BROOKS, B. R., BRUCCOLERI, R. E., OLAFSON, B. D., STATES, D. J., SWAMINATHAN, S., KARPLUS, M., *J. Comput. Chem.* **1983**, *4*, 187–217.
- 48 HALL, R. J., HINDLE, S. A., BURTON, N. A., HILLIER, I. H., *J. Comput. Chem.* **2000**, *21*, 1433–1441.
- 49 JORGENSEN, W. L., *Acc. Chem. Res.* **1989**, *22*, 184–189.
- 50 STANTON, R. V., PERÄKYLÄ, M., BAKOWIES, D., KOLLMAN, P. A., *J. Am. Chem. Soc.* **1998**, *120*, 3448–3457.
- 51 KOLLMAN, P. A., KUHN, B., DONINI, O., PERÄKYLÄ, M., STANTON, R. V., BAKOWIES, D., *Acc. Chem. Res.* **2001**, *34*, 72–79.
- 52 PERÄKYLÄ, M., KOLLMAN, P. A., *J. Am. Chem. Soc.* **2000**, *122*, 3436–3444.
- 53 DONINI, O., DARDEN, T., KOLLMAN, P. A., *J. Am. Chem. Soc.* **2000**, *122*, 12270–12280.
- 54 KUHN, B., KOLLMAN, P. A., *J. Am. Chem. Soc.* **2000**, *122*, 3909–3916.
- 55 ZHANG, Y., LIU, H., YANG, W., *J. Chem. Phys.* **2000**, *112*, 3483–3492.
- 56 LIU, H., ZHANG, Y., YANG, W., *J. Am. Chem. Soc.* **2000**, *122*, 6560–6570.
- 57 ZHANG, Y., LEE, T.-S., YANG, W., *J. Chem. Phys.* **1999**, *110*, 46–54.
- 58 DINNER, A., BLACKBURN, G. M., KARPLUS, M., *Nature* **2001**, *413*, 752–755.
- 59 BASH, P. A., FIELD, M. J., DAVENPORT, R. C., PETSKO, G. A., RINGE, D., KARPLUS, M., *Biochemistry* **1991**, *30*, 5826–5832.
- 60 LYNE, P. D., MULHOLLAND, A. J., RICHARDS, W. G., *J. Am. Chem. Soc.* **1995**, *117*, 11345–11350.
- 61 PERÄKYLÄ, M., KOLLMAN, P. A., *J. Am. Chem. Soc.* **1997**, *119*, 1189–1196.
- 62 LYNE, P. D., HODOSCEK, M., KARPLUS, M., *J. Phys. Chem. A* **1999**, *103*, 3462–3471.

- 63 JONSSON, T., GLICKMAN, M. H., SUN, S., KLINMAN, J. P., *J. Am. Chem. Soc.* **1996**, *118*, 10319–10320.
- 64 HWANG, J.-K., WARSHEL, A., *J. Am. Chem. Soc.* **1996**, *118*, 11745–11751.
- 65 THOMAS, A., JOURAND, D., BRET, C., AMARA, P., FIELD, M. J., *J. Am. Chem. Soc.* **1999**, *121*, 9693–9702.
- 66 ALHAMBRA, C., GAO, J., CORCHADO, J. C., VILLÁ, J., TRUHLAR, D. G., *J. Am. Chem. Soc.* **1999**, *121*, 2253–2258.
- 67 ALHAMBRA, C., CORCHADO, J. C., SÁNCHEZ, M. L., GAO, J., TRUHLAR, D. G., *J. Am. Chem. Soc.* **2000**, *122*, 8197–8203.
- 68 ALHAMBRA, C., SÁNCHEZ, M. L., CORCHADO, J., GAO, J., TRUHLAR, D. G., *Chem. Phys. Lett.* **2001**, *347*, 512–518.
- 69 FAULDER, P. F., TRESADERN, G., CHOHAN, K. K., SCRUTTON, N. S., SUTCLIFFE, M. J., HILLIER, I. H., BURTON, N. A., *J. Am. Chem. Soc.* **2001**, *123*, 8604–8605.
- 70 MULHOLLAND, A. J., RICHARDS, W. G., *Proteins* **1997**, *27*, 9–25.
- 71 MULHOLLAND, A. J., LYNE, P. D., KARPLUS, M., *J. Am. Chem. Soc.* **2000**, *122*, 534–535.
- 72 GARCIA-VILOCA, M., GONZÁLES-LAFONT, Á., LLUCH, J. M., *J. Am. Chem. Soc.* **2001**, *123*, 709–721.
- 73 CUI, Q., KARPLUS, M., *J. Am. Chem. Soc.* **2001**, *123*, 2284–2290.
- 74 GERLT, J. A., GASSMAN, P. G., *Biochemistry* **1993**, *32*, 11943–11952.
- 75 FREY, P. A., WHITT, S. A., TOBIN, J. B., *Science* **1994**, *264*, 1927–1930.
- 76 DAVIDSON, M. M., GOULD, I. R., HILLIER, I. H., *J. Chem. Soc. Chem. Commun.* **1995**, 63–64.
- 77 MARTÍ, S., ANDRÉS, J., MOLINER, V., SILLA, E., TUÑÓN, I., BERTRÁN, J., *J. Phys. Chem. B* **2000**, *104*, 11308–11315.
- 78 MARTÍ, S., ANDRÉS, J., MOLINER, V., SILLA, E., TUÑÓN, I., BERTRÁN, J., *J. Am. Chem. Soc.* **2001**, *123*, 1709–1712.
- 79 WORTHINGTON, S. E., KRAUSS, M., *J. Phys. Chem. B* **2001**, *105*, 7096–7098.
- 80 WORTHINGTON, S. E., ROITBERG, A. E., KRAUSS, M., *J. Phys. Chem. B* **2001**, *105*, 7087–7095.
- 81 GUO, H., CUI, Q., LIPSCOMB, W. N., KARPLUS, M., *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9032–9037.
- 82 CUI, Q., ELSTNER, M., KAXIRAS, E., FRAUENHEIM, T., KARPLUS, M., *J. Phys. Chem. B* **2001**, *105*, 569–585.
- 83 AMARA, P., FIELD, M. J., in: *Computational Molecular Biology*, LESZCZYNSKI, J. (ed.), Elsevier, Amsterdam, **1999**, p. 1.
- 84 KOLLMAN, P. A., KUHN, B., PERÄKYLÄ, M., *J. Phys. Chem. B* **2002**, *106*, 1537–1542.
- 85 GAO, J., *Acc. Chem. Res.* **1996**, *29*, 298–305.
- 86 THOMPSON, M. A., SCHENTER, G. K., *J. Phys. Chem.* **1995**, *99*, 6374–6386.
- 87 HUGHES, J. M., HUTTER, M. C., REIMERS, J. R., HUSH, N. S., *J. Am. Chem. Soc.* **2001**, *123*, 8550–8563.
- 88 HOIJOU, H., INOUE, Y., SAKURAI, M., *J. Phys. Chem. B* **2001**, *105*, 867–879.
- 89 WARSHEL, A., CHU, Z. T., *J. Phys. Chem. B* **2001**, *105*, 9857–9871.
- 90 CUI, Q., KARPLUS, M., *J. Phys. Chem. B* **2000**, *104*, 3721–3743.

6

Quantum-mechanical/Molecular-mechanical Methods in Medicinal Chemistry

FRANCESCA PERRUCCIO, LARS RIDDER, and ADRIAN J. MULHOLLAND

6.1

Introduction

It is widely recognized that computer modeling methods have important roles to play in many aspects of medicinal chemistry [1]. A particular recent focus has been the development and application of “embedding” methods which embed quantum-mechanical treatment of a small subsystem within a simpler representation of its surroundings [2, 3]. One promising approach is to combine the quantum mechanical (QM) method with a molecular mechanics (MM) treatment of the environment, so-called QM/MM methods. These methods have the potential to contribute to problems for which electronic-structure calculations, including environmental effects, are essential or desirable. Such applications include modeling of enzyme-catalyzed reaction mechanisms and calculation of the spectroscopic properties of proteins. In ligand design, knowledge of the chemical mechanism of an enzyme target can be invaluable in inhibitor design, and simulations with combined QM/MM techniques are of increasing importance in the investigation of mechanistic questions. In addition, such methods can enable the derivation of descriptors based on structure-activity relationships which take into account the effects of the protein on bound ligands. They also enable the analysis of effects such as ligand polarization on binding affinity, which are not treated by typical empirical potential functions.

In this chapter, we will review some basic features of embedding methods, with a particular emphasis on some QM/MM techniques. We will discuss the basic theory, and some practical aspects of how to perform QM/MM calculations on biological

systems. Finally, we review a number of recent applications. Most applications to date have been to study the mechanisms of enzyme-catalyzed reactions, an obviously suitable area for these hybrid techniques, because standard molecular mechanics potential functions cannot be applied to processes of chemical change. We cover here outlines of investigations of enzyme mechanisms of particular interest in medicinal chemistry, and mention also some other applications of interest in this important and growing field.

6.2

Theory

6.2.1

Methodology

The essence of the QM/MM approach is the partitioning of the system considered into a small region of most interest (where the quantum-mechanical description is required), and the bulk of the system (simply treated by use of molecular mechanical force fields) [2, 4, 5]. The major interest in the development of QM/MM approaches has focused on simulating enzyme reactions [3, 4, 6]. QM/MM techniques can, however, be applied to a wide range of applications in which quantum-mechanical treatment is required for a small part within a large system. Examples of such applications include reactions in other condensed phases (for example in DNA) [7], in solution [8–10], in solids (for example in zeolites) [11, 12], at surfaces [13], in clusters and in transition metal complexes [14, 15], in studies of solvent and solvation effects [10, 16–20], electronic excitation in large molecules [21, 22] and in solution, calculation of absorption energies in zeolites [23, 24], and studies of biological binding interactions [25–28]. In subsequent sections we will describe the use of QM/MM methods in studies of enzymes. Enzymes are large systems, but to investigate their chemical mechanisms we are mainly interested in a small region, the active site where the reaction occurs.

6.2.2

Basic Theory

Warshel and Levitt's study of lysozyme initiated the field of QM/MM methods [29]. Many groups [10, 30–33] have covered the basic theory of QM/MM methods in detail, so we will outline the theoretical basis only briefly. In accordance with Field et al. [30], the energy of the whole system, E , is written in terms of an effective Hamiltonian, \hat{H}_{eff} , and the electronic wavefunction of the QM atoms, ψ :

$$E = \frac{\langle \psi | \hat{H}_{\text{eff}} | \psi \rangle}{\langle \psi | \psi \rangle}$$

Considering the effective Hamiltonian equal to:

$$\hat{H}_{\text{eff}} = \hat{H}_{\text{QM}} + \hat{H}_{\text{MM}} + \hat{H}_{\text{QM/MM}} + \hat{H}_{\text{boundary}}$$

the total energy of the system is described by four contributions:

$$E = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}} + E_{\text{boundary}}$$

The energy of the QM atoms, E_{QM} , is calculated in a standard molecular orbital calculation. The energy of the atoms in the MM region, E_{MM} , is given by a molecular mechanics force field – it is defined by a potential function including terms for bond stretching, bond-angle bending, dihedral and “improper” dihedral angles, electrostatic interactions (point partial charges represent the MM atoms), and van der Waals interactions. The boundary energy, E_{boundary} , arises when the surroundings of the simulation system must be accounted for. The QM/MM interaction energy, $E_{\text{QM/MM}}$, consists of two terms: the electrostatic interactions and the van der Waals interactions. For calculation of the van der Waals interactions with MM atoms, QM atoms are treated as normal MM atoms [30]. Appropriate van der Waals parameters are therefore assigned for each QM atom, and the energies of these QM/MM interactions are calculated by a molecular mechanics procedure. The van der Waals terms describe dispersion and exchange-repulsion interactions between QM and MM regions and they are significant at close range, playing an important role in determining geometries and interaction energies. In semiempiri-

cal molecular orbital methods such as MNDO [34], AM1 [35], and PM3 [36] electrostatic interactions are treated differently from the treatment in *ab initio* methods. Electrostatic interactions are treated in *ab initio* QM/MM methods by including the MM atomic point charges directly through one-electron integrals, and the interactions of the classical charges with the nuclei of the QM system. In semiempirical molecular orbitals, QM/MM electrostatic interactions are calculated by including the point charges of the MM atom as atomic cores, which represent the nucleus and the inner electrons of an atom combined in the semiempirical methods. This QM/MM approach, in which the MM environment directly influences the QM system, is distinct from simple schemes such as the ONIOM method, in which the interactions between the higher and lower level regions are included at the lower level. Thus, a two-level ONIOM QM/MM calculation (as currently implemented) does not account for polarization in the QM region because of the MM electrostatic potential, an effect which may be significant in a polar system such as an enzyme [22, 37, 38].

6.2.3

QM/MM Partitioning Schemes

The partitioning of the system in a QM/MM calculation is simpler if it is possible to avoid separating covalently bonded atoms at the border between the QM and the MM regions. An example is the enzyme chorismate mutase [39] for which the QM region could include only the substrate, because the enzyme does not chemically catalyze this pericyclic reaction. In studies of enzyme mechanisms, however, this situation is exceptional, and usually it will be essential, or desirable, to include parts of the protein (for example catalytic residues) in the QM region of a QM/MM calculation, i.e. the boundary between the QM and MM regions will separate covalently bonded atoms (Fig. 6.1).

Separation of covalently bonded atoms into QM and MM regions introduces an unsatisfied valence in the QM region; this can be satisfied by several different methods. In the frozen-orbital approach a strictly localized hybrid sp^2 bond orbital containing a single electron is used at the QM/MM junction [29]. Fro-

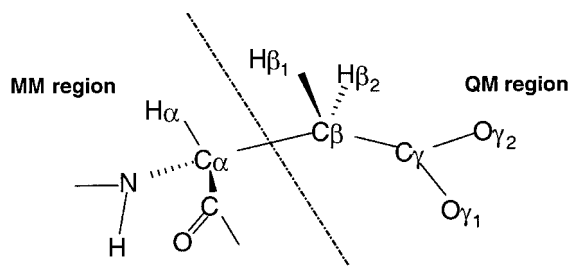


Fig. 6.1 An aspartate amino acid partitioned into quantum and classical (MM) regions. The functional group of the side chain, involved in the chemical reaction, lies within the quantum region and the backbone atoms are treated by using a molecular mechanics force field.

zen localized bond orbitals are obtained from calculations on small models and they do not change in the QM/MM calculations. This approach has been used at the semiempirical level [40–43], and also developed for *ab initio* calculations [44]. Gao et al. have proposed a similar generalized hybrid orbital method for QM/MM calculations at the semiempirical level of theory; this method uses modified semiempirical parameters at the QM/MM junction to enable the transferability of the localized orbitals [45].

A limitation of using the “frozen orbital” approach to describe the connection between QM atoms and MM atoms is the lack of flexibility in the electronic chemical changes as a consequence of freezing the frontier orbital. An alternative approach is to introduce so-called link atoms, which are generally treated as hydrogen atoms (to satisfy the valence of the boundary atom) in the QM system. In the QM/MM method of Field et al. link atoms do not interact with the MM region. Link atoms have the theoretical disadvantage of introducing extra degrees of freedom [46–48]. It is possible that interactions arising as a result of the presence of the link atom might be overcounted [49]. In the implementation of Field et al. [30], all MM bonding terms (bonds, angles, and dihedral energies) involving QM atoms were retained where such a term involved at least one MM atom. The link atom does not interact with MM atoms electrostatically or through van der Waals terms. Reuter et al. and Antes and Thiel [46, 50] have investigated the interactions of link atoms and how to treat them [51] and found it preferable not to exclude the link atom from inter-

actions with MM charges, for example all QM atoms should interact with the same set of MM atoms. In addition, it has been suggested that MM groups hosting the link atoms should have charges set to zero, to avoid unrealistic electrostatic interactions with the link atom [6, 52].

A related development for semiempirical QM/MM calculations is the “connection atom”, developed by Antes and Thiel, which is parameterized to reproduce the structural and electronic properties of a methyl group [46].

In the pseudobond method of Yang and coworkers [47] a pseudobond is formed with one free-valence atom with an effective core potential (optimized to reproduce the length and strength of the real bond). This core potential can be applied in Hartree-Fock and density functional calculations and is designed to be independent of the choice of the MM force field.

6.3

Practical Aspects of Modeling Enzyme Reactions

6.3.1

Choice and Preparation of the Starting Structure

A high-quality three-dimensional structure of an enzyme complex is the first requirement in setting up QM/MM calculations on an enzymatic reaction. A crystal structure of an enzyme with no ligands bound at the active site can be a poor model of the enzyme-substrate complex, because conformational changes might result from substrate binding. It is vital that the starting structure represents the reacting form, for example a point on the pathway of the chemical reaction within the enzyme. In practice, the structure could be the enzyme complexed with an inhibitor, a substrate, a product, or a transition state/intermediate analog.

Structures of actual enzyme-substrate complexes are generally difficult to determine, because the reaction occurs too quickly, but techniques now available occasionally enable study of these complexes [53]. Protein X-ray crystallography has several limitations, for example, it often gives little or no information about the positions of protons (because of the low electron density of hydrogen atoms) in a particular protein. This can cause prob-

lems in assigning protonation states to protein residues and in deciding between rotamers and tautomers. Hydrogen atoms can be added to the model by using standard computational procedures, for example the HBUILD routine in CHARMM [54], but it is essential that the correct protonation states are chosen in constructing a realistic model. Calculation of pK_a values by finite-difference Poisson-Boltzmann methods [55], or MM or QM/MM molecular dynamics simulations on different possible charge states (to examine which remain closest to, and are most consistent with, the crystal structure) can be useful in deciding which protonation states are appropriate for the amino acids residue in the protein [6, 56].

It also should be remembered that a crystal structure represents an average over all the molecules in the crystal and the time course of data collection. When alternative conformations for a side chain or ligand have been included in the crystallographic structure, one must decide which conformation to use in a simulation.

6.3.2

Definition of the QM Region

A crucial issue in QM/MM calculations is the choice of the QM system. In modeling an enzymatic reaction, it is important to construct a consistent hypothesis about which protein residues and solvent molecules are chemically involved in the reaction of interest. At this stage, kinetic data and site-directed mutagenesis experiments can provide useful information. Because QM/MM calculations can be very demanding in terms of computer time, it is necessary to limit the number of atoms in the QM region. Thus, only the side chains of the important residues should be included, and sometimes only the reactive portion of the substrate, when the whole substrate seems to be too large. Enlarging the QM region does not necessarily result in a more accurate model, particularly when low levels of QM theory are used [57]. In practice, covalent links between QM and MM atoms are preferably introduced some distance from the site where chemical and electronic changes of interest occur. Also, they should not be located close to highly charged QM groups and should not dis-

rupt conjugated systems. A carbon–carbon single bond some distance from the chemical change (for example between CA and CB of an amino acid side chain) is a good place for the QM/MM boundary.

6.3.3

Choice of the QM Method

QM/MM methods enable the investigation of large biomolecules, but require a balanced compromise between the level of QM theory, the size of the simulation system, and the nature of the calculations. It is essential to use a level of QM theory that represents accurately the system under consideration. Semiempirical methods are not suitable for modeling many systems – they do not, for example, perform well for strong hydrogen bonds, many transition structures, or for molecules containing atoms for which they are poorly parameterized. Sometimes, hydrogen bonds are represented more accurately as QM/MM interactions, compared with when they are treated entirely within the semiempirical QM region. Some conformational properties (e.g. rotation around peptide bonds) are better reproduced by MM force fields than with semiempirical methods [4]. Semiempirical methods often overestimate energy barriers to reactions [58–60]. They do, on the other hand, enable the treatment of larger QM systems and can be used in more extensive calculations. Occasionally they give better results than low-level *ab initio* QM/MM treatments. *Ab initio* calculations at the Hartree-Fock level perform inadequately for many reactions and inclusion of electronic correlation makes the calculations very time-consuming. It is often necessary to use large basis sets to obtain accurate results. Currently, *ab initio* QM/MM calculations on enzymatic reactions entail high computational expense, and they are generally limited to optimization of the geometry of the most important structures along the reaction pathway (for example reactants and intermediates). Although density-functional theory-based QM/MM methods can, potentially, achieve significantly greater accuracy than semiempirical methods, at a smaller computational cost than correlated *ab initio* calculations [32, 48, 61, 62], these methods can increase computational expense dramatically compared with

semiempirical and small basis set Hartree-Fock *ab initio* QM/MM methods. Density functional and higher *ab initio* levels of theory might, therefore, be most useful in tests of the more approximate semiempirical QM/MM methods.

6.4

Techniques for Reaction Modeling

6.4.1

Optimization of Transition Structures and Reaction Pathways

One of the most accurate ways of calculating a reaction potential-energy profile is to use reaction path-following methods. An example is the intrinsic reaction coordinate (IRC) [63] method. An IRC calculation of the reaction path starts from a transition state geometry towards the adjacent minima using a steepest descent minimization method with the system described by mass-weighted coordinates. The reaction path towards each minimum is initially determined by minimizing along the eigenvector that corresponds to the transition state's imaginary frequency [63]. Further steps along the IRC involve calculation of the gradient at each point calculated. These methods require calculation of the Hessian matrix; for biomolecules this can be computationally demanding.

Adiabatic mapping methods are an alternative, more approximate choice for modeling transition states and reaction pathways in large systems such as proteins. When we have set the reaction coordinate as a combination of internal coordinates corresponding to the chemical reaction that is occurring (for example to describe which bond(s) is (are) forming and breaking), we can calculate the energy of the system at each value of the reaction coordinate. The basic idea of the adiabatic mapping method is to restrain the appropriate internal coordinate(s) and change them gradually from the initial value (in a fully minimized structure of the reactants) to their final values (for example in the products) in a stepwise manner (complex combinations of internal coordinates can, in principle, be chosen). At each step, the rest of the system is allowed to relax using minimization methods. This

approach is only valid if one conformation of the protein is involved in the reaction. The RESD procedure [64] in CHARMM [54] is an example of how to impose restraints on (combinations of) internal coordinates. Adiabatic mapping approaches tend to overestimate energy barriers if atom movements involved in the reaction are not included in the reaction coordinate. Recalculating the path backwards from products is a useful test to ensure that a proper adiabatic, reversible profile has been obtained. Reaction coordinate calculations must start from a fully minimized structure.

Geometry optimization algorithms for large molecules have been developed by many groups [38, 65–68]. An example is the software package GRACE [69, 70] developed in association with CHARMM [54], which is designed specifically for QM/MM calculations on enzyme reactions [69]. In the GRACE approach a large molecular system is divided into two subsections – the “environment” surrounding the “core”. The algorithm searches for a saddle point (transition structure) in the “core” degrees of freedom, while maintaining the potential energy of the environment at a minimum (i.e. in the environment the “fast-cycling” subset is optimized to a minimum before each energy and gradient is evaluated for the “core”, the “slow-cycling” subset). The advantage of the approach used in the GRACE package is that a Hessian matrix needs to be stored and maintained for the core only.

An alternative means of calculating reaction pathways is employed in so-called global methods. These methods treat the entire path as a succession of points [71] which are found simultaneously. Methods of this type (for example the conjugate peak refinement algorithm [72], available in the TRAVEL module of CHARMM, which has the advantage of requiring only first derivatives of the energy) have been used to determine reaction paths in a number of proteins [4, 73].

6.4.2

Dynamics and Free Energy Calculations

The reaction modeling techniques described so far (transition structure optimization, adiabatic mapping, and reaction path modeling) rely on the assumption that a single protein structure

properly represents the enzyme during the reaction of interest. For well-defined active sites, which comprise a single binding orientation of the substrate and which are inaccessible to bulk solvent, this is a reasonable approximation. *para*-Hydroxybenzoate hydroxylase and phenol hydroxylase are good examples of systems on which adiabatic mapping, involving a single protein conformation, have been applied successfully, as shown by good correlations between experimental activation energies and calculated barriers for the hydroxylation step [74, 75]. For larger, aspecific, and solvent-accessible active sites, however, multiple configurations might have to be taken into account.

Different conformations of the enzyme and possible conformations and orientations of a ligand, as a substrate, can be sampled in molecular dynamics (that is classical dynamics) simulations (MD), or with Monte Carlo methods. MD on the basis of a molecular mechanical force field is an extremely powerful and widely used method, which has applications in several areas of biophysics, e.g. protein folding, protein dynamics, protein-protein and protein-ligand interactions, and calculation of binding affinity. It can also be used to generate an ensemble of structures that could subsequently be used for QM/MM reaction-pathway calculations. Alternatively, MD can be performed with a QM/MM potential, which, for example, enables evaluation of a potential of mean force (the free energy profile along a chosen coordinate). Although a QM potential has the disadvantage of being computationally more expensive than an MM force field, QM potentials are generally more reliable for geometries further away from minima (MM force fields are generally derived on the basis of minimum energy structures).

A useful compromise between speed and accuracy is provided by semiempirical methods. In this context, semiempirical methods can be used as “fitting tools” rather than predictive methods [76]. Optimization of the semiempirical parameters to reproduce experimental or high-level *ab initio* results, for a specific reaction, can yield an accurate potential for the problem of interest at relatively low cost [77]. This approach has been shown to be successful in a QM/MM framework also [78] and is a powerful tool in the accurate study of enzyme reactions [57, 79] (see also Section 6.5.2).

As QM/MM dynamics simulations are relatively expensive, it is useful to limit the size of the simulation system, for example by

using a stochastic boundary approach [80]. This approach implies that a spherical portion of the protein, centered on the region of interest (the active site), is selected and solvated (in a sphere of explicit water molecules). During dynamics simulations, a solvent boundary potential prevents “evaporation” of water. A buffer zone is defined as the outer layer of the sphere, in which positional restraints maintain the correct protein structure and Langevin forces are applied to correct for exchange of energy with the environment (which is not treated explicitly). This procedure, designed to reduce the size of simulation systems, has been shown to provide good results comparable with full-size simulation systems (involving periodic boundaries) [80].

Although, in theory, reactions within the QM region of a QM/MM model can be observed in a dynamics simulation, many reaction barriers are too high to be frequently crossed in a simulation with a practical length. By using free-energy simulation methods [81] or umbrella sampling techniques [82], more efficient sampling is performed along an approximate reaction coordinate, to yield a potential of mean force. These methods require a good (approximate) reaction coordinate, which might sometimes be difficult to define. Also, the importance of the solvent contribution to the barrier (represented only in an average way in the potential of mean force) is a matter of ongoing debate [83]. These methods have, nevertheless, been widely used and shown to be useful in the context of QM/MM simulations of reactions in solution [78, 84] and enzyme-catalyzed reactions [57, 85].

In umbrella sampling an “umbrella” potential is added which, in a dynamics simulation, directs the system into regions of a reaction coordinate that would otherwise not be adequately sampled. The (probability) distributions along the reaction coordinate observed in such a biased simulation are corrected afterwards for the effect of the umbrella potential, leading to the “true” distribution. In practice, a series of simulations is performed with different restraints corresponding to different areas along the reaction coordinate. The statistics obtained from the various simulations are combined into a single free energy profile, for example by using weighted histogram analysis methods [86, 87].

6.5

Some Recent Applications

The great potential of QM/MM calculations has attracted much attention in the past decade and the number of studies published in recent years is now so large that it is not possible to cover them all here. Among recent QM/MM applications at different levels are *para*-hydroxybenzoate hydrolase [56, 74], citrate synthase [4–6, 52], uracil-DNA glycosylase [88], neuraminidase [89, 90], aldose reductase [91], human thrombin [92], glutathione *S*-transferases [57], and HIV protease [93].

In this section some of these studies will be discussed to illustrate the advantages of this approach in medicinal chemistry, and the different kinds of procedure that have been used.

6.5.1

Human Aldose Reductase

Aldose reductase (ALR2) is a monomeric, NADPH-dependent oxidoreductase catalyzing the reduction of a wide variety of carbonyl compounds to the corresponding alcohols. It is known [5] that under hyperglycemic conditions it participates in the polyol pathway to reduce *D*-glucose to *D*-sorbitol. It has been found that when tissues contain a high level of glucose, sorbitol builds up and apparently damages the membranes lining body tissues. Because this process is thought to be one of the factors contributing to diabetic neuropathy, a nerve disorder caused by diabetes, much effort is underway to develop effective inhibitors [6].

Although crystal structures and kinetic data reveal some aspects of the reaction mechanism in ALR2, details of the catalytic step are still unclear. The mechanism has been shown to involve stereospecific hydride attack from NADPH on the carbonyl carbon of the substrate, and protonation of the carbonyl oxygen from a nearby proton donor residue in the active site. Ternary complex crystal structures enabled identification of a specific anion-binding site in the active site formed by NADPH, His110, and Tyr48, suggesting that the reaction proceeds through a negatively charged species. Semiempirical QM/MM calculations and *ab initio* QM calculations were performed by Varnai et al. [91] to

elucidate whether the protonation of the carbonyl oxygen of the substrate precedes or follows hydride attack from NADPH on the carbonyl carbon of the substrate [10], and the identity of the proton donor to the carbonyl oxygen.

The nature of the proton donor in ALR2 is of great importance in understanding the catalytic mechanism of the enzyme. Potential proton donor residues in the active site are Tyr48 (part of a conserved triad Tyr48-Lys77-Asp43) and His110. Several reaction pathways were investigated on the basis of two models in which either Tyr48 (Model A) or protonated His110 (Model B) act as the proton donor. The catalytic step of the reduction of the substrate, D-glyceraldehyde, to glycerol was analyzed first in model calculations on fragments of the active site using AM1, PM3, and HF/3-21G* to obtain information about the catalytic region. The AM1/CHARMM22 QM/MM potential was used for QM/MM calculations [30]. Standard all-atom CHARMM22 parameters [71] were used, except for the substrate for which MP2/6-31G(d) electrostatic potential-derived charges were used for the molecule. The simulated system was an 18 Å radius sphere containing 174 protein residues, the cofactor, the substrate, and 611 water molecules (water molecules from the crystal structures plus addition of solvent water molecules by superimposition of a 20 Å radius sphere of pre-equilibrated TIP3P water molecules [94, 95]. The QM region contained the substrate, the nicotinamide ring of the cofactor, and the side chains of residues Tyr48, His110, Lys77, and Asp43. The rest of the system was treated by MM force fields.

Reaction pathways were obtained by adiabatic mapping using the RESD module in CHARMM [24]. The reaction coordinates for the two steps were the differences between the breaking and forming bond distances for hydride and proton transfer, respectively. Transition states were refined as saddle points with the TRAVEL module of CHARMM [72]. The results showed that the substrate binds to the enzyme by hydrogen bonding in an orientation that facilitates the stereospecific catalytic step in both models. The catalytic mechanism with Tyr48 as the proton donor proceeds through a negatively charged intermediate (in accord with experimental results) but has a high activation energy. In this case hydride transfers occurs before protonation. The calculations indicate that the His110 in the protonated form (Model B)

is a better proton donor than Tyr48; the lowest energy pathway in this mechanism proceeds via a protonated intermediate, a conclusion that is apparently not in agreement with experimental data. With His110 as the proton donor, the lowest-energy pathway has the protonation step preceding hydride transfer. According to these calculations, if His110 is present in its protonated form in the native complex it is the energetically favored proton donor compared with the Tyr48 in the active site with neutral His110.

6.5.2

Glutathione S-Transferases

The family of glutathione *S*-transferases (GST) has an important place in the large array of biotransformation enzymes that metabolize and detoxify drugs and other xenobiotics. Biotransformation enzymes are of increasing toxicological and pharmacological interest, because they determine to a large extent how fast, and via which metabolic pathway, xenobiotics are metabolized. In general these enzymes have broad substrate specificities, are present as multiple classes of isoenzymes, often subject to polymorphisms, and sometimes (e.g. the enzymes cytochrome P450 and GST) catalyze multiple types of reaction.

In a recent application, the conjugation of glutathione to phenanthrene 9,10-oxide, catalyzed by GST M1-1 from rat, was studied by QM/MM-based umbrella sampling [57]. Phenanthrene 9,10-oxide is the model substrate for epoxide ring opening reactions by GST. Some aspects of this reaction are of particular interest. First, as is often the case for biotransformation enzymes, the active site is highly solvent-accessible and proper inclusion of solvent effects is required for accurate modeling. In this example solvent was included explicitly, by use of a stochastic boundary approach. A second aspect of interest is the use of a genetic algorithm to calibrate the semiempirical AM1 treatment of the QM region. The QM region of the GST model contained the thiolate sulfur of glutathione, which is central in the conjugation reaction. Sulfur is a versatile, and therefore difficult, element in the context of semiempirical methods. In this specific case it was shown that reparametrization of just the sulfur, keeping the stan-

standard AM1 parameters for other elements (H, C, and O) present in the system, could significantly improve the accuracy of the results for the conjugation reaction studied. This was appropriate, because the standard AM1 parameters for H, C, and O [35] were derived independently from those for sulfur [96]. Partial reparametrization seems to be a good approach for large QM systems that contain elements in different chemical configurations, for which it may be difficult to improve on the general AM1 parameters.

Free energy profiles for the conjugation of glutathione to the epoxide moiety of phenanthrene 9,10-oxide were obtained by means of umbrella sampling along an approximate reaction coordinate (the difference between the $S_{\text{thiolate}}-C_{\text{epoxide}}$ and $C_{\text{epoxide}}-O_{\text{epoxide}}$ distances). The barriers in the calculated free energy profiles agreed with the experimental rate constant for the overall reaction, which supported the QM/MM model of this reaction and which confirms the epoxide ring-opening step to be rate limiting in the enzyme-catalyzed reaction.

The model was analyzed to obtain detailed insight into several aspects of the reaction. First, an atomic structure for the transition state was obtained as an average structure of the dynamic trajectory restrained to the top of the free energy profile. The transition state structures are indicative of interactions with key active-site residues, indicating important catalytic effects. This gave valuable information to supplement insight obtained from X-ray structures. Analysis of hydrogen bonding by solvent molecules along the reaction coordinate indicated a significant change in solvation of both the thiolate moiety of glutathione and the epoxide oxygen of phenanthrene 9,10-oxide, indicating that solvent effects have a dramatic effect on the energetics of the reaction, in agreement with experimental results. Finally, approximate effects of mutations were established. One mutation (Asp8Asn) represents a difference between two isoenzymes (M1-1 and M2-2) with markedly different stereoselectivity with respect to the products formed. The modeled mutation seemed to have a differential effect on the barriers towards these products, which suggests that this mutation is an important determinant of different diastereoselectivity between the isoenzymes. This illustrates the potential of QM/MM modeling in enabling understanding of the phenotypical consequences of genetic variations.

6.5.3

Influenza Neuraminidase

Neuraminidases are enzymes present in viruses, bacteria, and parasites. They are implicated in serious diseases such as cholera, meningitis and pneumonia. Neuraminidase from influenza virus aids the transmission of the virus between cells and maintains viral infectivity. In different strains of influenza several amino acids are conserved, especially in the active site, giving rise to hopes of finding a single inhibitor (and so a drug) for all the neuraminidase enzymes from influenza strains. The crucial question is whether a covalent bond is formed between the enzyme and the reaction intermediate.

Thomas et al. [90] investigated the reaction catalyzed by neuraminidase from influenza virus. QM/MM calculations were performed using AM1 in the QM part and treating the MM region with either the CHARMM22 [97] or OPLS-AA force fields [98]. The system under study [90] was partitioned into two concentric spheres: a sphere of 11 Å (containing 860 MM atoms free to move) and a sphere of 35 Å (containing 15,250 fixed MM atoms belonging to the protein and one to two solvation layers at the protein surface). The QM atoms included 47, 63, and 73 atoms (depending upon the stage of reaction being studied). Free energies were calculated by umbrella sampling for different steps along the reaction coordinate. *Ab initio* QM calculations were used to validate the semiempirical results and quantum dynamical effects were analyzed by path integral simulations [90].

The calculations found there was no covalent intermediate in the viral neuraminidase reaction and the intermediate was more likely to be hydroxylated directly. Because there is only a small energy difference between the two options (formation of a covalent bond or direct hydroxylation) Thomas et al. proposed it might be possible to design inhibitors covalently bound to the enzyme.

6.5.4

Human Thrombin

Thrombin is a serine protease and is one of the key enzymes in the blood coagulation system. It plays fundamental roles in the human body, but the major interest is in the control of throm-

bus formation. Inhibition of thrombin is an interesting target in the design of new antithrombotics. Mlinsek et al. [92] constructed a model based on X-ray inhibitor-enzyme complexes to predict K_i values for binding of an inhibitor to the active site of thrombin. The second stage of the study was identification of structural and electrostatic characteristics of the inhibitor that are important in their binding within the enzyme.

The QM/MM procedure was used to calculate the molecular electrostatic potential (MEP) at the van der Waals surfaces of atoms in the enzyme active site. The inhibitor was treated quantum mechanically and the rest of the complex by use of molecular mechanics force fields. This approach was used to retain 3D structural information about the bound inhibitor. The MEP results were used as input for the counter-propagation artificial neural network (CP-ANN). A genetic algorithm was then used to identify the atoms that affect the binding process. This study also identified the most important amino acid residues in the active site for inhibitor binding. The results were in good agreement with existing knowledge about the thrombin active site and its mechanism.

6.5.5

Human Immunodeficiency Virus Protease

Liu et al. [93] studied the catalytic mechanism of HIV protease by QM/MM molecular dynamics simulations. They performed calculations using the biomolecular simulation program package GROMOS [99] interfaced with MOPAC [60]. The PM3 semiempirical Hamiltonian was combined with the MM GROMOS87 force field and the SPC water model. The entire enzyme dimer with a bound hexapeptide was included in a periodic box of dimensions $51 \text{ \AA} \times 54 \text{ \AA} \times 72 \text{ \AA}$ with 5427 SPC water molecules. The QM region included the side chains of two catalytic residues (aspartate Asp25 and aspartic acid Asp25'), the scissile peptide, and a lytic water molecule. QM/MM dynamics simulations were performed for a total of few hundred picoseconds with a time step of 0.5 fs. Several possible reaction pathways were investigated by umbrella sampling. The calculations suggested a different mechanism from that proposed using structural data alone [93].

6.6

Conclusions

QM/MM methods have developed to the stage where they have become a practical and useful tool for investigation of a wide range of systems. In studies of enzyme reactions they have proved suitable for identification of the catalytic function of protein residues in the active site and for distinguishing between alternative possible reaction intermediates; they have also been used to reveal important interactions within the enzyme. QM/MM calculations can be performed at semiempirical, *ab initio* or density-functional QM levels. Transition state structures can be optimized, molecular dynamics simulations can be performed, free energy differences (e.g. activation-free energy) and quantum effects (e.g. tunneling and zero point corrections [100]) can be calculated. QM/MM methods do, however, require careful and consistent application, for example careful consideration must be given to the selection of the QM atoms, the location and treatment of the QM/MM boundary and the choice of theoretical model suitable for modeling a particular system. In addition to studies of the chemical mechanisms of biological reactions, QM/MM methods will undoubtedly play an increasingly important role in other areas of medicinal chemistry, e.g. modeling ligand-binding interactions.

6.7

References

- 1 A. R. LEACH, *Molecular Modeling: Principles and Applications*, Pearson Education, Harlow, 2001.
- 2 A. J. MULHOLLAND, G. H. GRANT, W. G. RICHARDS, *Protein Eng.* **1993**, 6, 133–147.
- 3 A. J. MULHOLLAND, *The QM/MM Approach to Enzymatic Reactions*, in: *Theoretical Biochemistry – Processes and Properties of Biological Systems*, L. A. ERIKSSON (ed.), Elsevier, Amsterdam, 2001, pp. 597–653.
- 4 A. J. MULHOLLAND, M. KARPUS, *Biochem. Soc. Trans.* **1996**, 24, 247–254.
- 5 A. J. MULHOLLAND, W. G. RICHARDS, *Modeling the Citrate Synthase Reaction. QM/MM and Small Model Calculations*, in: *Transition State Modeling for Catalysis*, D. G. TRUHLAR, K. MOROKUMA (eds), American

- Chemical Society, Washington, DC, 1999, pp. 448–461.
- 6 A. J. MULHOLLAND, W. G. RICHARDS, *Proteins Structure Function Genetics*, 1997, 27, 9–25.
 - 7 A. H. ELCOCK et al., *J. Am. Chem. Soc.* 1995, 117, 4706–4707.
 - 8 T. R. FURLANI, J. L. GAO, *J. Org. Chem.* 1996, 61, 5492–5497.
 - 9 D. WEI, D. R. SALAHUB, *Chem. Phys. Lett.* 1994, 224, 291–296.
 - 10 J. GAO, *Acc. Chem. Res.* 1996, 29, 298–305.
 - 11 N. A. BURTON et al., *Modelling Transition States in Condensed Phase Reactivity Studies*, in: *Transition State Modeling for Catalysis*, D. G. TRUHLAR, K. MOROKUMA (eds), American Chemical Society, Washington, DC, 1999, pp. 401–410.
 - 12 A. H. DeVRIES et al., *J. Phys. Chem. B* 1999, 103, 6133–6141.
 - 13 J. R. SHOEMAKER, L. W. BURGGRAB, M. S. GORDON, *J. Phys. Chem. A* 1999, 103, 3245–3251.
 - 14 L. GUIDONI, P. MAURER, S. PIANI, U. RÖTHLISBERGER, *Quant. Struct.-Act. Rel.* 2002, 21, 119–127.
 - 15 T. K. Woo et al., *J. Phys. Chem. B* 1997, 101, 7877–7880.
 - 16 J. GAO, X. XIA, *Science* 1991, 258, 631–635.
 - 17 M. GARCIA-VILOCA, A. GONZALEZ-LAFONT, J. M. LLUCH, *J. Am. Chem. Soc.* 1999, 121, 9198–9207.
 - 18 P. L. CUMMINS, J. E. GREARY, *J. Comput. Chem.* 1997, 18, 1496–1512.
 - 19 G. A. KAMINSKI, W. L. JORGENSEN, *J. Phys. Chem. B* 1998, 102, 1787–1796.
 - 20 C. D. BERWEGER, W. F. VAN GUNSTEREN, F. MÜLLER-PLATHE, *J. Chem. Phys.* 1998, 108, 8773–8781.
 - 21 I. P. MERCER, I. R. GOULD, D. R. KLUG, *J. Phys. Chem. B* 1999, 103, 7720–7727.
 - 22 R. D. J. FROESE, K. MOROKUMA, *Chem. Phys. Lett.* 1999, 305, 419–424.
 - 23 T. CLARK et al., *J. Mol. Model.* 1999, 5, 1–7.
 - 24 P. E. SINCLAIR et al., *J. Chem. Soc. Faraday Trans.* 1998, 94, 3401–3408.
 - 25 W. G. HAN, E. TAJKHORSHID, S. SUHAI, *J. Biomol. Struct. Dyn.* 1999, 16, 1019–1032.
 - 26 M. J. RAMOS et al., *Int. J. Quantum Chem.* 1999, 74, 299–314.
 - 27 J. D. MARECHAL et al., *J. Comput. Chem.* 2000, 21, 282–294.
 - 28 A. CARTIER et al., *Theor. Chem. Acc.* 1999, 101, 241–245.
 - 29 A. WARSHHEL, M. LEVITT, *J. Mol. Biol.* 1976, 103, 227–249.
 - 30 M. J. FIELD, P. A. BASH, M. KARPPLUS, *J. Comput. Chem.* 1990, 11, 700–733.
 - 31 D. BAKOWIES, W. THIEL, *J. Phys. Chem.* 1996, 100, 10580–10594.
 - 32 P. D. LYNE, M. HODOSCEK, M. KARPPLUS, *J. Phys. Chem. A* 1999, 103, 3462–3471.
 - 33 R. V. STANTON, L. R. LITTLE, K. M. MERZ, *J. Phys. Chem.* 1995, 99, 483–486.
 - 34 M. J. S. DEWAR, W. THIEL, *J. Am. Chem. Soc.* 1977, 99, 4899–4907.
 - 35 M. J. S. DEWAR et al., *J. Am. Chem. Soc.* 1985, 107, 3902–3909.
 - 36 J. J. P. STEWART, *J. Comput. Chem.* 1989, 10, 209–220.
 - 37 S. DAPPRICH et al., *Theochem – J. Mol. Struct.* 1999, 462, 1–21.
 - 38 F. MASERAS, K. MOROKUMA, *J. Comput. Chem.* 1995, 16, 1170–1179.

- 39 P. LYNE, A. J. MULHOLLAND, W. G. RICHARDS, *J. Am. Chem. Soc.* **1995**, *117*, 11345–11350.
- 40 V. THERY et al., *J. Comput. Chem.* **1994**, *15*, 269–282.
- 41 G. G. FERENCZY, G. NARAY-SZABO, P. VARNAI, *Int. J. Quantum Chem.* **1999**, *75*, 215–222.
- 42 G. G. FERENCZY et al., *J. Comput. Chem.* **1992**, *13*, 830–837.
- 43 G. G. FERENCZY et al., *J. Comput. Chem.* **1998**, *19*, 38–50.
- 44 X. ASSFELD, J. L. RIVAIL, *Chem. Phys. Lett.* **1996**, *263*, 100–106.
- 45 J. L. GAO et al., *J. Phys. Chem. A* **1998**, *102*, 4714–4721.
- 46 I. ANTES, W. THIEL, *J. Phys. Chem. A* **1999**, *103*, 9290–9295.
- 47 Y. K. ZHANG, T. S. LEE, W. T. YANG, *J. Chem. Phys.* **1999**, *110*, 46–54.
- 48 M. EICHINGER et al., *J. Chem. Phys.* **1999**, *110*, 10452–10467.
- 49 K. P. EURENIUS et al., *Int. J. Quantum Chem.* **1996**, *60*, 1189–1200.
- 50 I. ANTES, W. THIEL, *On the Treatment of Link Atoms in Hybrid Methods*, in: *Combined Quantum Mechanical and Molecular Mechanical Methods*, J. GAO, M. A. THOMPSON (eds), American Chemical Society, Washington, DC, **1998**, pp. 50–65.
- 51 N. REUTER et al., *J. Phys. Chem. A* **2000**, *104*, 1720–1735.
- 52 A. J. MULHOLLAND, P. D. LYNE, M. KARPLUS, *J. Am. Chem. Soc.* **2000**, *122*, 534–535.
- 53 G. PETSKO, *Nature Structural Biology* **1996**, *3*, 565–566.
- 54 B. R. BROOKS et al., *J. Comput. Chem.* **1983**, *4*, 187–217.
- 55 J. ANTOSIEWICZ et al., *J. Comput. Chem.* **1996**, *17*, 1633–1644.
- 56 L. RIDDER et al., *J. Mol. Graphics Model.* **1999**, *17*, 163–175.
- 57 L. RIDDER, I. M. C. M. RIETJENS, J. VERVOORT, A. J. MULHOLLAND, *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.
- 58 A. J. MULHOLLAND, W. G. RICHARDS, *J. Phys. Chem. B* **1998**, *102*, 6635–6646.
- 59 A. J. MULHOLLAND, W. G. RICHARDS, *Int. J. Quantum Chem.* **1994**, *51*, 161–172.
- 60 J. J. P. STEWART, *J. Comput. Aided Mol. Des.* **1990**, *4*, 1–105.
- 61 R. V. STANTON, D. S. HARTSOUGH, K. M. MERZ, *J. Comput. Chem.* **1995**, *16*, 113–128.
- 62 X. P. LONG et al., *J. Mol. Struct.* **1997**, *412*, 121–133.
- 63 K. FUKUI, *Acc. Chem. Res.* **1981**, *14*, 363–368.
- 64 K. P. EURENIUS et al., *Int. J. Quantum Chem.* **1996**, *60*, 1189–1200.
- 65 W. THIEL, *Theochem – J. Mol. Struct.* **1997**, *398*, 1–6.
- 66 P. DERREUMAUX et al., *J. Comput. Chem.* **1994**, *15*, 532–552.
- 67 O. FARKAS, H. B. SCHLEGEL, *J. Chem. Phys.* **1999**, *111*, 10806–10814.
- 68 S. PATCHKOVSKII, W. THIEL, *J. Comput. Chem.* **1996**, *17*, 1318–1327.
- 69 A. J. TURNER, V. MOLINER, I. H. WILLIAMS, *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323–1331.
- 70 V. MOLINER, A. J. TURNER, I. H. WILLIAMS, *Chem. Commun.* **1997**, *14*, 1271–1272.
- 71 O. S. SMART, *Chem. Phys. Lett.* **1994**, *222*, 503–512.
- 72 S. FISCHER, M. KARPLUS, *Chem. Phys. Lett.* **1992**, *194*, 252–261.
- 73 S. FISCHER, S. MICHNICK, M. KARPLUS, *Biochemistry* **1993**, *32*, 13830–13837.
- 74 L. RIDDER et al., *J. Am. Chem. Soc.* **1998**, *120*, 7641–7642.
- 75 L. RIDDER et al., *J. Am. Chem. Soc.* **2000**, *122*, 8728–8738.

- 76 A. GONZALEZ-LAFONT, T.N. TRUONG, D.G. TRUHLAR, *J. Phys. Chem.* **1991**, *95*, 4618–4627.
- 77 I. ROSSI, D.G. TRUHLAR, *Chem. Phys. Lett.* **1995**, *233*, 231–236.
- 78 P.A. BASH et al., *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 3698–3703.
- 79 E.Y. LAU et al., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9937–9942.
- 80 C.L. BROOKS III, M. KARPLUS, *J. Mol. Biol.* **1989**, *208*, 159–181.
- 81 C.L. BROOKS III, M. KARPLUS, B.M. PETTITT, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Wiley, New York, **1988**.
- 82 G.M. TORRIE, J.P. VALLEAU, *J. Comput. Phys.* **1977**, *23*, 187.
- 83 A. WARSHHEL, J. BENTZIEN, *Energetics and Dynamics of Transition States of Reactions in Enzymes and Solutions*, in: *Transition State Modeling for Catalysis*, D.G. TRUHLAR, K. MOROKUMA (eds), American Chemical Society, Washington, DC, **1999**, 489–499.
- 84 L. L. HO, A.D. MACKERELL JR, P.A. BASH, *J. Phys. Chem.* **1996**, *100*, 4466–4475.
- 85 F. PROUST-DE MARTIN, R. DUMAS, M.J. FIELD, *J. Am. Chem. Soc.* **2000**, *122*, 7688–7697.
- 86 S. KUMAR et al., *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- 87 C. BARTELS, M. KARPLUS, *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- 88 A.R. DINNER, G.M. BLACKBURN, M. KARPLUS, *Nature* **2001**, *413*, 752–755.
- 89 J.A. BARNES, I.H. WILLIAMS, *Biochem. Soc. Trans.* **1996**, *24*, 263–268.
- 90 A. THOMAS et al., *J. Am. Chem. Soc.* **1999**, *121*, 9693–9702.
- 91 P. VARNAI, W.G. RICHARDS, P.D. LYNE, *Proteins – Structure Function Genetics* **1999**, *37*, 218–227.
- 92 G. MLINSEK et al., *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1286–1294.
- 93 H. LIU, F. MÜLLER-PLATHE, W.F. VAN GUNSTEREN, *J. Mol. Biol.* **1996**, *261*, 454–469.
- 94 W.L. JORGENSEN et al., *J. Chem. Phys.* **1983**, *79*, 926–935.
- 95 E. NERIA, M. KARPLUS, *J. Chem. Phys.* **1996**, *105*, 10812–10818.
- 96 M.J.S. DEWAR, Y.-C. YUAN, *Inorg. Chem.* **1990**, *29*, 3881.
- 97 A.D. MACKERELL et al., *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- 98 W.L. JORGENSEN, D.S. MAXWELL, J. TIRADORIVES, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- 99 W.R.P. SCOTT et al., *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- 100 D.G. TRUHLAR, B.C. GARRETT, *J. Phys. Chem. B* **2000**, *104*, 1069–1072.

Molecular Properties

7

Atoms in Medicinal Chemistry

RICHARD F. W. BADER, CHERIF F. MATTA, and FERNANDO J. MARTIN

7.1

Why Define Atoms in Molecules?

The notion of a functional group is the operational concept of experimental chemistry. In chemistry we recognize the presence of a group in a given molecule and predict its effect on the static and dynamic properties of the molecule on the basis of the characteristic properties assigned to that group. In many instances the properties of a group are not only characteristic but seem to be transferable from one molecule to another. It is now possible to use theory to define a functional group and determine its transferability [1]. In instances of demonstrated transferability, the groups become molecular building blocks, enabling one to construct a molecule from known fragments and predict its resulting properties. Such transferability of properties is particularly true of the groupings of atoms, e.g. amino acids, nucleotides, simple sugars, and phosphates, that are the building blocks of biological macromolecules. The sole purpose of a theory of atoms in molecules is to emulate experiment in this regard – to define atoms and hence functional groupings of atoms with definable sets of properties that, when joined together, form a molecule of known or desired structure so that its properties can be understood and predicted from theory.

The atoms defined in the quantum theory of atoms in molecules (QTAIM) satisfy these requirements [1]. The atoms of theory are regions of real space bounded by a particular surface defined by the topology of the electron density and they have all the properties essential to their role as building blocks:

- All properties are additive, every property of the molecule equalling the sum of the corresponding values for its constituent atoms or groups. It is important to emphasize that every measurable property of a molecule has this additive behavior, including, for example, the dipole moment. The contribution of an amino acid side chain to a molecular dipole moment is definable from theory and is shown to be relevant when classifying the workings of the genetic code.
- The atoms and functional groups of theory maximize the transferability of their properties from one molecule to another. The reason for this is readily understood. Two pieces of matter or two atoms are identical and have identical properties only if they have identical charge distributions, that is, they are indistinguishable in real space. Because an atom of theory is defined by its charge distribution as a bounded region of real space, its form necessarily reflects its properties. What is remarkable is the exceptional transferability that the charge distribution of an atom or a functional grouping of atoms may exhibit. This has been shown to be particularly true for the groupings of atoms that correspond to the building blocks of biological macromolecules. Whereas such a finding must come as no surprise to a chemist used to understanding the properties of a protein in terms of its amino acid residues for example, it requires a theory of atoms in molecules to implement and make quantitative use of this knowledge.

7.2

Theory of Atoms in Molecules

The quantum theory of atoms in molecules is described in texts and several reviews [1–4]. A qualitative survey of the essential definitions and their application to problems in the field of medicinal chemistry are given here with two purposes:

- To indicate to the reader that the theory is based on the observable distribution of charge and its theoretical consequences. It is thus not only model-free but relates directly to the measurable properties of a system.

- To acquaint the reader with the potential uses of the theory in medicinal chemistry, a theory couched in the language and concepts of chemistry; a theory of atoms with characteristic properties, of the bonds that link them, and of structure and its stability.

7.2.1

Definition of Atoms and Molecular Structure

Matter is made manifest through the space-filling distribution of charge. A charge distribution consists of point-like nuclei embedded in a diffuse cloud of negative charge whose spatial distribution is described by the electron density, $\rho(\mathbf{r})$. The topology of the electron density is dominated by the attractive force exerted on it by the nuclei, a force that endows the density with its principal topological feature – that it has a maximum value at the position of each nucleus, a feature that is true for any plane containing a nucleus, as displayed for two planes of the molecule BF_3 (Fig. 7.1a and b). An immediate consequence of this topological feature of the density is the natural association of an atom with a region of space, each region being dominated by a given nucleus, with boundaries evident in the minima that occur between the nuclear maxima. The boundaries are determined by the balance in the forces that the neighboring nuclei exert on the density.

The definition of an atom and its surface are made both qualitatively and quantitatively apparent in terms of the patterns of trajectories traced out by the gradient vectors of the density, vectors that point in the direction of increasing ρ . Trajectory maps, complementary to the displays of the density, are given in Fig. 7.1c and d. Because ρ has a maximum at each nucleus in any plane that contains the nucleus (the nucleus acts as a *global attractor*), the three-dimensional space of the molecule is divided into atomic basins, each basin being defined by the set of trajectories that terminate at a given nucleus. An *atom is defined as the union of a nucleus and its associated basin*. The saddle-like minimum that occurs in the planar displays of the density between the maxima for a pair of neighboring nuclei is a consequence of a particular kind of critical point (CP), a point where all three derivatives of ρ vanish, that

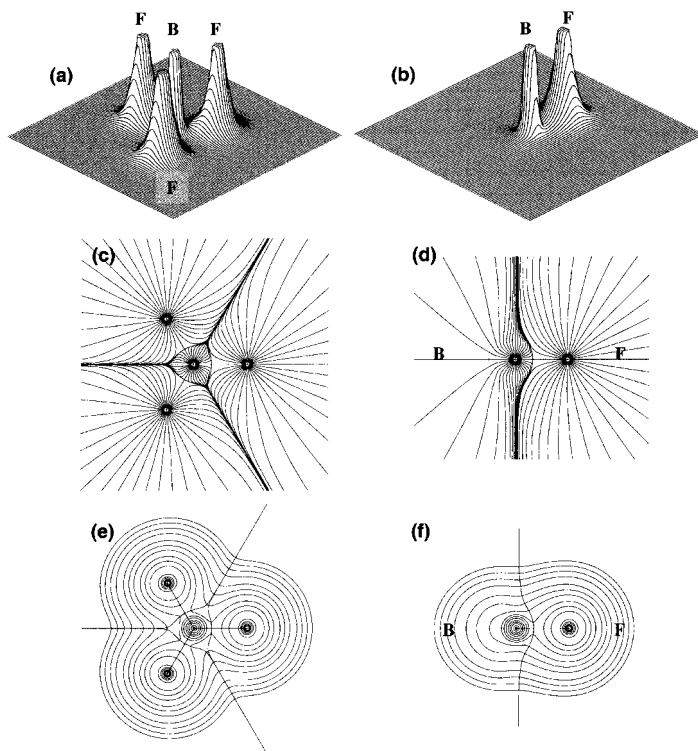


Fig. 7.1 The electron density $\rho(\mathbf{r})$ is displayed in the σ_h and σ_v symmetry planes of BF_3 in (a) and (b), respectively. The density is a maximum at the position of each nucleus (values of ρ greater than 2.5 au are not shown in the relief maps) and has a saddle between B and each of the F nuclei. The minimum in ρ at a saddle point denotes the position of a bond critical point (BCP). The trajectories traced out by the vectors $\nabla\rho$ are illustrated in (c) and (d) for the same planes as in (a) and (b). All the paths in the neighborhood of a given nucleus terminate at the maximum value of ρ found at each nucleus and define the atomic basin. (a) and (b) show two orthogonal views of the same BCP. They indicate that ρ is a *minimum* at the BCP along the internuclear axis, the curvature is positive, and two trajec-

tories *originate* at a BCP and terminate at the two associated nuclei. They define the bond path, a line of maximum electron density linking bonded nuclei. The curvature of ρ is negative in every direction perpendicular to the bond path and ρ appears as a *maximum* at the BCP. An infinite set of trajectories thus *terminate* at the BCP and define the interatomic surface. A BCP is located at the intersection of the bond path and the interatomic surface. (e) and (f) show the electron density in the same two planes overlaid with the bond paths and the interatomic boundaries. The values of the contours in this and subsequent figures increase from the outermost 0.001 au contour inwards in steps of 2×10^n , 4×10^n and 8×10^n au with n beginning at -3 and increasing in steps of unity.

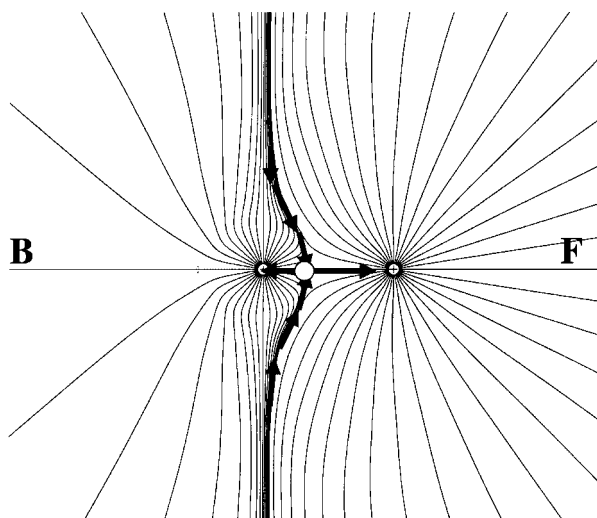


Fig. 7.2 A display of the trajectories of $\nabla\rho$ for the same plane as in Fig. 7.1d, complemented with arrows denoting the two unique trajectories that originate at the BCP, marked by an open circle, and terminate at each of the neighboring nuclei. They define the bond path.

Also indicated by arrows are the two trajectories that terminate at the BCP in this symmetry plane. They are members of the infinite set of such trajectories that define the interatomic surface of zero-flux in $\nabla\rho$ between the boron and fluorine atoms.

is, where $\nabla\rho = 0$. The density is a minimum at this CP along a line linking the two nuclei, a line that is uniquely defined by the pair of trajectories that originate at this CP and terminate at the two adjoining nuclear maxima, as indicated by a pair of arrows on the trajectory plot for one B–F interaction in Fig. 7.2. The line so defined represents a line of maximum density relative to any neighboring line and for a bound molecule it is referred to as a *bond path*, and the associated CP a *bond-critical point*. A bond path provides a universal indicator of bonding, linking all pairs of bonded atoms, irrespective of the nature of the interaction [5]. A *molecular graph*, defined by a molecule's connected set of bond paths, defines a molecular structure with particular critical points defining the presence of bonded rings and cages. Nuclear motion can induce topological changes in the density that correspond to the making and breaking of chemical bonds and to a change in molecular structure. Thus the topology of the charge distribution and its

change, as induced by nuclear motion, embodies the essential elements of structure and structural change [6].

The density is a maximum in all directions perpendicular to the bond path at the position of a bond CP, and it thus serves as the terminus for an infinite set of trajectories, as illustrated by arrows for the pair of such trajectories that lie in the symmetry plane shown in Fig. 7.2. The set of trajectories that terminate at a bond-critical point define the *interatomic surface* that separates the basins of the neighboring atoms. Because the surface is defined by trajectories of $\nabla\rho$ that terminate at a point, and because trajectories never cross, an interatomic surface is endowed with the property of *zero-flux – a surface that is not crossed by any trajectories of $\nabla\rho$* , a property made clear in Fig. 7.2. The final set of diagrams in Fig. 7.1 depict contour maps of the electron density overlaid with trajectories that define the interatomic surfaces and the bond paths to obtain a display of the atomic boundaries and the molecular structure.

These definitions apply to any atomic system, molecule or crystal. Fig. 7.3a illustrates their application to the charge distribution of the guanine-cytosine base-pair. Fig. 7.3b shows the molecular structure defined by the bond paths and the associated CPs that clearly and uniquely define the three hydrogen bonds that link the two bases. Fig. 7.3c shows the atomic boundaries and bond paths overlaid on the electron density in the plane of the nuclei. All properties of the atoms can be determined, enabling one, for example, to determine separately the energy of formation of each of the three hydrogen bonds.

7.3

Definition of Atomic Properties

Quantum mechanics applies to a segment of a system, that is, to an open system, if the segment is bounded by a surface of zero flux in the gradient vector field of the density. Thus the quantum mechanical and topological definitions of an atom coincide [1]. The quantum mechanical rules for determining the average value of a property for a molecule, as the expectation value of an associated operator, apply equally to each of its constituent atoms.

The atomic contributions are defined for every measurable property, including those induced by externally applied electric and magnetic fields and they are additive, summing to the corresponding value for the total system. It is not immediately obvious how one defines spatially additive contributions to properties such as the energy that involve attractive and repulsive interactions between all the particles in a molecule. That it should be possible to do so is, however, demanded by experiment. There is a long history, extending over 100 years, underlying the realization that the properties of a total system are the sum of its atomic contributions, including demonstrations of group additivity for heats of formation. The demonstration, as detailed elsewhere, that every measurable property of a system is expressible in terms of a corresponding “dressed” density distribution whose integration over an atomic basin yields the atom’s additive contribution to that property, is one of the most important results obtained from the physics of an open system [7, 8].

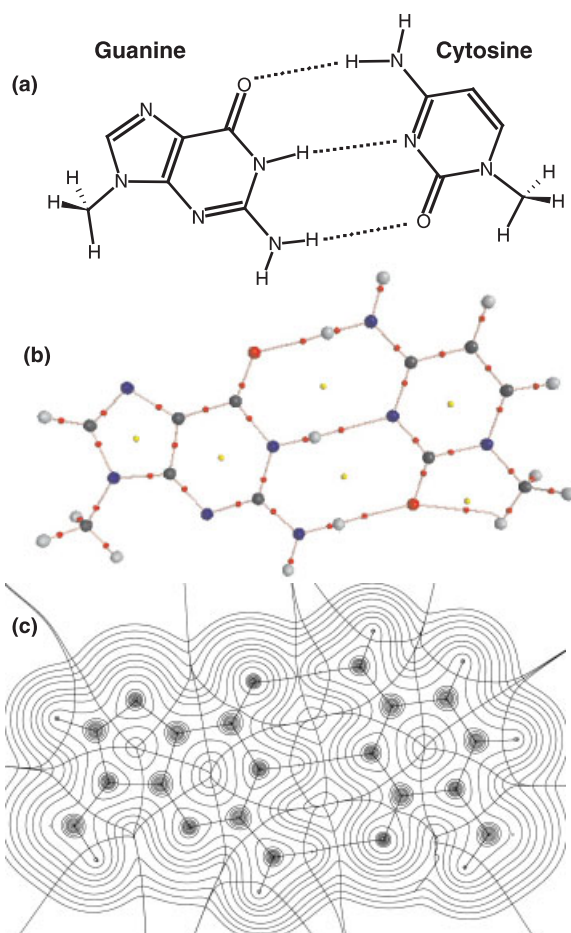
The atomic properties satisfy the necessary physical requirement of paralleling the transferability of their charge distributions – atoms that look the same in two molecules contribute identical amounts to all properties in both molecules, including field-induced properties. Thus the atoms of theory recover the experimentally measurable contributions to the volume, heats of formation, electric polarizability, and magnetic susceptibility in those cases where the group contributions are found to be transferable, as well as additive additive [4]. The additivity of the atomic properties coupled with the observation that their transferability parallels the transferability of the atom’s physical form are unique to QTAIM and are essential for a theory of atoms in molecules that purports to explain the observations of experimental chemistry.

7.3.1

Atomic Charges, Multipole Moments and Volumes

The charge associated with a given atom is an often used and important concept, but one that has physical meaning only when defined for a bounded, open system. The *atomic charge* for an atom A , denoted by $q(A)$, is defined as $q(A) = Z_A - N(A)$, the dif-

ference between the nuclear charge and the atomic population, the latter obtained by integration of the electron density over the atomic basin. Atomic charges obtained by a fitting of an electrostatic potential or charges defined in terms of atomic centered basis functions, such as Mulliken charges, in addition to lacking a physical basis, are strongly dependent on the basis set and are not, even in favorable cases, transferable between similar bonding environments. Models of overlapping atoms or overlapping orbital contributions do not account for the essential observation that atoms and functional groups can have characteristic properties despite changes in their immediate environments, inadequacies that have been previously detailed [9].



The atomic charges $q(A)$ minimize the basis set dependence and they are maximally transferable between molecules. They have also been demonstrated to be insensitive to conformational changes in the side chain of an amino acid [9]. The same applies to higher atomic multipole moments that, like $q(A)$, are defined by averaging the appropriate operator, dipole, quadrupole, etc., over the density in the atomic basin. In several studies Popelier has demonstrated that the QTAIM atomic multipoles when used in a multipole moment expansion, reproduce the exact *ab initio* electrostatic potential to any required accuracy [10–12]. Studies already completed for the free amino acids [13] and for the bound amino acid residues [14] provide a set of highly transferable atomic charges and moments that can be used in a multipole moment expansion for rapid determination of electrostatic potential. Similar sets of atomic moments will be forthcoming for the nucleotides, simple sugars, and phosphates. Atomic properties are transferable only if the associated geometrical data are transferable and the QTAIM results are obtained from calculations in which the molecular geo-

Fig. 7.3 Definition of the molecular structure of the guanine-cytosine base-pair: (a) identifies the atoms in terms of the usual chemical structure, a structure recovered in its entirety in the molecular graph defined by the topology of the density expressed in terms of the bond paths in (b), which is color coded – C (black), H (gray), N (blue), O (red); small dots for critical points, red for bond, and yellow for ring. There are four hydrogen bonds, three linking the two base molecules and an intramolecular bond in the cytosine moiety. They form three bonded rings in addition to the three rings associated with the pyrimidine and imidazole groups. A ring structure requires the presence of a ring CP (RCP), as indicated by a small yellow dot in the interior of each of the six bonded rings. An RCP is the topological opposite of a bond CP having two positive curvatures

whose associated trajectories originate at the RCP and span the ring surface (the density is a minimum at a ring CP in the surface) and a single negative curvature lying along the ring axis perpendicular to the surface. (c) is a display of the density in the plane of the nuclei overlaid with the bond paths and atomic boundaries. The bond path defining the intramolecular hydrogen bond in cytosine does not appear in (c) because of its very curved nature, evident in (b). In addition to the determination of the additive atomic contributions to each property, each bonded interaction can be characterized in terms of the properties of the density and energy at the BCP. A bond order that measures the physical delocalization of the electrons between the basins of a pair of atoms can be defined in terms of the electronic pair density (Section 7.6.2).

metries are fully optimized [14, 15] and the transferability is thus demonstrated and not a consequence of an assumed set of geometrical data, as has been done in the past.

Atomic volumes play an important role in relating physico-chemical properties to biological effects. Most atoms in molecules are not entirely bounded by interatomic surfaces and an *atomic volume* is defined as a measure of the space enclosed by the intersection of the atom's zero-flux surfaces with some outer envelope of the density. The envelope with a value of 0.001 au is generally chosen as this has been shown to yield molecular sizes in good agreement with experimentally assigned van der Waals radii [16, 17]. A related property is the van der Waals surface area, which QTAIM determines by integrating an atom's exposed contribution to a molecule's isovalued surface.

A molecular dipole moment, because it is a measurable property, can be partitioned into atomic and group contributions. A dipole moment is a product of a charge multiplied by the displacement of the negative from the positive charge centroids. It is independent of the position of the origin that is chosen to define the coordinates of a system only if the system has a zero net charge. Because atoms or groupings of atoms in molecules do not usually meet this requirement, it would seem that one cannot assign a unique atomic contribution to a molecular dipole moment. This, however, is not true when one takes proper account of the charge transferred across the atom's interatomic surfaces, as required by the physics of an open system [18, 19]. The contribution of each atom to a dipole moment is given by a sum of two terms. One is a measure of the displacement of the centroid of negative charge within the atom's basin relative to the position of its nucleus, the atomic moment. The second term measures the dipoles resulting from the charge transferred across each of the atom's interatomic surfaces. One recalls that a bond CP serves as the terminus for the set of trajectories that define an interatomic surface. Thus multiplication of the charge transferred across a given surface by the displacement of its associated CP from the nucleus determines the contribution of the charge-transfer term to the dipole moment. In this manner one can determine the contribution from the side chain of an amino acid, or of an amino acid residue in a polypeptide, to the dipole moment of the molecule [13].

7.4

QTAIM and Correlation of Physicochemical Properties

7.4.1

Use of Atomic Properties in QSAR

The correlation of atomic properties defined within QTAIM with experimental physicochemical properties to obtain quantitative structure-activity relationships (QSAR) is demonstrated with three examples [13]. Partial molal volumes of the amino acids are known experimentally and play an important role in understanding their properties and interactions in solution [20]. Group additivity of *partial molal volumes* V^0 at infinite dilution is well documented [21] and provides an ideal vehicle for demonstrating the ability of QTAIM to both recover experimental additivity schemes and provide a rigorous atomic basis. The experimental group contributions to V^0 are equated to the sum of two contributions – $V^0(\text{int})$, representing the intrinsic physical volume occupied by the molecule, and $V^0(\text{elect})$ the decrease in volume occupied by the surrounding solvent molecules resulting from the forces of electrorestriction. In the modeling of these two contributions $V^0(\text{int})$ is equated to the sum of the molecule's atomic volumes determined by the 0.001 au van der Waals density envelope. To account for the reduction in volume from electro-restriction that results from charges of either sign, one defines a *charge separation index*, CSI, the sum of the absolute charges on each of the atoms in the molecule [13]. These two atomic contributions yield a linear regression equation with the experimental V^0 values with $r^2=0.983$ with a negative coefficient for the CSI term, as expected for its effect in reducing the volume. Not only does the theoretically based correlation expression recover the experimental values to within the experimental variance, theory enables calculation of the contributions of individual groups to V^0 and comparison of these with the empirically fitted experimental values. The agreement between the two sets of group contributions is excellent, lending credence to the fundamental correctness of equating the molar volume to two primary contributions [13]. Some examples of the agreement of experimental and theoretical group contributions to V^0 ($\text{cm}^3 \text{mol}^{-1}$), with the latter values in parentheses, are: 28.2 (27.7) for $|\text{CH}_2\text{OH}$, 25.8 (25.5) for $|\text{COOH}$,

and 26.5 (25.5) for $|\text{CH}_3$. These group contributions to the molal volumes are a further example of the atoms of theory matching experimentally determined group contributions, previous examples being heats of formation, electric polarizability, and magnetic susceptibility.

The solvation of amino acid residues plays an important role in determining the conformation of proteins in aqueous solution. Experiments have been conducted to determine the contribution of the side chain to the free energy of transfer from the gas phase to aqueous solution using the side chain R capped with a hydrogen atom, a quantity that should model the hydrophilic or hydrophobic character of the side chain. Here again one finds that the CSI for the side chain correlates very well with the measured values of the free energy, yielding a linear regression relation with $r^2=0.93$. The reader is referred to the original paper for details [13].

The ability of QTAIM to determine atomic and group contributions for all measurable properties enables one to look for structure-property relationships for quantities other than atomic charges and volumes. In particular, one can determine the contributions of the side chain to the molecular dipole moment of the free amino acid. The dipole moment of a charged molecule depends on the choice of origin and the partitioning is applied only to side chains that have no formal charge. Although the main chain and side chain groups of an amino acid are not separately neutral, their net charges are in general small, the average magnitude equaling 0.07 e. Although the group contributions are precisely defined, irrespective of the residual charges on the two chains, the smallness of the charge implies that the group contributions provide good approximations of the dipole moment of each group.

The significance of the second letter in the triplet genetic code in determining the physical properties of an amino acid is well documented [22–25]. When the amino acid side chains are sorted according to the magnitude of their group contribution to the molecular dipole moment, and juxtaposing the second mRNA code letter of each, a pattern not dissimilar to that discussed by Wolfenden et al. emerges [25]. Amino acids with side chains with small dipolar polarizations ($0.09 \leq |\mu| \leq 0.29$ au) all have a pyrimidine base (either uracil or cytosine) as the second code letter. These include the side chains of leucine, isoleucine, alanine,

valine, and phenylalanine, all of which are hydrocarbons with no hetero-atoms. The amino acids with the highest dipolar polarizations ($1.67 \leq |\mu| \leq 1.85$ au), which include glutamine, asparagine, and histidine, all have adenine – a pyrimidine base – as the second code letter and all have two hetero atoms. The intermediate range includes amino acids with either a purine base (tyrosine, methionine, tryptophan, and cysteine) or a pyrimidine base (threonine and serine). Each side chain in the intermediate range has a single hetero atom. It is significant that this sorting lists consecutively the two amino acids with aromatic phenyl rings (phenylalanine and tyrosine).

These are three examples of the use of atomic properties to obtain quantitative structure-activity relationships (QSAR) or structure-function relationships. One should bear in mind that all properties have an atomic basis, making a multitude of new relationships possible. The atomic contribution to the polarizability, for example, is definable and shown to be transferable [26–28], offering the possibility of improving the use of an electrostatic potential map from zero- to first-order estimates of energies of interaction.

7.4.2

Use of Bond Critical Point Properties in QSAR

Next to the definition of an atom, the most dominant topological feature of the density is the presence of a bond critical point (BCP) between every pair of bonded atoms and the properties of the density at a BCP provide a further set of summarizing physical quantities that can be correlated with a molecule's observable properties. Primary among the uses of the properties of the density at a bond critical point (BCP) is its ability to concisely characterize the bonding between atoms linked by a bond path. The most important of these *bond indices* are the density, ρ_b , the Laplacian of the density, $\nabla^2\rho_b$ and the bond ellipticity, ε [1]. Other properties, related to the energy, e.g. the kinetic energy density and the electronic energy, can also be evaluated at a bond CP. Popelier has proposed that one constructs a BCP space wherein a given point fixes the values of a chosen set of bond indices for a given interaction [29]. The bond indices cluster in three major domains of this space that provide a corresponding classification

of bonding [1, 30]: *closed-shell interactions*, which include ionic, hydrogen-bonded, and van der Waals bonding; *shared interactions*, which encompass what is commonly referred to as covalent and polar bonding, the shared interactions being demonstrably distinct and separate from the domain of values encompassing *metallic interactions*. There are, of course, no boundaries between the domains, bonds with transitional characteristics causing each domain to merge smoothly with neighboring domains.

Popelier has proposed a measure of similarity that operates in the BCP space [29]. The use of this space enables one to represent molecules compactly and reliably in terms of topological information readily extracted from electron densities obtained from *ab initio* wave functions. The method is called quantum topological molecular similarity (QTMS) and is directly applicable to the construction of QSAR schemes [31]. There are a number of ways, each of increasing mathematical sophistication, of obtaining a measure of molecular similarity using the chemical descriptor vectors that span the BCP space. The reader is referred to the original papers and only the most direct measure is given here to illustrate the ideas. An Euclidean measure of the distance between two CPs in the BCP space is given by the square root of the sum of the squares of the differences in each of their bond indices. The distance $d(A,B)$ between two molecules A and B in the BCP space is then obtained by summing the individual distances for corresponding CPs between the two molecules. The smaller the distance $d(A,B)$, the more similar the two molecules are. Molecular ordering required for QSAR is obtained by listing the molecules in order of increasing $d(A,B)$ from a reference molecule forming a bound of the resulting sequence. The method, in combination with the partial least-squares procedure, is shown to produce statistically valid QSAR for measured acidities of five sets of carboxylic acid systems as a function of substitution. The approach can also be used to isolate the common reactive center in a series of molecules.

7.4.3

QTAIM and Molecular Similarity

Two objects are similar and have similar properties to the extent that they have similar distributions of charge in real space. Thus chemical similarity should be defined and determined using the atoms of QTAIM whose properties are directly determined by their spatial charge distributions [32]. Current measures of molecular similarity are couched in terms of Carbo's molecular quantum similarity measure (MQSM) [33–35], a procedure that requires maximization of the spatial integration of the overlap of the density distributions of two molecules the similarity of which is to be determined, and where the product of the density distributions can be weighted by some operator [36]. The MQSM method has several difficulties associated with its implementation [31]:

- the dependence of the value of the measured index on the particular choice of superposition of the molecular geometries;
- the time and cost of the required integration of products of molecular densities;
- the dominance of the core densities in determining the degree of overlap of the molecular densities; and finally
- the question of which regions of the molecules to be compared should be included in the maximization of the overlap.

These problems are avoided when the atomic or BCP properties of QTAIM are used in similarity or QSAR studies.

A comparative study of the charge distributions of enkephalin, morphine, and the oripavine molecule PEO, undertaken to identify a possible common binding site for their agonistic action at opioid receptors, serves as an example of the determination of similarity afforded by the use of the atoms of QTAIM [37]. Tyrosine is a precursor in all three structures and it is this region of the three molecules that has high similarity when the charge distributions are compared in terms of the atomic boundaries and bond paths. Fig. 7.4 illustrates the charge distributions of tyrosine and protonated forms of morphine and of PEO. They illustrate how a set of connected zero-flux surfaces can be found to isolate within each molecule the region that is most similar in all three molecules. The apparent similarity of the atoms within this maximum region in morphine and PEO is quantified by the

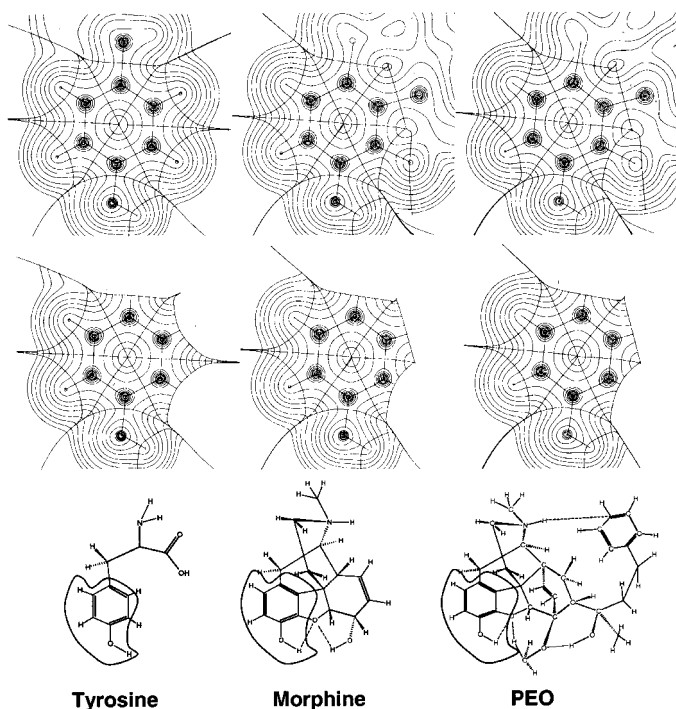


Fig. 7.4 The top row illustrates the electron densities for portions of the precursor tyrosine, for morphine, and for PEO. Note, as demonstrated in the second row, how the boundaries of the atoms enable

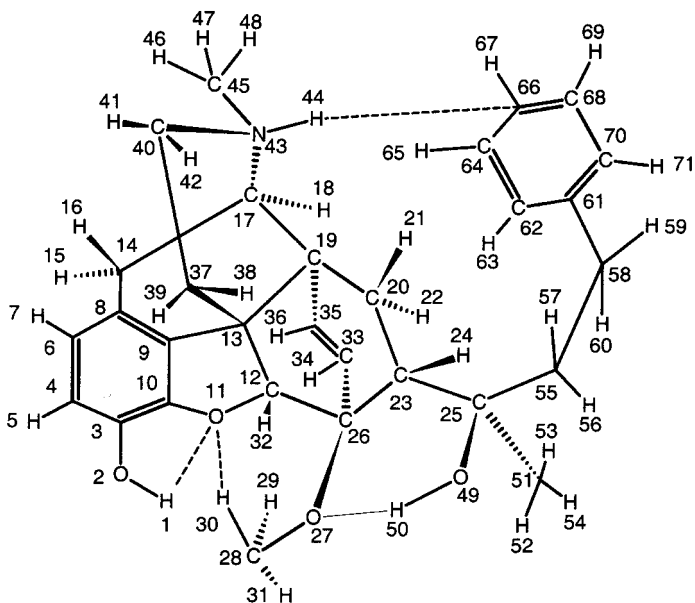
one to uniquely isolate and define, as a bounded region of real space, the group that has maximum similarity in all three molecules, with the third row locating this common group in the three molecules.

corresponding high similarity in the atomic properties, as summarized in Tab. 7.1 which lists the charge $q(A)$, the magnitude of the first moment $|\mu(A)|$, the magnitude of the quadrupolar polarization $Q(A)$, the energy $E(A)$, and the volume $v(A)$ for each atom A . Rather than assign a single number as a measure of the extent of similarity, one instead relates any observed differences between the physiological activity of the two molecules directly to the residual differences between their local properties. Thus the atoms of QTAIM, in addition to identifying a common active site within a set of molecules, enable maximum quantitative comparison of their properties.

Tab. 7.1 Comparison of atomic properties for the maximal transferable group in morphine and PEO.*

Atom	$q(\Omega)$	$q(\Omega)$	$ \mu(\Omega) $	$ \mu(\Omega) $	$ Q(\Omega) $	$ Q(\Omega) $	$E(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$\nu(\Omega)$
	PEO	Morphine	PEO	Morphine	PEO	Morphine	PEO	Morphine	PEO	Morphine
H1	0.655	0.656	0.133	0.133	0.042	0.043	-0.3176	-0.3171	17.2	17.2
O2	-1.283	-1.283	0.332	0.334	0.804	0.806	-75.3741	-75.3874	117.5	117.4
C3	0.660	0.665	0.758	0.760	3.135	3.128	-37.4492	-37.4509	57.4	57.4
C4	0.102	0.105	0.157	0.160	3.368	3.369	-37.7959	-37.8024	78.6	78.2
H5	0.015	0.019	0.098	0.098	0.454	0.452	-0.6291	-0.6274	46.0	45.7
C6	0.069	0.070	0.049	0.052	3.381	3.365	-37.7981	-37.8041	78.6	78.6
H7	-0.027	-0.023	0.098	0.097	0.480	0.479	-0.6466	-0.6449	48.5	48.3
C8	-0.026	-0.022	0.152	0.149	3.193	3.173	-37.8882	-37.8987	66.6	66.2
C9	-0.023	-0.027	0.187	0.158	2.987	3.062	-38.0415	-38.0415	58.3	58.0
C10	0.526	0.505	0.731	0.727	3.340	3.410	-37.6097	-37.6250	58.1	59.4
O11	-1.262	-1.253	0.238	0.242	0.880	0.833	-75.3627	-75.3687	97.7	97.0
C12	0.571	0.558	0.597	0.607	1.148	0.919	-37.5042	-37.4946	39.7	42.9
C13	0.046	0.059	0.036	0.048	0.304	0.393	-37.8884	-37.8550	40.0	41.4
C14	0.182	0.186	0.096	0.097	0.382	0.431	-37.7173	-37.7305	52.6	53.1
H15	-0.059	-0.010	0.105	0.105	0.401	0.401	-0.6455	-0.6416	46.4	46.5
H16	-0.017	-0.054	0.096	0.098	0.450	0.449	-0.6679	-0.6635	45.2	46.0
N43	-1.310	-1.292	0.022	0.041	0.221	0.301	-55.0165	-55.0374	63.3	63.8
H44	0.489	0.463	0.159	0.166	0.078	0.050	-0.4124	-0.4295	22.7	25.6
C45	0.547	0.541	0.485	0.490	1.041	1.097	-37.5110	-37.5141	58.2	58.9
av dev		0.012	0.006	0.006		0.041		0.009		0.7

* See following page for numbering scheme.



Numbering Scheme for PEO

7.5

Use of QTAIM in Theoretical Synthesis of Macromolecules

The transferability of group properties defined for proper open systems has been extensively studied and documented [18, 26, 38, 39]. One distinguishes two types of transferable behavior – *near perfect transferability* and *compensatory transferability*. Of these, the latter is more common. The extensive tabulations of experimentally derived group properties by Benson et al. [40] demonstrate that group additivity of thermodynamic properties must often be only apparent, because it is found to apply when a group is unavoidably perturbed by a change in its environment. These are examples of *compensatory transferability*. Electronic charge can be transferred from one group to another when they are linked to form a new molecule but charge is necessarily conserved in the process. In compensatory transferability, the perturbation caused by the conjoining of the two groups results in the conservation of all properties and not just the charge. Thus energy is conserved if the energy lost by one group equals the energy gained by the other.

Compensatory transferability is widespread and found even when each group undergoes nontrivial changes. For example, the

formation of pyridine from the |CHCHCH| group of benzene and the |CHNCH| group of pyrazine results in transfer of 0.024 e from the former to the latter and to compensating energy changes of the order of $\pm 22 \text{ kcal mol}^{-1}$; these sum to yield a heat of formation for pyridine that differs by only $0.1 \text{ kcal mol}^{-1}$ from the mean value of the heats of formation of benzene and pyrazine. The classic example of group additivity found for the homologous series of saturated hydrocarbons has been shown to be the result of both compensatory and near perfect transferability [39].

7.5.1

Assumed Perfect Transferability in the Synthesis of a Polypeptide

Many examples of atomic charge distributions that are essentially transferable between different systems or between different sites within a given system have now been documented and shown to contribute identical additive amounts to all of a system's properties [18, 19, 26, 37–39]. Near perfect transferability is both expected and found in the “building blocks” of biological molecules and one may take advantage of this to construct a polypeptide by the simple joining of the density distributions of amino acid residues previously determined in a suitable model [41]. The charge distributions and atomic properties of the residues |Aa| of all the genetically-encoded amino acids have been determined in the tripeptide Gly' |Aa| Gly'' where all geometrical properties are optimized. The tripeptide $\text{H}_2\text{NCH}_2\text{C}(=\text{O})|\text{NHC}\alpha\text{HRC}(=\text{O})|\text{NHCH}_2\text{CO}_2\text{H}$, is referred to as the glycine mold. Each vertical bar denotes an amidic surface of zero-flux and R denotes the side chain of the |Aa| residue [14, 38]. The geometrical properties of both the main chain groups and the common bonds and functional groups of the side chains are found to be transferable. This finding is a necessary prerequisite for the parallel transferability of the bond and atomic properties [14, 15]. The residue for serine determined in this manner and cut from the glycine mold at its two amidic interatomic surfaces is illustrated in Fig. 7.5. The two surfaces defining |Aa|, one bordering the amino N atom (the N-surface), the other bordering the keto carbon atom (the C-surface) of |Aa|, should form a pair of *complementary surfaces*, because ideally they are different sides of identical amidic surfaces, the concavity of the C-surface

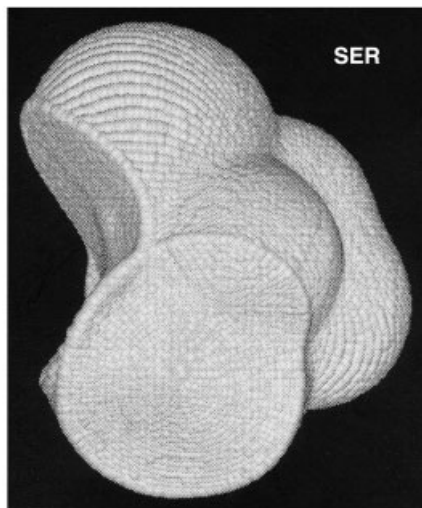


Fig. 7.5 The serinyl group $|\text{NHCH}(\text{CH}_2\text{OH})\text{C}(=\text{O})|$ cut from the “glycine mold” represented by the intersection of its van der Waals 0.001 au isodensity surface with the $-\text{C}(=\text{O})|$ or “C-surface” at the top left and the $|\text{NH}-$ or “N-surface” at the bottom center. These are the complementary sides of the amidic zero-flux surface characteristic of a polypeptide. All properties of the residue are defined and make additive contributions to the molecule constructed from it. The residue has a net charge of $-0.006 e$.

complementing the convexity of the N-surface. True complementarity of surfaces thus requires that the flux in the electric field through the N-surface, resulting from the transfer of density to N from its bonded keto carbon, should be equal and opposite to the flux through the C-surface. The condition of zero net flux in the electric field through the two surfaces bounding $|Aa|$ requires, as a consequence of Gauss’ divergence theorem, that the net charge on each $|Aa|$ residue be zero. The magnitude of the average net charge for the $|Aa|$ residues cut from the glycine mold, as obtained by the summation of the atomic charges, is $0.002 e$, a value that lies within the integration error in the determination of the atomic populations. For a charged residue such as $\text{His}^{(+)}$ or $\text{Tyr}^{(-)}$ the net charge is found to equal $+1.00$ or $-1.00 e$, respectively. Recovery of the proper net charge, 0 or $\pm 1 e$ are among the properties of the $|Aa|$ residues that demonstrate the complementarity of the two amidic surfaces and the consequent near perfect transferability of the amino acid residues.

The winter flounder “antifreeze” protein (AFP), characterized by Sicheri and Yang [42], consists of 37 residues of eight amino acids in an α -helix configuration. The AFP protein was synthesized by the conjoining of the $|Aa|$ residues determined in the glycine mold, with the exception of the two residues at each of the termini, HOOC-Asp|Thr| and $|\text{Ala|Arg-NH}_2$ and the synthesized protein fragment is left with open amidic surfaces on the

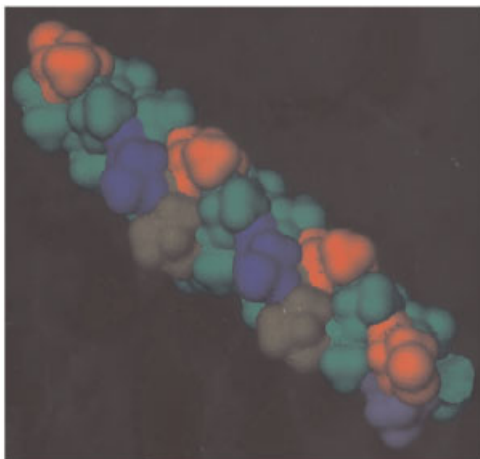


Fig. 7.6 The 33-residue fragment of the winter flounder antifreeze protein (AFP), constructed by the conjoining of the |Aa| amino acid residues defined within the “glycine mold”, pictured in terms of its 0.001 au isodensity envelope, its van der Waals envelope. This is a view showing the ice-binding motif. It is believed that the AFP strand binds

to the 201 plane of ice by establishing hydrogen bonding interactions between the hydroxyl group of threonine and the oxygen atoms in ice [44]. Color code for exposed residues: Ala (green), Ser (orange), Asp (gray), Leu (olive green), Asn (blue), Thr (red), Glu (pale green). Ser is the residue between Ala (green) and Asp (gray) at bottom end.

terminal |Ala| and |Thr| groups. The advantage of this mode of synthesis is the lack of need to “cap” a terminal residue, because the properties of an open system remain unchanged if its bounding surfaces remain unchanged and an uncapped residue is ignorant of the absence of the missing bonded neighbor [38]. In this way one can construct only that portion of a macromolecule that is of interest, as in this example wherein the groups responsible for repression of ice formation are found in the interior threonine groups [43]. The synthesized AFP is pictured in Fig. 7.6 in terms of its 0.001 au iso-valued density envelope. One notes the virtually seamless nature of the joining of the amidic surfaces of zero-flux of the |Aa| residues. Other, more quantitative measures of the errors incurred in the joining of the groups, including the volume of non-overlap of the surfaces, are computed and found to be acceptably small [14].

All properties of the protein are determined by simply summing the corresponding atomic properties. The electrostatic po-

tential field for any portion can be expanded in terms of the atomic multipole moments, particularly in the neighborhood of the Thr residues, because it is their hydrogen-bonded interaction with water molecules that is thought to be responsible for preventing the formation of the ice structure. The total charge accumulated by assembling the 33 residues is only 0.089 e, too small an error to noticeably affect the calculation of the electrostatic field. Methods that consider the joining of overlapping fragments in a theoretical construction of a molecule are unable to ensure the absence of an accumulated net charge.

Matta [37] used the conjoining of |Aa| residues to obtain the properties of the charge distributions of the proteins Leu- and Met-enkephalin that consist of five residues, in a study of the agonistic activity of the oripavine PEO and morphine. Assembly of the five residues with the required terminal groups resulted in net charges of +0.051 e and -0.043 e, respectively.

7.5.2

The Assembly of Buffered Open Systems in a Macrosynthesis

In his study of the opioids Matta introduced an alternative method for obtaining the charge distribution of a large molecule, by using the properties of open systems, and applied it to the PEO molecule [37]. In Matta's approach a large molecule is broken down into smaller molecules that are amenable to computation with a large basis set, the smaller molecules being chosen so that each contains a group in common with the large molecule of interest, bounded by a set of buffer atoms or groups. Because of the unique ability of QTAIM to define atomic boundaries, the desired groups are then extracted from each of the computed small molecule densities and assembled by matching of their outer atomic surfaces to construct the desired product molecule. Differences between the two densities are acceptably small, a conclusion made quantitative by comparison of the atomic properties given in Tab. 7.2. The averages of the absolute deviations ($\text{av}|\text{dev}|$) among the two sets of properties are extremely small. The error in the total volume of the assembled PEO molecule using the van der Waals 0.001 au envelope is 0.13% and the error in its total charge is 0.02 e, another example of how the use

Tab. 7.2 Comparison of properties for some atoms in intact and assembled PEO.

Atom	$q(\Omega)$ intact	$q(\Omega)$ assemb.	$ \mu(\Omega) $ intact	$ \mu(\Omega) $ assemb.	$ Q(\Omega) $ intact	$ Q(\Omega) $ assemb.	$E(\Omega)$ intact	$E(\Omega)$ assemb.	$\nu(\Omega)$ intact	$\nu(\Omega)$ assemb.
H1	0.653	0.655	0.133	0.133	0.042	0.042	-0.3185	-0.3176	17.3	17.2
O2	-1.282	-1.282	0.330	0.332	0.806	0.804	-75.3652	-75.3741	117.3	117.5
C8	-0.025	-0.026	0.150	0.152	3.179	3.193	-37.8827	-37.8882	66.6	66.6
O11	-1.264	-1.262	0.234	0.238	0.884	0.880	-75.3553	-75.3627	97.8	97.7
C13	0.047	0.045	0.044	0.036	0.277	0.304	-37.8814	-37.8884	40.0	40.0
C25	0.628	0.614	0.661	0.663	0.883	0.895	-37.3737	-37.3719	35.6	35.8
O27	-1.274	-1.270	0.196	0.205	1.001	0.995	-75.3634	-75.3944	88.0	89.6
C35	0.003	-0.004	0.075	0.096	2.942	3.011	-37.8848	-37.8933	70.9	71.4
C40	0.456	0.452	0.522	0.526	1.120	1.151	-37.5684	-37.5765	49.9	49.8
N43	-1.312	-1.310	0.026	0.022	0.209	0.221	-55.0455	-55.0165	63.3	63.3
H44	0.491	0.489	0.158	0.159	0.077	0.078	-0.4110	-0.4124	22.5	22.7
C45	0.551	0.547	0.491	0.485	1.046	1.077	-37.5039	-37.5110	57.8	58.2
H46	0.006	0.016	0.098	0.098	0.422	0.419	-0.6383	-0.6334	43.6	43.1
H47	0.012	0.014	0.103	0.103	0.414	0.422	-0.6311	-0.6307	46.6	45.8
H48	0.015	0.014	0.102	0.102	0.418	0.422	-0.6299	-0.6305	46.5	45.6
H50	0.675	0.671	0.112	0.113	0.072	0.074	-0.3106	-0.3120	11.0	11.4
C61	-0.023	-0.018	0.108	0.081	2.492	2.555	-37.8847	-37.8677	61.4	61.4
C66	0.010	0.012	0.090	0.088	3.479	3.461	-37.8320	-37.8143	79.7	79.5
av dev		0.004		0.005		0.017		0.0088		0.3

of proper open systems in a theoretical synthesis does not result in significant charge accumulation.

The success of this method of construction is again a result of the high transferability of atoms defined as open systems, the same property that underlies their use in molecular recognition, the atomic boundaries uniquely defining corresponding fragments in two or more molecules. The assembly of judiciously chosen buffered open systems in the construction of a macromolecule can minimize or, occasionally, essentially remove the error incurred in the matching of their zero-flux surfaces. One should bear in mind that every property of a molecule, and not just those determined by the density, when synthesized from open systems is obtained by simply summing the properties of its constituent groups.

7.6

The Laplacian of the Density and the Lewis Model

The theory as presented so far is clearly incomplete. The topology of the density, while recovering the concepts of atoms, bonds and structure, gives no indication of the localized bonded and non-bonded pairs of electrons of the Lewis model of structure and reactivity, a model secondary in importance only to the atomic model. The Lewis model is concerned with the pairing of electrons, information contained in the electron pair density and not in the density itself. Remarkably enough however, the essential information about the spatial pairing of electrons is contained in the Laplacian of the electron density, the sum of the three second derivatives of the density at each point in space, the quantity $\nabla^2\rho(\mathbf{r})$ [44].

It is a property of the second derivative of a scalar function such as ρ that it determines where the function is locally concentrated and locally depleted, in the absence of corresponding maxima and minima in the function itself. Consider a function $f(x)$ with both maxima and minima. At a maximum in $f(x)$ the curvature is negative, $d^2f(x)/dx^2 < 0$, and the value of $f(x)$ is greater than the average of its values at $x+dx$ and $x-dx$. The reverse is true at a minimum in $f(x)$ where $d^2f(x)/dx^2 > 0$ and the value of $f(x)$ is less than the average of its values at $x+dx$ and $x-dx$. It is

in this sense that $f(x)$ is respectively, said to be locally concentrated or locally depleted at x , a statement that remains true even in for a point removed from a maximum or a minimum in $f(x)$. These considerations carry over to three-dimensional space and $\nabla^2\rho(\mathbf{r}) < 0$ implies that $\rho(\mathbf{r})$ is greater at \mathbf{r} than the average of its values in the immediate neighborhood of \mathbf{r} , with the reverse being true for $\nabla^2\rho(\mathbf{r}) > 0$. It is useful to define the function $L(\mathbf{r}) = -\nabla^2\rho(\mathbf{r})$, because a maximum in $L(\mathbf{r})$ then corresponds to a maximum in the concentration of electronic charge [1].

The topology of $L(\mathbf{r})$ is completely different from that for the density itself, with local maxima corresponding to the presence of Lewis bonded and nonbonded electron pairs. It has been known for some time that $L(\mathbf{r})$ recovers the shell structure of an atom in terms of a corresponding number of alternating pairs of shells of charge concentration and charge depletion and that on bonding with other atoms the outer or valence shell of charge concentration loses its uniformity resulting in the formation of local maxima, that is, local charge concentrations (CC) [44]. The number, relative size, and orientation of these CC enable faithful mapping of the localized bonded and nonbonded Lewis pairs assumed in the VSEPR model of molecular geometry [45].

7.6.1

The Laplacian and Acid-Base Reactivity

The phenomenon of electron pairing is a consequence of the Pauli exclusion principle. The physical consequences of this principle are made manifest through the spatial properties of the density of the Fermi hole. The Fermi hole has a simple physical interpretation – it provides a description of how the density of an electron of given spin, called the reference electron, is spread out from any given point, into the space of another same-spin electron, thereby excluding the presence of an identical amount of same-spin density. If the Fermi hole is maximally localized in some region of space all other same-spin electrons are excluded from this region and the electron is localized. For a closed-shell molecule the same result is obtained for electrons of β spin and the result is a localized α, β pair [46].

The spatial properties of the Fermi hole when coupled with the definition of an atom in a molecule provide a physical measure of electron delocalization. If the hole remains largely localized within the basin of a given atom the electrons are correspondingly localized in the basin of that atom. Correspondingly, the electron can go wherever its Fermi hole goes. If the Fermi hole of an electron when referenced to a given atom is delocalized into the basin of a second atom the electron is shared between them and the spreading of the Fermi hole provides a quantitative measure of electron delocalization. The numerical value of the delocalization index defined in this manner, between two atoms with similar electronegativity, is equal to the number of shared pairs of electrons expected on the basis of the Lewis model [47].

Under the conditions of maximum localization of the Fermi hole, one finds that the conditional pair density reduces to the electron density ρ . Under these conditions the Laplacian distribution of the conditional pair density reduces to the Laplacian of the electron density [48]. Thus the CCs of $L(\mathbf{r})$ denote the number and preferred positions of the electron pairs for a fixed position of a reference pair, and the resulting patterns of localization recover the bonded and nonbonded pairs of the Lewis model. The topology of $L(\mathbf{r})$ provides a mapping of the essential pairing information from six- to three-dimensional space and the mapping of the topology of $L(\mathbf{r})$ on to the Lewis and VSEPR models is grounded in the physics of the pair density.

The integral of $L(\mathbf{r})$ over an atomic basin must vanish as a consequence of the zero-flux surface condition and consequently the creation of regions with $L(\mathbf{r}) > 0$ must be coupled with the creation of others with $L(\mathbf{r}) < 0$. Just as the local maxima in $L(\mathbf{r})$ denote concentrations of electronic charge and hence the presence of sites of basic or nucleophilic activity, so the corresponding “holes” in $L(\mathbf{r})$, regions of local depletion in electronic charge correspond to sites of acidic or electrophilic activity. Thus an important feature of a Laplacian map, such as that displayed in Fig. 7.7a for the carbamoyl sarcosine molecule, an inhibitor of creatine, is its ability to locate the “lumps” and “holes” in electronic charge distribution, the sites of electrophilic and nucleophilic attachment. Numerous examples have been given wherein the relative orientation of approaching reactants can be predicted on the basis of the alignment of their respective maxima and

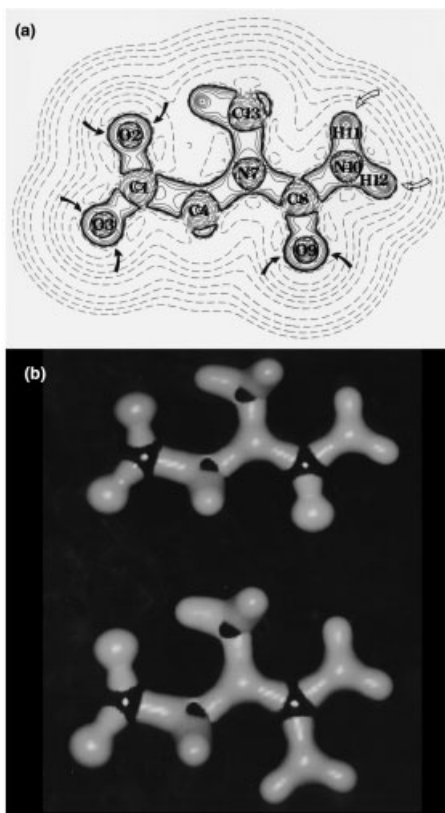


Fig. 7.7 (a) Contour map of $L(\mathbf{r})$ in the plane of the nuclei for the inhibitor carbamoyl sarcosine. Solid contours denote regions of charge concentration, dashed contours regions of charge depletion, the absolute values of the contours being the same as in Fig. 7.1. The non-bonded charge concentrations on the ketonic oxygen atoms are indicated by solid arrows, the “holes” on the amido hydrogens by open arrows. The out-of-plane holes on the doubly bonded carbon atoms are made evident in the reactive surface plot given in (b). (b) Reactive surface plots for creatine (bottom) and its inhibitor shown in (a) wherein one NH_2 of the guanidino

group is replaced by a keto oxygen atom. The two surfaces are similar with the exception of the region of the group replacement. Note the presence of the pronounced holes in the reactive surface at the positions of the carbon atoms in the carboxylic acid and guanidino groups, being largest for the latter group. These holes expose the carbon atom to nucleophilic attack or attachment from either side of the molecular plane. Although not clearly evident in this view, there are two nonbonded charge concentrations on each of the guanidino N atoms, one each side of the molecular plane.

minima in $L(\mathbf{r})$ [1]. The alignment of the very pronounced toroidal charge concentration on one Cl atom of Cl_2 with the axial charge depletion on a Cl in another molecule [44, 49] accounts for the bent geometry of the Cl_2 dimer in the gas phase and for the resulting non-van der Waals layered structure found in solid chlorine [50].

7.6.2

Molecular Complementarity

One can distinguish two measures of molecular complementarity in the fitting of one molecule on to another [32]. One, *van der Waals complementarity*, is determined by the size and shape of the atoms or groups that are brought into contact. It is determined by the complementarity of the physical shapes of the two molecules. It is of importance where non-directional van der Waals or dispersion forces are responsible for the mating of the molecules, as found, for example, in the alignment of the hydrocarbon chains forming biological membranes. These interactions increase in parallel with the area of contact between the interacting molecules. Consequently, the 0.001 au isodensity surface that defines a molecule's van der Waals shape [16] and the area of this surface as determined by the QTAIM programs [51] are necessary for predicting the relative orientation and resulting strength of the interaction. It is also documented that the atomic volumes of QTAIM correlate with additive contributions to the molecular polarizability [26], enabling one to use the atomic volumes to obtain quantitative estimates of the strength of such interactions [52].

It is *Lewis complementarity*, on the other hand, that is operative when the mating of the molecules is determined by acid-base interactions, one that is described and predicted by the complementary mating of the lumps with the holes in the two associated Laplacian distributions. A molecule's *reactive surface* is defined by the zero envelope of the Laplacian distribution, the envelope that separates the shells of charge concentration from those of charge depletion. The reactive surfaces make immediately clear the locations of the lumps, the nucleophilic sites, and the holes, the electrophilic sites, that are brought into juxtaposi-

tion of the two molecules in interactions governed by acid-base interactions. The reactive surfaces for creatine and its inhibitor carbamoyl sarcosine are shown in Fig. 7.7b to afford comparison of their reactive sites [32]. A comparison of such surfaces for morphine and PEO [37] illustrates the similarity of the reactive surface of the transferable portion of these two molecules and makes clear their common agonistic behavior. When the part of a substrate that binds to an active site has been identified, one could in principle obtain a copy of the reactive surface of the active site by performing a *complementary mapping* of the substrate's Laplacian distribution.

MacDougall and Henze have written and made available a new molecular visualization tool called EVolVis which enables one to explore interactively a molecule's electronic charge density [53]. It is particularly well suited for generating and studying displays of a molecule's reactive surface defined by the Laplacian of the charge density; it is so visually rewarding that it prompted a review in *Nature* by a member of the department of the history of art at Oxford [54]. MacDougall and Henze give a number of examples of EVolVis displays of the reactive surfaces of biological molecules, identifying, in particular, the key reactive sites in penamecillin.

7.7 Conclusions

This chapter has provided an introduction to the ideas underlying the quantum theory of atoms in molecules, the theory that gives theoretical expression to chemical concepts and enables one to employ these concepts in a quantitative manner for prediction and for understanding of chemical problems. The theory is particularly well-suited to problems in medicinal chemistry where the important role of building block molecules enables one to make maximum use of the transferability of atoms and groups defined as open quantum systems.

7.8

References

- 1 R. F. W. BADER, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, UK, 1990.
- 2 P. L. A. POPELIER, *Atoms in Molecules: An Introduction*, Pearson Education, Harlow, UK, 2000.
- 3 R. F. W. BADER, in: *Encyclopedia of Computational Chemistry*, Vol. 1, P. v. SCHLEYER (ed.), John Wiley and Sons, Chichester, UK, 1998, pp. 64–86.
- 4 R. F. W. BADER, P. L. A. POPELIER, T. A. KEITH, *Angew. Chem. Int. Ed. Engl.* **1994**, *106*, 647.
- 5 R. F. W. BADER, *J. Phys. Chem. A* **1998**, *102*, 7314.
- 6 R. F. W. BADER, T. T. NGUYEN-DANG, Y. TAL, *Rep. Prog. Phys.* **1981**, *44*, 893.
- 7 R. F. W. BADER, *Can. J. Chem.* **1998**, *76*, 973.
- 8 R. F. W. BADER, *Phys. Rev. B* **1994**, *49*, 13348.
- 9 C. F. MATTA, R. F. W. BADER, *Proteins Structure Function Genetics* **2000**, *40*, 310.
- 10 D. S. KOSOV, P. A. L. POPELIER, *J. Chem. Phys.* **2000**, *113*, 3969.
- 11 D. S. KOSOV, P. L. A. POPELIER, *J. Phys. Chem. A* **2000**, *104*, 7339.
- 12 P. A. L. POPELIER, L. JOUBERT, D. S. KOSOV, *J. Phys. Chem. A* **2001**, *105*, 8254.
- 13 C. F. MATTA, R. F. W. BADER, *Proteins Structure Function Genetics*, **2002**, submitted.
- 14 F. J. MARTIN, PhD Thesis, *Theoretical Synthesis of Macromolecules From Transferable Functional Groups*, McMaster University, 2001.
- 15 C. F. MATTA, R. F. W. BADER, *Proteins Structure Function Genetics*, **2002**, *48*, 519.
- 16 R. F. W. BADER, M. T. CARROLL, J. R. CHEESEMAN, C. CHANG, *J. Am. Chem. Soc.* **1987**, *109*, 7968.
- 17 R. F. W. BADER, W. H. HENNEKER, P. E. CADE, *J. Chem. Phys.* **1967**, *46*, 3341.
- 18 R. F. W. BADER, T. A. KEITH, *J. Chem. Phys.* **1993**, *99*, 3683.
- 19 R. F. W. BADER, C. F. MATTA, *Int. J. Quantum Chem.* **2001**, *85*, 592.
- 20 F. J. MILLERO, A. L. SURODO, C. SHIN, *J. Phys. Chem.* **1978**, *82*, 784.
- 21 H.-J. E. HINZ, *Thermodynamic Data for Biochemistry and Biotechnology*, Springer, Berlin, 1986.
- 22 C. R. WOESE, *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 71.
- 23 M. V. VOLKENSTEIN, *Nature* **1965**, *207*, 294.
- 24 M. V. VOLKENSTEIN, *Biochim. Biophys. Acta* **1966**, *119*, 421.
- 25 R. WOLFENDEN, L. ANDERSSON, P. M. CULLIS, C. C. B. SOUTHGATE, *Biochemistry* **1981**, *20*, 849.
- 26 R. F. W. BADER, K. M. GOUGH, K. E. LAIDIG, T. A. KEITH, *Mol. Phys.* **1992**, *75*, 1167.
- 27 A. J. STONE, C. HÄTTIG, G. JANSEN, J. G. ÁNGYÁN, *Mol. Phys.* **1996**, *89*, 595.
- 28 C. HÄTTIG, G. JANSEN, B. A. HESS, J. G. ÁNGYÁN, *Mol. Phys.* **1997**, *91*, 145.
- 29 P. L. A. POPELIER, *J. Phys. Chem. A* **1999**, *103*, 2883.
- 30 R. F. W. BADER, H. ESSÉN, *J. Chem. Phys.* **1984**, *80*, 1943.
- 31 S. E. O'BRIEN, P. L. A. POPELIER, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764.

- 32 R. F. W. BADER, P. L. A. POPELIER, C. CHANG, *J. Mol. Struct. (Theochem.)* **1992**, 255, 145.
- 33 R. CARBO, L. LEYDA, M. ARNAU, *Int. J. Quantum Chem.* **1980**, 17, 1185.
- 34 R. CARBO-DORCA, E. BESALU, *J. Mol. Struct. (Theochem.)* **1998**, 451, 11.
- 35 R. CARBO-DORCA, L. AMAT, E. BESALU, X. GIRONES, D. ROBERT, *J. Mol. Struct. (Theochem.)* **2000**, 504, 181.
- 36 X. GIRONES, L. AMAT, D. ROBERT, R. CARBO-DORCA, *J. Comput.-Aided Mol. Design* **2000**, 14, 477.
- 37 C. F. MATTA, *J. Phys. Chem. A* **2001**, 105, 11088.
- 38 R. F. W. BADER, F. J. MARTIN, *Can. J. Chem.* **1998**, 76, 284.
- 39 R. F. W. BADER, D. BAYLES, *J. Phys. Chem. A* **2000**, 104, 5579.
- 40 S. W. BENSON, F. R. CRUICKSHANK, D. M. GOLDEN, G. R. HAUGEN, H. E. O'NEAL, A. S. RODGERS, R. SHAW, R. WALSH, *Chem. Rev.* **1969**, 69, 279.
- 41 C. CHANG, R. F. W. BADER, *J. Phys. Chem.* **1992**, 112, 6530.
- 42 F. SICHERI, D. S. C. YANG, *Nature* **1995**, 375, 427.
- 43 D. S. C. YANG, W.-H. HON, S. BUBANKO, Y. XUE, J. SEETHARAMAN, C. L. HEW, F. SICHERI, *J. Biophys.* **1998**, 74, 2142.
- 44 R. F. W. BADER, P. J. MACDOUGALL, C. D. H. LAU, *J. Am. Chem. Soc.* **1984**, 106, 1594.
- 45 R. J. GILLESPIE, *Molecular Geometry*, Van Nostrand Reinhold, London, **1972**.
- 46 R. F. W. BADER, M. E. STEPHENS, *J. Am. Chem. Soc.* **1975**, 97, 7391.
- 47 X. FRADERA, M. A. AUSTEN, R. F. W. BADER, *J. Phys. Chem. A* **1999**, 103, 304.
- 48 R. F. W. BADER, G. L. HEARD, *J. Chem. Phys.* **1999**, 111, 8789.
- 49 M. T. CARROLL, J. R. CHEESEMAN, R. OSMAN, H. WEINSTEIN, *J. Phys. Chem.* **1989**, 93, 5120.
- 50 V. TSIRELSON, P. F. ZOU, T.-H. TANG, R. F. W. BADER, *Acta Crystallogr. A* **1995**, 51, 143.
- 51 F. W. BIEGLER-KÖNIG, R. F. W. BADER, T.-H. TANG, *J. Comput. Chem.* **1982**, 13, 317.
- 52 F. W. BIEGLER-KÖNIG, J. SCHÖNBOHM, D. BAYLES, *J. Comput. Chem.* **2001**, 22, 545.
- 53 P. J. MACDOUGALL, C. E. HENZE, *Theor. Chim. Acc.* **2001**, 105, 345.
- 54 M. KEMP, *Nature* **2001**, 412, 588.

8

The Use of the Molecular Electrostatic Potential in Medicinal Chemistry

JANE S. MURRAY and PETER POLITZER

8.1

Introduction

The electrons and nuclei of an atom or molecule (or other system) are charged particles, and therefore create an electrical potential in the surrounding space. Because we normally treat the electrons as a continuous but static distribution of negative charge, with the nuclei forming a rigid framework, the resulting “electrostatic” potential $V(\mathbf{r})$ is, by Coulomb’s law:

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (1)$$

where Z_A is the charge on nucleus A, located at \mathbf{R}_A , and $\rho(\mathbf{r})$ is the electronic density. $V(\mathbf{r})$ is a real physical property, which can be determined experimentally, by diffraction methods [1, 2], and computationally.

The sign of $V(\mathbf{r})$ at any point in space depends on whether the nuclear contribution (positive) or the electronic (negative) dominates. For neutral, spherically-averaged free atoms, $V(\mathbf{r})$ is positive everywhere, decreasing monotonically with radial distance from the nucleus [3, 4]. When atoms interact to form molecules, however, the concomitant polarization of their charge distributions (relatively minor but very important) results in the development of regions of negative potential. These occur primarily:

- near the lone pairs of electronegative atoms such as F, N, O, Cl, Br, etc.;

- near the π regions of unsaturated hydrocarbons, e.g. ethylene, acetylene, benzene, etc.; and
- near the C–C bonds of strained hydrocarbons.

Each such region must have one or more local minima, V_{\min} , at which the potential reaches its most negative values. These have often been used, with some success, to identify and rank sites with regard to reactivity toward electrophiles [5–8]. The converse – taking local maxima as indicative of susceptibility to nucleophilic attack – is not, however, valid. Pathak and Gadre have shown that the only maxima of $V(\mathbf{r})$ are those associated with the nuclei [9]; these do not correlate with reactivity.

The use of the electrostatic potential to interpret and predict molecular reactive behavior stems from the pioneering work of Scrocco and Tomasi [5, 6]; this has been described in very interesting fashion by Tomasi et al. [10]. The insight and increased understanding that could be obtained by this approach quickly led to its adoption by other research groups, and a steadily growing number of studies followed. Initially, $V(\mathbf{r})$ was computed primarily in planes through the molecule and the focus was on locating the V_{\min} , although it soon became apparent that the overall pattern of positive and negative regions in a plane above the molecule could be very helpful in understanding its interactions, especially in “recognition” processes, e.g. between enzymes and substrates or drugs and receptors. (For reviews, see Naray-Szabo and Ferenczy [2], Scrocco and Tomasi [6], and Politzer et al. [7, 11].) Gradually, three-dimensional modes of representing $V(\mathbf{r})$ were introduced, for example by means of contours showing a single selected value [2, 12], or by plotting $V(\mathbf{r})$ on an appropriately-defined molecular surface [8, 13]. Our own applications of the electrostatic potential to problems in medicinal chemistry have evolved similarly, also paralleling remarkable improvements in software and methodology. We have now progressed to the stage of detailed characterization of $V(\mathbf{r})$ on molecular surfaces by evaluation of a group of statistically-defined quantities [14–17]. This will be discussed later in this chapter.

The significance of the electrostatic potential is not limited to reactivity. It is indeed a fundamental quantity, in terms of which such intrinsic atomic and molecular properties as energies and electronegativities can be expressed *rigorously*. (For detailed discussions see Politzer et al. [18–23] and March [24].) In this chap-

ter, however, we will focus on $V(\mathbf{r})$ in relation to molecular interactive behavior, proceeding from a historical perspective and giving examples from our own work of the different modes of analyzing $V(\mathbf{r})$ and applying it in medicinal chemistry.

8.2 Methodology

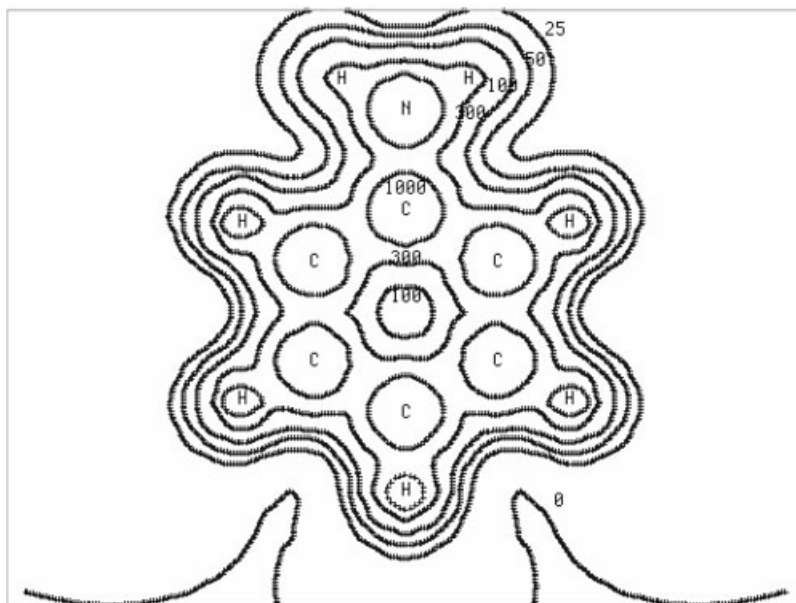
To calculate the molecular electrostatic potential, it is necessary to have a structure for the molecule in question. This can be obtained experimentally or by optimization of geometry by use of software such as Gaussian 98 [25]. A molecular wave function is then computed, from which $V(\mathbf{r})$ can be determined by means of Eq. (1), also by Gaussian 98. This general procedure is classified as a rigorous evaluation of $V(\mathbf{r})$, even though the wave function was necessarily approximate, because Eq. (1) was applied. (Alternative approaches have also been introduced which involve approximations to Eq. (1); these have been described elsewhere [2, 6, 7, 10, 26].) It has been found that a generally satisfactory representation of $V(\mathbf{r})$ in the outer regions of a molecule is produced at the Hartree-Fock level, even with minimum basis sets, and by density-functional and some semi-empirical methods [6–8, 13]. As will be pointed out, it is the outer regions that are relevant to our purposes. Accordingly the results to be discussed in this chapter will be primarily Hartree-Fock, minimum basis set; we do recommend, however, that the latter include polarization functions for second-row atoms, e.g. STO-5G* rather than STO-5G.

Fig. 8.1 shows examples, for aniline (1), of two-dimensional representations of the electrostatic potential, in planes through or above the molecule; this was originally the typical mode of its presentation.

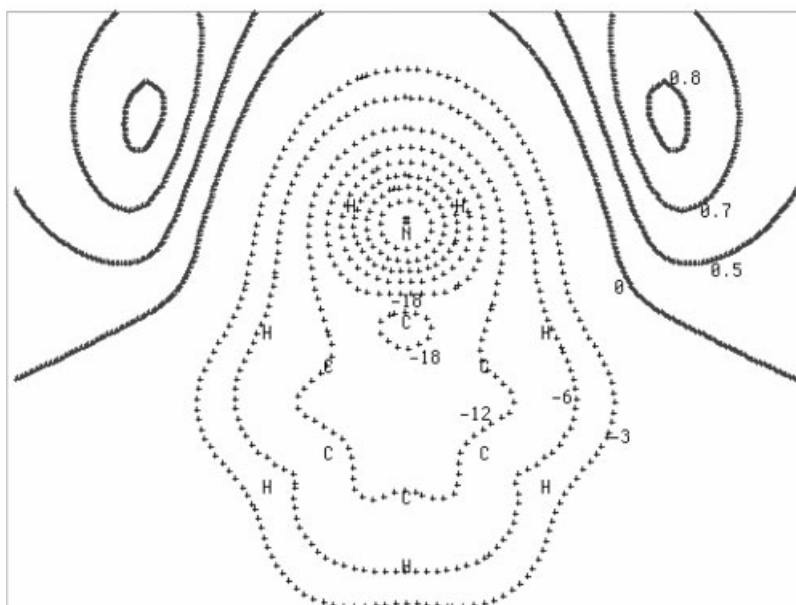


1

Fig. 8.1a shows $V(\mathbf{r})$ computed in the plane of the aromatic ring. It is dominated by positive regions, which are, of course,



a)



b)

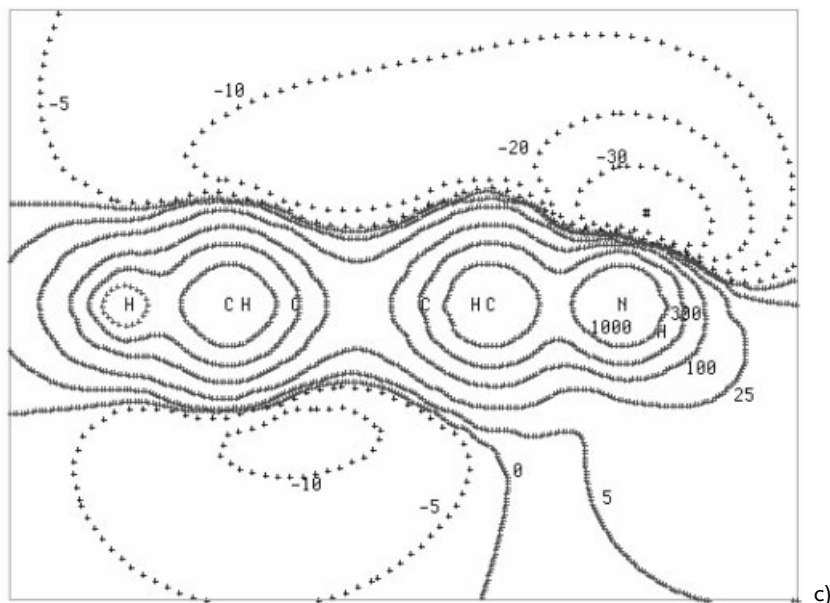


Fig. 8.1 Calculated electrostatic potential for aniline (**1**), in kcal mol⁻¹, in (a) the plane containing the aromatic ring and amine nitrogen, (b) the plane 1.5 Å above the aromatic ring (on the side with the nitrogen lone pair), and (c) the plane perpendicular to the aromatic ring and slicing through the C–N bond and the *para* carbon. The nuclear positions or their pro-

jections are shown by their atomic symbols. In (b), the contours above the amine nitrogen lone pair after -18 are increasingly negative by increments of -6. The V_{\min} in this plane is designated # and has a value of -48 kcal mol⁻¹. In (c), V_{\min} is also designated # and corresponds to the overall spatial minimum of -72 kcal mol⁻¹.

particularly strong around the nuclei that are situated in the plane. Fig. 8.1b shows $V(\mathbf{r})$ in the plane 1.5 Å above the ring, now largely negative. It is fairly uniform above most of the ring, reflecting the π electrons, and reaches its most negative value near the amine nitrogen. A strongly negative region is characteristically associated with electronegative heteroatoms; in this instance, it is because of the concentrated electronic charge of the nitrogen lone pair. Fig. 8.1c shows the potential in a plane perpendicular to the ring, passing through the C–N bond and the *para* carbon. This view shows that the negative regions above and below the ring differ – the former being stronger and overlapping with that produced by the nitrogen lone pair. The infor-

mation in Fig. 8.1 a–c together gives a fairly good picture of $V(\mathbf{r})$ in the space surrounding aniline; the molecule attracts electrophiles through its negative potentials above and below the ring, with the nitrogen being the most favorable site for their approach. It is clear, however, that a misleading picture can emerge if the planes are not chosen judiciously. For example, if one were to plot $V(\mathbf{r})$ only in the plane of the ring, important negative regions would not be seen. Choosing the planes appropriately obviously becomes more difficult when the molecule is less symmetrical.

The values associated with the contours in Fig. 8.1 correspond to the interaction energies of a proton with the *unperturbed* charge distribution of the molecule. It must, of course, be recognized that the latter will not remain unperturbed as the proton approaches. (There have been several attempts to take such polarization effects into account, for instance by means of perturbation theory [7, 10, 27, 28].) Nevertheless, the V_{\min} can be quite effective in ranking protonation sites if these are chemically similar, for example the nitrogens in a series of azines [8, 29, 30]. Problems can arise, however, when the charge-transfer capabilities of the sites inherently differ significantly, e.g. NH_3 compared with PH_3 [31, 32].

The difficulties associated with polarization and charge transfer are exacerbated when the electrophile is larger than a proton. It is generally best, therefore, to view $V(\mathbf{r})$ as indicating the most favorable initial path(s) of approach of an electrophile, or alternatively, as a guide to noncovalent interactions or the early stages of bond-forming processes [11, 16, 17, 26, 31–33]. In these situations the separation between the reactants is sufficient to minimize polarization and charge transfer. Thus the focus should be upon $V(\mathbf{r})$ in the outer regions of the molecule of interest.

An effective approach is to compute $V(\mathbf{r})$ on an appropriately-defined molecular surface, because this is what is “seen” or “felt” by the other reactant. Such a surface is of course arbitrary, because there is no rigorous basis for it. A common procedure has been to use a set of fused spheres centered on the individual nuclei, with van der Waals or other suitable radii [34–37]. We prefer, however, to follow the suggestion of Bader et al. [38] and take the molecular surface to correspond to an outer contour of the electronic density. This has the advantage of reflecting features such

as lone pairs and strained bonds that are specific to a particular molecule but which would not be shown by fused spheres.

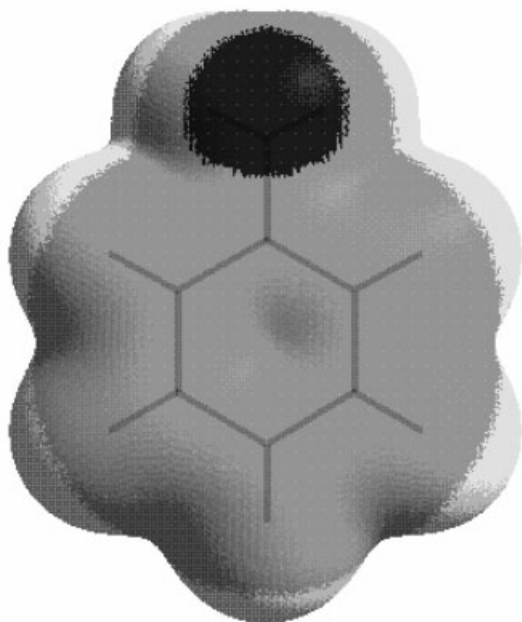
Fig. 8.2 depicts the electrostatic potential computed on the $\rho(\mathbf{r})=0.001$ electrons bohr⁻³ contour encompassing aniline. (Other low-value contours, e.g. $\rho(\mathbf{r})=0.002$ electrons bohr⁻³, serve equally well [39].) Fig. 8.2a shows the side having the nitrogen lone pair, and can be compared with Fig. 8.1b. In Fig. 8.2b, which is the other side of the ring, we see that the amine hydrogens are the most positive, i.e. have the strongest surface maxima. (Pathak and Gadre's proof that the only three-dimensional maxima of $V(\mathbf{r})$ are those associated with the nuclei [9], mentioned in Section 8.1, does not preclude maxima on the molecular surface.) Thus the advantages of computing the potential on surfaces such as that in Fig. 8.2 include:

- largely avoiding the problems associated with polarization and charge transfer;
- being able to treat reactivity toward nucleophiles as well as electrophiles; and
- gaining an appreciation of the actual shape of the molecule.

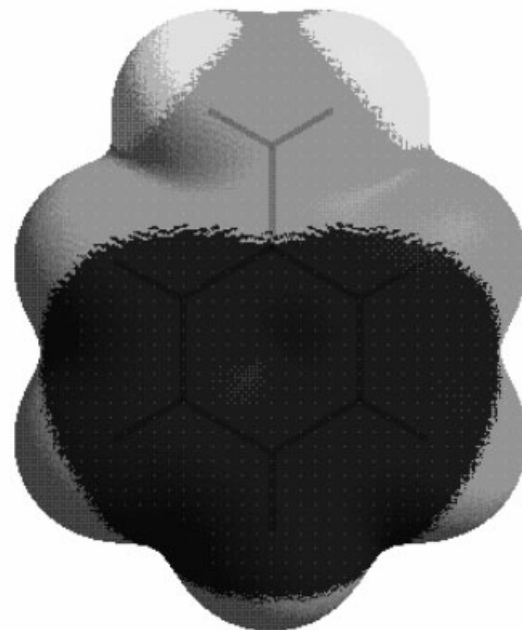
8.3

An Example that Focuses on V_{\min} – the Carcinogenicity of Halogenated Olefins and their Epoxides

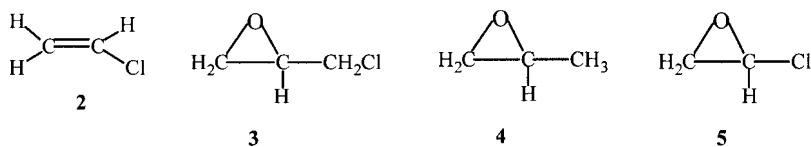
One of our earliest applications of the electrostatic potential in medicinal chemistry was in the screening of halogenated olefins and epoxides for suspect carcinogens. These compounds have been used in the manufacture of textiles and plastics, as solvents, pesticides, dry-cleaning fluids, refrigerants, flame retardants, and degreasing agents, and as intermediates in syntheses [40]. Some, including vinyl chloride (2), epichlorohydrin (3) and propylene oxide (4) [40–42], among others, are carcinogenic. Vinyl chloride, for example, has been shown to cause liver angiosarcomas, lung adenocarcinomas, and other tumors in animals [43, 44], and has been implicated in liver, lung, and lymphatic cancers in man [45, 46].



a)



b)



The carcinogenic behavior is believed to depend upon four factors:

- *Ease of olefin epoxidation.* The carcinogenic forms of the halogenated olefins are believed to be the corresponding epoxides, produced metabolically by microsomal monooxygenases [40–42, 47, 48]. For example, chlorooxirane (5) is the metabolite believed to be responsible for the toxicity of vinyl chloride. Accordingly, in screening olefins for carcinogenicity, we focused on their epoxides.
- *Tendency for epoxide protonation.* When the epoxide is formed *in vivo* it is believed that protonation of the oxygen occurs. This idea is supported by the fact that there are subcellular regions with high proton activity [49–52]. Protonation of an epoxide oxygen weakens the C–O bonds and promotes ring opening [53–57].
- *Reactivity of protonated epoxide.* There is believed to be a zone of intermediate epoxide reactivity that is optimum for initiating the carcinogenic process [57–59]. The molecule must be able to undergo the necessary reactions, yet not be so active it will interact prematurely with other cellular species.
- *Nature of the critical cellular adduct.* It is to be expected that the properties and behavior of the adduct formed with DNA will be of key importance. For example, the carcinogenicity of vinyl chloride has been attributed to its 7-*N*-(2-oxoethyl) derivative of guanine [60, 61].

Fig. 8.2 Calculated electrostatic potential on the molecular surface of aniline (1), in kcal mol⁻¹. (a) The side of the aromatic ring with the nitrogen lone pair. Ranges: black is more negative than -25; gray is between -25 and 0; off-white is between 0 and 20. (b) The side of the aromatic ring opposite to the nitro-

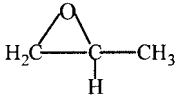
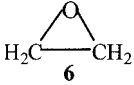
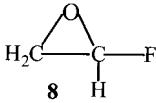
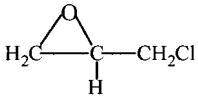
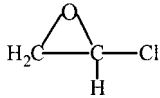
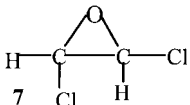
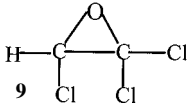
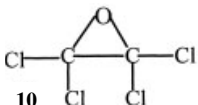
gen lone pair. Ranges: black is between -25 and 0; gray is between 0 and 20; off-white is more positive than 20. The constraint of being limited to black, gray, or white necessitates having different ranges in (a) and (b). In color, greater resolution is, of course, possible.

In developing a screening technique for determining whether a particular epoxide is likely to be carcinogenic, we addressed the second of the four factors mentioned above, its tendency to undergo oxygen protonation. For approximately thirty epoxides we computed electrostatic potentials in planes perpendicular to the three-membered rings and passing through the oxygens. We found there are usually two V_{\min} near each oxygen, reflecting its two lone pairs [62–64]. The values obtained at the HF/STO-5G level for some of these molecules are listed in Tab. 8.1. The magnitudes of the oxygen minima reflect the nature of the substituents on the epoxide ring. As a reference point we can take the totally unsubstituted ethylene oxide (6), for which $V_{\min} = -51 \text{ kcal mol}^{-1}$. Tab. 8.1 shows that a weakly electron-donating substituent, e.g. the methyl in propylene oxide (4), results in V_{\min} becoming slightly more negative. Electron-withdrawing substituents such as chlorine and fluorine have the opposite effect, which increases with the extent of halogenation (compare the V_{\min} of 5 and 7) and is stronger when the substitution is directly on the ring rather than on an exocyclic carbon (compare the V_{\min} of 3 and 5). Consistent with other studies of halogenated compounds [64–66], chlorine weakens the oxygen V_{\min} substantially more than does fluorine (compare 5 and 8), despite the greater intrinsic electronegativity of fluorine. This can be explained by invoking chlorine's greater charge capacity [67, 68], which enables it to accept more charge than can the small fluorine atom.

When the epoxides are ranked in order of their V_{\min} , as in Tab. 8.1, it is found that the established carcinogens are those with the more negative values, e.g. 3–6 and 8. The inactive or weakly active carcinogens have less negative V_{\min} ; examples in Tab. 8.1 include 7, 9 and 10. This became our criterion for deciding whether a particular epoxide was likely to be carcinogenic [62, 64]. Epoxides with V_{\min} more negative than a threshold potential of approximately $-30 \text{ kcal mol}^{-1}$ (HF/STO-5G) were viewed as suspect and recommended for further investigation using animal testing procedures. In the course of this study we computed the oxygen V_{\min} for more than sixty epoxides, many of them requested by the Environmental Protection Agency.

The use of V_{\min} in this manner does not encounter the problem of inherent differences in charge-transfer capacities (Sec-

Tab. 8.1 Calculated oxygen V_{min} for some epoxides.^a

Molecule	V_{min} (kcal mol ⁻¹)
 4	-53.4
 6	-51.3
 8	-47.7
 3	-43.1
 5	-38.1
 7	-23.1
 9	-17.1
 10	-9.2

a) Data taken from [64].

tion 8.2), because the sites being compared are all epoxide oxygens. The variations in V_{\min} are therefore relatively minor and a result only of the groups attached to the ring.

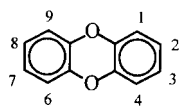
8.4

An Example Focusing on the General Patterns of Molecular Electrostatic Potentials – Toxicity of Dibenzo-*p*-dioxins and Analogs

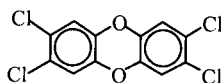
The electrostatic potential is highly suitable for analyzing processes in which the initial step is the “recognition” by some system, such as an enzyme or receptor, that an approaching molecule, e.g. a substrate or drug, has certain key features that will promote (or hinder) their interaction, which is electrostatic in the early stages. For this purpose $V(\mathbf{r})$ is computed in the outer regions of the molecule, perhaps in a plane but preferably on its surface, because this is what the enzyme, receptor, etc., “sees” or “feels.” There have been numerous such studies, some of which have been summarized in a variety of reviews [1, 2, 6, 7, 11, 69].

An example from our work involved dibenzo-*p*-dioxin (**11**), its halogenated derivatives, and some analogs. Whereas the parent compound **11** is non-toxic, several of its halogenated derivatives (which can be substituted at various combinations of positions 1–4 and 6–9) have a wide range of toxicity, the most severe being that of the notorious 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD, **12**). TCDD, and some of the others, have been linked with carcinogenesis, hepatotoxicity, gastric lesions, loss of lymphoid tissue, urinary tract hyperplasia, chloracne, and acute loss of weight [70]. The mechanism by which these effects occur is believed to involve initial interaction with a cytosolic receptor, which has been suggested to be porphine-like (**13**) [71].

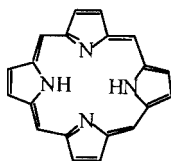
It was pointed out by Poland and Knutson that some molecular structural features seem to be associated with high levels of both toxicity and receptor binding [70]. These include planarity and rectangular shape, halogenation at a minimum of three of the four lateral positions (2, 3, 7, and 8), but at least one unsubstituted ring position. Activity increases from fluorine to bromine.



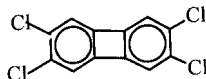
11



TCDD, 12



13



14

We have analyzed the electrostatic potentials of a group of twelve dibenzo-*p*-dioxins, with different amounts of substitution and toxicity, and several analogs, e.g. **14**. In most of these studies, $V(\mathbf{r})$ was computed in planes that were 1.75 Å above the nuclear framework and perpendicular to the molecular plane and passing through the oxygens [64, 72–75], but eventually we progressed to using molecular surfaces [76]. The same general patterns were observed by use of both approaches.

The parent, unsubstituted molecule, **11**, has strong and extensive negative regions associated with the oxygen lone pairs, and weaker ones above and below the aromatic rings, because of their π electrons. Halogen substitution significantly weakens the former and eliminates the latter. Thus the potential on the molecular surface of TCDD (**12**) is negative above the chlorines, quite weakly so above the oxygens, and positive everywhere else.

We found this $V(\mathbf{r})$ pattern to be characteristically associated with high toxicity and receptor binding – negative regions above all or most of the lateral positions at the ends of the molecule separated by a large positive area, with (at most) limited and weak negative potentials above the oxygens. The latter are apparently not even needed, because **14** (which otherwise fits the TCDD pattern very well) has biological activity similar to that of TCDD. The reason why the oxygen potentials seem to inhibit activity, if not considerably weaken it, has been suggested by modeling studies [71] to be their unfavorable interaction with the lone pairs of the receptor's nitrogens (**13**).

8.5

Statistical Characterization of the Molecular Surface Electrostatic Potential – the General Interaction Properties Function (GIPF)

In the final stage of our dibenzo-*p*-dioxin work, summarized in Section 8.4, we computed $V(\mathbf{r})$ on the molecular surfaces, which shall henceforth be designated $V_S(\mathbf{r})$, but our analysis was essentially qualitative. As we began to focus increasingly upon $V_S(\mathbf{r})$, however, while proceeding to other projects, we became interested in retrieving, quantitatively, more of the wealth of information it contains. The obvious starting point was the extrema of $V_S(\mathbf{r})$, i.e. $V_{S,\max}$ and $V_{S,\min}$. We were able to relate these to hydrogen-bond-donating and -accepting tendencies; specifically, we showed that established empirical measures of hydrogen-bond acidity and basicity can be expressed analytically in terms of $V_{S,\max}$ and $V_{S,\min}$, respectively [13, 26, 77]. (For an overview of this and other aspects of our applications of surface potentials to hydrogen bonding, see Politzer and Murray [26].)

In progressing beyond $V_{S,\max}$ and $V_{S,\min}$ our first objective was to find a reasonable measure of internal charge separation (local polarity). Although the dipole moment is often used for this purpose, it has serious limitations, most notably that it can be zero, because of symmetry, even for many molecules that clearly have considerable charge separation, good examples being **11** and **12**. Accordingly we proposed the quantity Π , the average deviation of $V_S(\mathbf{r})$, defined by [78]:

$$\Pi = \frac{1}{n} \sum_{i=1}^n |V_S(\mathbf{r}_i) - \bar{V}_S| \quad (2)$$

where \bar{V}_S is the average of $V_S(\mathbf{r})$ over a grid of n points covering the entire surface:

$$\bar{V}_S = \frac{1}{n} \sum_{i=1}^n V_S(\mathbf{r}_i) \quad (3)$$

We found that Π varied between approximately 2.0 for alkane hydrocarbons to the mid-twenties for molecules with several strongly electron-withdrawing groups, e.g. some polynitro derivatives [79]. Equation (2) shows it to be size-independent, unlike

the dipole moment; thus $\Pi=21.6$ for the very small H_2O molecule (HF/STO-5G*) [78]. We have demonstrated that Π correlates with several empirical indices of polarity [14, 78].

The introduction of Π as an analytical tool enabled us to identify a significant feature common to several different families of drugs we have investigated:

- anticonvulsants of various kinds, e.g. hydantoins, barbiturates, carbamazepines, succinimides, etc. [80];
- tetracyclines (antibiotics) [81];
- reverse transcriptase inhibitors (anti-HIV agents) [82]; and
- cocaine analogues [83].

We found that each of these types of drug, which encompass quite a diversity of molecules, is characterized by a markedly limited range of Π values. For example, 16 of the 19 anticonvulsants have Π between 10.0 and 13.0. The situation for the tetracyclines is similar; for the reverse transcriptase inhibitors and the cocaine analogs the magnitudes of Π are somewhat smaller but again within a narrow range. Thus the internal charge separations effective for each type of activity are quite restricted. This suggests that each drug type requires an optimum balance between hydrophilicity and hydrophobicity, to enable the necessary migration between media of different polarity.

After Π , we gradually introduced several more statistically-defined quantities, designed to aid more complete characterization of $V_S(\mathbf{r})$ over the entire molecular surface [14–17, 84]. These include the average positive and negative potentials, \bar{V}_S^+ and \bar{V}_S^- :

$$\bar{V}_S^+ = \frac{1}{a} \sum_{j=1}^a V_S^+(\mathbf{r}_j) \quad (4)$$

$$\bar{V}_S^- = \frac{1}{\beta} \sum_{k=1}^{\beta} V_S^-(\mathbf{r}_k) \quad (5)$$

the positive, negative, and total variances, σ_+^2 , σ_-^2 , and σ_{tot}^2 :

$$\sigma_{\text{tot}}^2 = \sigma_+^2 + \sigma_-^2 = \frac{1}{a} \sum_{j=1}^a [V_S^+(\mathbf{r}_j) - \bar{V}_S^+]^2 + \frac{1}{\beta} \sum_{k=1}^{\beta} [V_S^-(\mathbf{r}_k) - \bar{V}_S^-]^2 \quad (6)$$

and an “electrostatic balance” term, ν :

$$\nu = \frac{\sigma_+^2 \sigma_-^2}{[\sigma_{\text{tot}}^2]^2} \quad (7)$$

The purpose of σ_+^2 , σ_-^2 , and σ_{tot}^2 is to indicate how variable are $V_S^+(\mathbf{r})$, $V_S^-(\mathbf{r})$ and $V_S(\mathbf{r})$. Although the formula for σ_{tot}^2 seems to somewhat resemble that for Π , the presence of the squared terms in Eq. (6) gives σ_{tot}^2 the possibility of having much larger magnitudes. The two quantities actually deal with quite different aspects of $V(\mathbf{r})$, and often do not even vary in the same direction [14, 16, 79, 84]. Finally, ν is viewed as a measure of the balance between the strengths of the positive and negative surface potentials. Eq. (7) shows that the upper limit of ν is 0.250, which is reached when $\sigma = \sigma_-^2$. Thus, as ν approaches 0.250 the molecule is increasingly able to interact to a similar extent through both its positive and negative regions (whether that be strongly or weakly).

The quantities defined by Eqs. (2)–(7) plus $V_{S,\text{max}}$, $V_{S,\text{min}}$, and the positive and negative areas, A_S^+ and A_S^- , enable detailed characterization of the electrostatic potential on a molecular surface. Over the past ten years, we have shown that subsets of these quantities can be used to represent analytically a variety of liquid-, solid-, and solution-phase properties that depend on noncovalent interactions [14–17, 84]; these include boiling points and critical constants, heats of vaporization, sublimation and fusion, solubilities and solvation energies, partition coefficients, diffusion constants, viscosities, surface tensions, and liquid and crystal densities.

Conceptually, we view our approach in terms of a general interaction properties function (GIPF):

$$\text{Property} = f[V_{S,\text{min}}, V_{S,\text{max}}, \bar{V}_S^+, \bar{V}_S^-, \Pi, \sigma_+^2, \sigma_-^2, \sigma_{\text{tot}}^2, \nu, A_S^+, A_S^-] \quad (8)$$

It should be emphasized, however, that most of our relationships involve, in different combinations, only three or four of the quantities on the right of Eq. (8). To develop an expression to represent a particular property we need a database of experimental values for it. For each compound in the database we compute $V(\mathbf{r})$ and all of the variables in Eq. (8). A statistical analysis package is then used to identify a subset of these to which the experi-

mental data can be fit with satisfactory accuracy. We use as few variables as possible, because one of our objectives is to gain insight into the key factors that determine the property. We also seek maximum generality, i.e. to cover a wide variety of compounds, even though the correlations would undoubtedly be better if different chemical classes were treated separately, e.g. hydrocarbons, alcohols, etc. Our correlation coefficients are typically between 0.95 and 0.99.

As examples, our GIPF relationships for diffusion constants in gelatin [85] and the solubility of C₆₀ in twenty organic solvents [86] are given in Eqs. (9) and (10):

$$\text{Diff. const.} = a_1 A_S^- - \beta_1 \sigma_+^2 + \gamma_1 \sigma_-^2 - \delta_1 \quad (9)$$

$$\text{Log}(C_{60} \text{ solub.}) = -a_2 [\sigma_{\text{tot}}^2 / A_S^{1.5}] + \beta_2 [v \sigma_{\text{tot}}^2]^{0.5} + \gamma_2 A_S^4 - \delta_2 \quad (10)$$

A_S is the total surface area and $a_i \dots \delta_i$ are all positive. The correlation coefficients for Eqs. (9) and (10) are 0.990 and 0.954, respectively. When applying such expressions it is, of course, imperative that the surface quantities are evaluated at the same computational levels as were used in their development.

Up to this point, GIPF expressions have been formulated for only one type of biological activity – the inhibition of reverse transcriptase (RT), the enzyme that promotes the reverse transcription of genomic RNA into double-stranded DNA, a key step in the replication of the human immunodeficiency virus, HIV [82, 87]. Analytical representations were obtained for the anti-HIV potencies of three families of RT inhibitors; the correlation coefficients are between 0.930 and 0.952. We are currently investigating the effects of applying the GIPF approach to certain portions of the molecules rather than their entireties. This might reveal the source of the activity, or alternatively, indicate it to be delocalized.

An important feature of the GIPF technique is that calculations for single molecules suffice to establish quantitative relationships for liquid-, solid-, and solution-phase properties, without taking explicit account of the media. This is clearly of considerable practical significance. Finally, it should be mentioned that changes in molecular conformation generally have little effect upon our computed surface quantities [88], unless some change in hydrogen bonding is involved.

8.6

Summary

The electrostatic potential, especially when computed on molecular surfaces, is a powerful tool for analyzing and interpreting reactive behavior. It is particularly effective for noncovalent interactions and the early stages of processes that eventually involve bond formation and/or rupture. We have shown that the potential has qualitative and quantitative predictive capacity. Because it can be obtained purely computationally, using optimized molecular geometries, the potential can be used to characterize compounds that have not yet been synthesized, and to design them to have specific desired features [89, 90]. With the continuing remarkable developments in methodology, software, and processor technology, the different applications of the electrostatic potential can be expected to increase further in scope and reliability, in medicinal chemistry as in other areas.

8.7

Acknowledgment

We greatly appreciate the assistance of Pat Lane in preparing the figures.

8.8

References

- 1 P. POLITZER, D. G. TRUHLAR (eds), *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
- 2 G. NARAY-SZABO, G. G. FERENCZY, *Chem. Rev.* **1995**, 95, 829.
- 3 H. WEINSTEIN, P. POLITZER, S. SREBRENİK, *Theor. Chim. Acta* **1975**, 38, 159.
- 4 K. SEN, P. POLITZER, *J. Chem. Phys.* **1989**, 90, 4370.
- 5 E. SCROCCO, J. TOMASI, in *Topics in Current Chemistry*, Vol. 42, Springer, Berlin, **1973**, p. 95.
- 6 E. SCROCCO, J. TOMASI, *Adv. Quantum Chem.* **1978**, 11, 115.
- 7 P. POLITZER, K. C. DAIKER, in *The Force Concept in Chemistry*, Chapter 6, B. M. DEB (ed.), Van Nostrand Reinhold, New York, **1981**.

- 8 P. POLITZER, J. S. MURRAY, in *Reviews in Computational Chemistry*, Chapter 7, Vol. 2, K. B. LIPKOWITZ, D. B. BOYD (eds), VCH, New York, 1991.
- 9 R. K. PATHAK, S. R. GADRE, *J. Chem. Phys.* **1990**, 93, 1770.
- 10 J. TOMASI, B. MENNUCCI, R. CAMMI, in *Molecular Electrostatic Potentials*, Chapter 1. J. S. MURRAY, K. SEN (eds) Elsevier, Amsterdam, 1996.
- 11 P. POLITZER, P. R. LAURENCE, K. JAYASURIYA, *Environ. Health Perspect.* **1985**, 61, 191.
- 12 J. B. MATTHEW, in *Theoretical Biochemistry and Molecular Biophysics*, Vol. 2: *Proteins*, D. L. BEVERIDGE, R. LAVERY (eds), Adenine, Schenectady, NY, 1991, p. 107.
- 13 J. S. MURRAY, P. POLITZER, in *Molecular Orbital Calculations for Biological Systems*, Chapter 3, A.-M. SAPSE (ed.), Oxford University Press, New York, 1998.
- 14 J. S. MURRAY, T. BRINCK, P. LANE, K. PAULSEN, P. POLITZER, *J. Mol. Struct. (Theochem.)* **1994**, 307, 55.
- 15 J. S. MURRAY, P. POLITZER, *J. Mol. Struct. (Theochem.)* **1998**, 425, 107.
- 16 P. POLITZER, J. S. MURRAY, *Trends Chem. Phys.* **1999**, 7, 157.
- 17 P. POLITZER, J. S. MURRAY, *Fluid Phase Equil.* **2001**, 185, 129.
- 18 P. POLITZER, *Israel J. Chem.* **1980**, 19, 224.
- 19 P. POLITZER, In *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Chapter 2, P. POLITZER, D. G. TRUHLAR (eds), Plenum, New York, 1981.
- 20 P. POLITZER, in *The Single-Particle Density in Physics and Chemistry*, Chapter 3, N. H. MARCH, B. M. DEB (eds), Academic, London, 1987.
- 21 P. POLITZER, J. S. MURRAY, in *Molecular Electrostatic Potentials*, Chapter 16, J. S. MURRAY, K. SEN (eds), Elsevier, Amsterdam, 1996.
- 22 P. POLITZER, P. LANE, J. S. MURRAY, in *Reviews in Modern Quantum Chemistry: A Celebration of the Contributions of R. G. Parr*, K. D. SEN (ed.), World Scientific, Singapore, 2002, in press.
- 23 P. POLITZER, J. S. MURRAY, *Theor. Chem. Acc.* **2002**, 108, 134.
- 24 N. H. MARCH, APPENDIX 6.1, *Electron Density Theory of Atoms and Molecules*, Academic, London, 1992.
- 25 M. J. FRISCH, G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J. R. CHEESEMAN, V. G. ZAKREZEWSKI, J. A. MONTGOMERY, R. E. STRATMANN, J. C. BURANT, S. DAPPICH, J. M. MILLAM, A. D. DANIELS, K. N. KUDIN, M. C. STRAIN, O. FARKAS, J. TOMASI, V. BARONE, M. COSSI, R. CAMMI, B. MENNUCCI, C. POMELLI, C. ADAMO, S. CLIFFORD, J. OCHTERSKI, G. PETERSSON, P. Y. AAYALA, Q. CUI, K. MOROKUMA, D. K. MALICK, A. D. RUBUCK, K. RAGHAVACHARI, J. B. FORESMAN, J. CIOSLOWSKI, J. V. ORTIZ, B. B. STEFANOV, G. LIU, A. LIASHENKO, P. PISKORZ, I. KOMAROMI, R. GOMPERS, R. L. MARTIN, D. J. FOX, T. KEITH, M. A. AL-LAHAM, C. Y. PENG, A. NANAYAKKARA, C. GONZALEZ, M. CHALLACOMBE, P. M. W. GILL, B. G. JOHNSON, W. CHEN, M. W. WONG, J. L. ANDRES, M. HEAD-GORDON, E. S. REPLOGLE, J. A. POPLE, Gaussian 98, Revi-

- sion A.5, Gaussian Inc., Pittsburgh, PA, 1998.
- 26 P. POLITZER, J. S. MURRAY, in *Computational Medicinal Chemistry and Drug Discovery*, J. P. POLLENAERE, P. BULTINCK, H. DEWINTER, W. LANGENAEKER (eds), Marcel Dekker, New York, 2002, in press.
- 27 M. M. FRANCL, *J. Phys. Chem.* 1985, 89, 428.
- 28 I. ALKORTA, J. J. PEREZ, H. O. VILLAR, *J. Mol. Graphics* 1994, 12, 3.
- 29 J. S. MURRAY, J. M. SEMINARIO, P. POLITZER, *J. Mol. Struct. (Theochem.)* 1989, 187, 95.
- 30 P. NAGY, K. NOVAK, G. SZASZ, *J. Mol. Struct. (Theochem.)* 1989, 201, 257.
- 31 T. BRINCK, J. S. MURRAY, P. POLITZER, *Int. J. Quantum Chem.* 1993, 48, 73.
- 32 J. S. MURRAY, P. POLITZER, in *Theoretical Organic Chemistry*, Chapter 7, C. PARKANYI (ed.), Elsevier, Amsterdam, 1998.
- 33 P. POLITZER, J. S. MURRAY, M. C. PONCHA, *Int. J. Quantum Chem.* 2002, 88, 19.
- 34 M. L. CONNOLLY, *J. Am. Chem. Soc.* 1985, 107, 1118.
- 35 Q. DU, G. A. ARTECA, *J. Comput. Chem.* 1996, 17, 1258.
- 36 J. BRICKMANN, T. EXNER, M. KEIL, R. MARHOFER, G. MOECKEL, in *Encyclopedia of Computational Chemistry*, Vol. 3, P. v. R. SCHLEYER (ed.), Wiley, New York, 1998, 1678.
- 37 M. L. CONNOLLY, in *Encyclopedia of Computational Chemistry*, Vol. 3, P. v. R. SCHLEYER (ed.), Wiley, New York, 1998, 1698.
- 38 R. F. W. BADER, M. T. CARROLL, J. R. CHEESEMAN, C. CHANG, *J. Am. Chem. Soc.* 1987, 109, 7968.
- 39 J. S. MURRAY, T. BRINCK, M. E. GRICE, P. POLITZER, *J. Mol. Struct. (Theochem.)* 1992, 256, 29.
- 40 Y.-T. WOO, D. LAI, J. C. ARCOS, M. F. ARGUS, *Chemical Induction of Cancer*, Vol. IIIB, Section 5.2.2.1, Academic Press, New York, 1984.
- 41 H. BARTSCH, C. MALAVEILLE, A. BARBIN, G. PLANCHE, *Arch. Toxicol.* 1979, 41, 249.
- 42 H. M. BOLT, R. J. LAIB, J. G. FILSER, *Biochem. Pharmacol.* 1982, 31, 1.
- 43 P. L. VIOLA, A. BIOGOTTI, A. CAPUTO, *Cancer Res.* 1971, 31, 516.
- 44 C. MALTONI, *Environ. Health Perspect.* 1977, 21, 1.
- 45 R. J. WAXWEILER, W. STRINGER, J. K. WAGONER, J. JONES, H. FALK, C. CARTER, *Ann. NY Acad. Sci.* 1976, 271, 40.
- 46 R. RAWLS, *Chem. Eng. News* 1980, April 7, 27–28.
- 47 S. OSTERMAN-GOLKAR, D. HULTMARK, D. SEGERBACK, C. J. CALLEMAN, R. GOTHE, *Biochem. Biophys. Res. Commun.* 1977, 76, 259.
- 48 S. BANERJEE, B. L. VAN DUUREN, *Cancer Res.* 1978, 38, 776.
- 49 A. D. MCLAREN, *Enzymologia* 1960, 21, 356.
- 50 A. SOLS, R. MARCO, *Curr. Top. Cell Regul.* 1970, 2, 227.
- 51 G. R. WELCH, M. N. BERRY, in *Coherent Excitations in Biological Systems*, Springer, New York, 1983, p. 95.
- 52 F. P. GUENGERICH, T. L. MACDONALD, *Acc. Chem. Res.* 1984, 17, 9.
- 53 R. T. MORRISON, R. N. BOYD, *Organic Chemistry*, 3rd edn, Allyn and Bacon, Boston, 1973.
- 54 P. POLITZER, K. C. DAIKER, V. M. ESTES, M. BAUGHMAN, *Int. J.*

- Quantum Chem., Quantum Biol. Symp.* **1978**, 5, 291.
- 55 P. POLITZER, V.M. ESTES, in *Catalysis in Chemistry and Biochemistry: Theory and Experiment*, B. PULLMAN (ed.), Reidel, Dordrecht, The Netherlands, **1979**, p. 305.
- 56 J.E. FERRELL, JR, G.H. LOEW, *J. Am. Chem. Soc.* **1979**, 101, 1358.
- 57 P.R. LAURENCE, T.R. PROCTOR, P. POLITZER, *Int. J. Quantum Chem.* **1984**, 26, 425.
- 58 R.B. JONES, W.C. MACKRODT, *Biochem. Pharmacol.* **1982**, 31, 3710.
- 59 Y.-T. WOO, J.C. ARCOS, D. LAI, *Handbook of Carcinogenic Testing*, Noyes, Park Ridge, NJ, **1985**.
- 60 E. SHERER, C.J. VAN DER LAKEN, L.M. GWINNER, R.J. LAIB, P. EMMELOT, *Carcinogenesis* **1981**, 2, 671.
- 61 R.J. LAIB, L.M. GWINNER, H.M. BOLT, *Chem. Biol. Interact.* **1981**, 37, 219.
- 62 P. POLITZER, P.R. LAURENCE, *Carcinogenesis* **1984**, 5, 845.
- 63 P. POLITZER, P.R. LAURENCE, *Int. J. Quantum Chem., Quantum Biol. Symp.* **1984**, 11, 155.
- 64 P. POLITZER, *Toxicol. Lett.* **1988**, 43, 257.
- 65 J.S. MURRAY, P. POLITZER, *Chem. Phys. Lett.* **1988**, 152, 364.
- 66 J.S. MURRAY, T. BRINCK, M.E. GRICE, P. POLITZER, *J. Mol. Struct. (Theochem)*, **1992**, 256, 29.
- 67 J.E. HUHEEY, *J. Phys. Chem.* **1965**, 69, 3284.
- 68 P. POLITZER, J.E. HUHEEY, J.S. MURRAY, M. GRODZICKI, *J. Mol. Struct. (Theochem.)* **1992**, 259, 99.
- 69 J.S. MURRAY, K. SEN (eds), *Molecular Electrostatic Potentials*, Amsterdam, Elsevier, **1996**.
- 70 A. POLAND, J.C. KNUTSON, *Ann. Rev. Pharmacol. Toxicol.* **1982**, 22, 517.
- 71 J.D. MCKINNEY, T. DARDEN, J.A. LYERLY, L.G. PEDERSEN, *Quant. Struct. Act. Relat.* **1985**, 4, 166.
- 72 J.S. MURRAY, B.A. ZILLES, K. JAYASURIYA, P. POLITZER, *J. Am. Chem. Soc.* **1986**, 108, 915.
- 73 J.S. MURRAY, P. POLITZER, *Theor. Chim. Acta* **1987**, 72, 507.
- 74 J.S. MURRAY, P. EVANS, P. POLITZER, *Int. J. Quantum Chem.* **1990**, 37, 271.
- 75 P. POLITZER, J.S. MURRAY, In *Theoretical Biochemistry and Molecular Biophysics: A Comprehensive Survey*, Vol. 2, D.L. BEVERIDGE, R. LAVERY (eds), Adenine Press, Schenectady, NY, **1992**, Chapter 13.
- 76 P. SJOBERG, J.S. MURRAY, T. BRINCK, P. EVANS, P. POLITZER, *J. Mol. Graphics* **1990**, 8, 81.
- 77 H. HAGELIN, T. BRINCK, M. BERTHELOT, J.S. MURRAY, P. POLITZER, *Can. J. Chem.* **1995**, 73, 483.
- 78 T. BRINCK, J.S. MURRAY, P. POLITZER, *Mol. Phys.* **1992**, 76, 609.
- 79 J.S. MURRAY, P. LANE, P. POLITZER, *Mol. Phys.* **1998**, 93, 187.
- 80 J.S. MURRAY, F. ABU-AWWAD, P. POLITZER, L.C. WILSON, A.S. TROUPIN, R.E. WALL, *Int. J. Quantum Chem.* **1998**, 70, 1137.
- 81 W. HUSSEIN, C.G. WALKER, Z. PERALTA-INGA, J.S. MURRAY, *Int. J. Quantum Chem.* **2001**, 82, 160.
- 82 O. GALVEZ GONZALEZ, J.S. MURRAY, Z. PERALTA-INGA, P. POLITZER, *Int. J. Quantum Chem.* **2001**, 83, 115.

- 83 Y. MA, J. S. MURRAY, P. POLITZER, unpublished work.
- 84 P. POLITZER, J. S. MURRAY, In *Recent Research Developments in Biological Applications of Quantum Chemistry Theory*, C. KUBLI-GARFIAS (ed.), Research Signpost, Trivandrum, India, 2002, in press.
- 85 P. POLITZER, J. S. MURRAY, P. FLODMARK, *J. Phys. Chem.* **1996**, *100*, 5538.
- 86 J. S. MURRAY, S. G. GAGARIN, P. POLITZER, *J. Phys. Chem.* **1995**, *99*, 12081.
- 87 P. POLITZER, J. S. MURRAY, Z. PERALTA-INGA, *Int. J. Quantum Chem.* **2001**, *85*, 676.
- 88 J. S. MURRAY, Z. PERALTA-INGA, P. POLITZER, *Int. J. Quantum Chem.* **1999**, *75*, 267.
- 89 P. POLITZER, J. S. MURRAY, M. C. CONCHA, T. BRINCK, *J. Mol. Struct. (Theochem.)* **1993**, *281*, 107.
- 90 P. POLITZER, J. S. MURRAY, T. BRINCK, P. LANE, in *Immunoanalysis of Agrochemicals: Emerging Technologies*, Chapter 8, J. O. NELSON, A. E. KARU, R. B. WONG (eds), American Chemical Society, Washington, 1995.

9 Applications of Quantum Chemical Methods in Drug Design

HANS-DIETER HÖLTJE and MONIKA HÖLTJE

9.1 Introduction

Quantum mechanics (QM) is an essential and widely used tool in computer-aided drug research. The comprehensive utility of QM is exemplified in this chapter by description of several case studies performed in our laboratory. In Section 9.2.1 we describe a routine which we have used to obtain accurate force-field parameters from results of *ab initio* calculations on small model structures. Probably the most important application of QM is explicit description of the electronic structure of a molecule. In this context, Section 9.2.2.1 demonstrates the consequences of π -delocalization and atom hybridization on molecular geometry and Section 9.2.2.2 highlights the potential importance of QM for conformational analysis. Another important area of QM application is calculation of atomic point charges (Section 9.2.3), which can be used to study binding properties such as intermolecular Coulomb forces, hydrogen bonding, and molecular electrostatic potentials (Section 9.2.4). Section 9.2.5 reports investigations on charge-transfer complexes which can play a role in molecular recognition processes. The chapter is concluded with an outlook. It should be noted that only a subjective selection of typical and fairly simple “every day” implementations is given in this chapter; no attempt is made to survey the comprehensive literature in this field. Rather, we intend to demonstrate the general utility of QM calculations in molecular modeling.

9.2

Application Examples

9.2.1

Force Field Parameters from *Ab Initio* Calculations

Because modeling studies on interaction of drugs with biological targets such as receptors, enzymes, or other biomacromolecules involve large chemical systems, thus molecular mechanics methods must be applied. In molecular mechanics molecules are considered as a collection of masses (atoms) held together by elastic or harmonic forces, analogous to a ball-and-spring model. These forces can be described by potential energy functions for structural features such as bond lengths, bond angles, torsional angles, and non-bonded interactions [1]. Often modelers face situations in which the force field in use lacks parameters for single atoms or particular combinations of atoms. In principle, new parameters can be introduced by analogy with existing values, by trial and error adjustment to fit experimental data, or by the use of quantum mechanics. The effort required to develop the new parameters should be related to the weight of the scientific question under consideration. If accurate parameters are needed, quantum mechanical calculations must be performed. In studies on dopamine D₃-receptor agonists recently performed in our group realistic molecular geometries for some of the ligands were lacking [2]. In detail several bond stretching and angle bending parameters for five-membered heterocycles were undefined in the consistent valence force field (CVFF) of Insight/Discover [3]. A variety of techniques for obtaining force field parameters has been proved. They can be deduced from two main strategies, as shown in the Fig. 9.1 a and b.

Both strategies have been applied successfully to a large variety of molecules. Particular problems arise in the parametrization of cyclic systems. Methods based on geometrical distortions, for example Hopfinger's method A [4], are unsatisfactory, because bond lengths and angles have to be treated independently. This problem is avoided by applying method B. Leonard and Ashman developed the method to parameterize force constants for bond stretching and angle bending for the MM2 force field [5]. They used the diagonal elements of an *ab initio* Hessian to calculate

the corresponding numerical values. The Hessian is the matrix of second derivatives of the energy with respect to Cartesian coordinates of the atoms, and the diagonal elements are the force constants associated with each of the variables. In this context the variables are the bond lengths, the bond angles, and the torsion angles determined in the Z-matrix defining the molecule. If the molecule is made up of n atoms, the variables are defined as $n - 1$ = bond lengths, $n - 2$ = bond angles, and $n - 3$ = torsion angles. After geometry optimization of a molecule, the Hessian matrix is available as output, but only for quite small molecules (in Gaussian 92 it was restricted to twenty variables, corresponding to only eight atoms). To bypass this problem, a frequency calculation must be added after geometry optimization.

Examining the eigenvalues of the matrix can ensure that the geometry optimization did not locate a saddle point, but rather a minimum, which is required for any further step. Both saddle point and minimum are so-called stationary points, characterized by a zero gradient (first derivative with respect to the atom coordinates). But a minimum of a multidimensional potential energy surface is characterized by exclusively positive eigenvalues of the Hessian matrix, whereas a saddle point corresponds to at least one negative eigenvalue (a first-order saddle point is characterized by one negative eigenvalue, a second-order saddle point by two negative eigenvalues and so forth). The easiest way to distinguish the type of stationary point is, however, to consider the number of frequencies (NIMAG) at the end of the Gaussian output. "NIMAG=0" means that no vibrational frequencies exist and a minimum has been found.

The next two steps in the procedure of Leonard and Ashman are the conversion of the diagonal elements from atomic units into force field units and calculation of scaling factors for bond lengths and angles. The calculated force constants had to be scaled down by approximately 25% and 70% to yield force constants comparable in numerical size with those included in MM2. Neither force constants nor scaling factors can be incorporated directly into a different force field. A modification of the described procedure that meets the requirements of CVFF was developed. Fragments with known force field parameters were chosen. After a full geometry optimization (HF/6-31G*) second derivatives and vibrational frequencies were calculated. The force

METHOD A

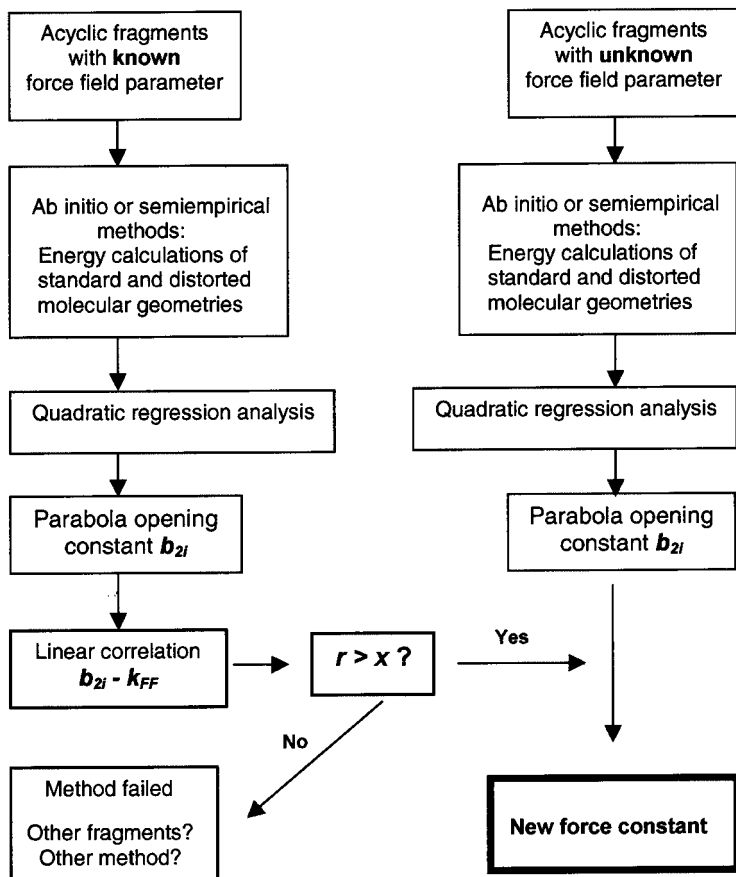


Fig. 9.1 (a) Schematic architecture of force field parameterization procedures for bond lengths and bond angles according to Hopfinger and Pearlstein [4]. In a first step model fragments with known parameters are varied slightly in their geometry and the resulting energy profile is calculated. The parabolic run of the energy curve enables calculation of

the parabola opening constant (b_{2i}) from quadratic regression analysis. If a linear correlation ($r > x$) of b_{2i} and the known force constant (k_{FF}) is achieved, the method is suitable for the estimation of missing force field parameters. Parameters for ring systems are derived from nonring model fragments.

METHOD B

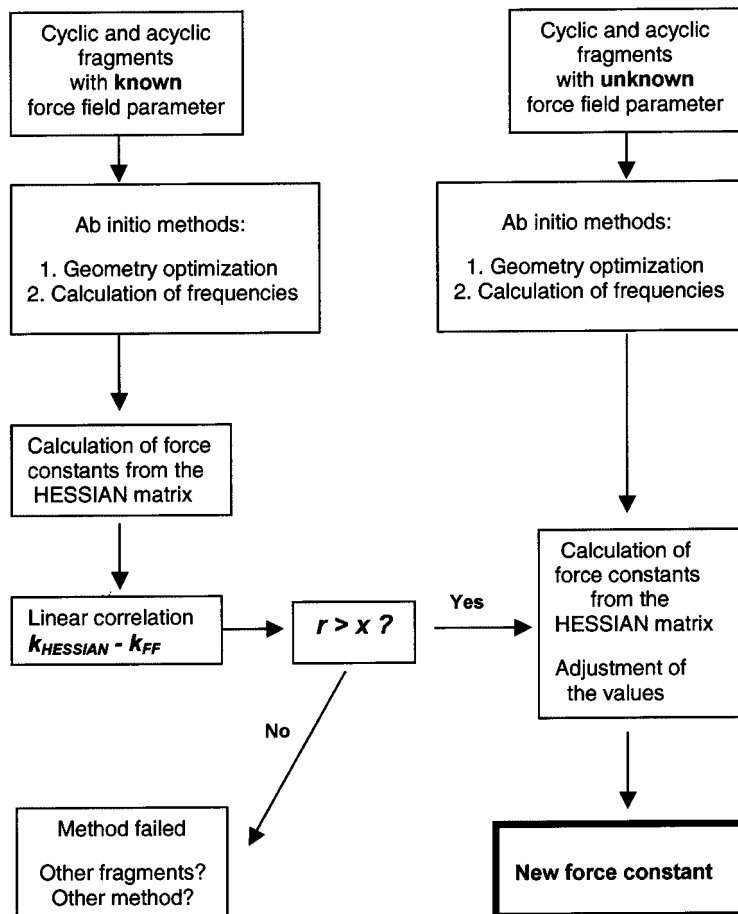


Fig. 9.1 (b) Schematic architecture of force field parameterization procedures for bond lengths and bond angles according to Leonard and Ashman [5]. Fragments used for parameterization are optimized. A frequency calculation is included. The Hessian matrix comprises the second partial derivatives of the wave functions with regard to dis-

placement of the atom coordinates. Diagonal elements are the force constants associated with each variable (bond lengths, angles, dihedrals). The extracted force constants have to be converted from atomic units to force field units and have to be adjusted to the force field values.

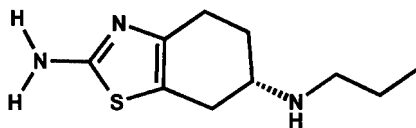
constants were extracted from the Gaussian output and converted from atomic into CVFF units. Now a linear regression of the calculated values versus the original CVFF force constants was performed. The resulting correlation functions confirmed a linear relationship between the Hessian diagonal elements and the CVFF force constants, and could be used to scale newly calculated force constants.

Beyond force constants, values for the unstrained, equilibrated bond lengths and angles are needed for a force field calculation. These values can be taken from crystal structures, but it must be considered that X-ray data can be deceptive. Many structural details result from interactions of the molecule with its neighbors (packing forces), which usually do not exist in the environment of interest. A second disadvantage is the latitude in the quality of crystal structures, even at low *R* factors. A third limitation is that, depending on the temperature, molecules are vibrating in the crystals. Because of these “rigid body motions”, X-ray bond lengths can be substantially too short. Another problem is that X-rays measure the center of the electron density, but not the nuclear position. Force field programs, however, are based on nuclear positions. The crystallographic measurement of electron densities tends to lengthen C–N and C–O bonds, because the lone pair protrudes, so that the center of electron density is further away from the carbon than is the nucleus. Bonds to hydrogen atoms are always too short in X-ray measurements, because the electron is pulled toward the atom to which it is bonded. Thus, to avoid the uncertainty of crystal data, one can use reference values from *ab initio* calculations. Nonetheless, calculated values for lengths of double, triple, and hydrogen bonds might be too short. Large basis sets with polarization functions and electron correlation are required for reliable descriptions.

9.2.1.1 Equilibrium Geometry for a Dopamine-D₃-Receptor Agonist

Molecular mechanics methods achieve good structural accuracy for classical molecules, whereas their reliability for species with particular combinations of atoms may be questionable, particularly for molecules containing heteroatoms, which affect geometry and conformation via the position of their lone-pairs. Force-field programs, for example, often fail to calculate the geometry

Fig. 9.2 Structural formula of the dopamine agonist pramipexol.



of particular nitrogen atoms. The development of a pharmacophore model in a study on dopamine agonists required prior optimization of the aminothiazole derivative pramipexol (Fig. 9.2) [6]. Force fields offer pure sp^3 or pure sp^2 hybridization for classification of the nitrogen in the NH_2 group. It is known, however, that the nitrogen atom in aniline (aminobenzene) has intermediate geometry between pyramidal and trigonal planar, because of a certain amount of sp^2 hybridization [7]. The NH_2 group of the aminothiazoles is connected to a heteroaromatic five-membered ring with sp^2 geometry, thus creating an environment similar to that of NH_2 in aniline.

Analyzing crystal data for aminothiazoles could not solve the problem, because the available structures have both planar and slightly tetrahedral geometry [8]. Planar geometry of the NH_2 group is supported by possible tautomeric exchange of hydrogen atoms between the NH - and the NH_2 group, whereas more pyramidal geometry is supported by the fact that the ring-nitrogen is more electronegative than the sulfur and thus causes the sulfur to push electrons into the ring, which might cause greater sp^3 hybridization on the NH_2 group, because its lone pair is not that much attracted by the ring. This assumption was confirmed by quantum chemical calculations. 2-Amino-1,3-thiazole was geometry optimized starting from two different points (pyramidal NH_2 and planar NH_2) using 3-31G* and 6-31G** Hartree-Fock methods. Both basis sets yielded a slightly pyramidal geometry as energetically more favorable. It should be kept in mind however, that solvation and the formation of hydrogen bonds can lead to a change in the tautomeric forms of heterocyclic compounds. In those circumstances a trigonal planar arrangement at the NH_2 group should be energetically more favorable, as already mentioned. Thus, despite the result of the QM optimization which corresponds to vacuum conditions, the presence of hydrogen bonding partners might introduce a planar geometry.

9.2.1.2 Searching for a Bioactive Conformation

This issue will be exemplified by our pharmacophore search for a series of 17 α -hydroxylase-17,20-lyase inhibitors [9]. 17 α -hydroxylase-17,20-lyase converts gestagens such as progesterone and pregnenolone to androgens by 17 α -hydroxylation, followed by the cleavage of the side-chain. Because of its key role in the biosynthesis of androgens, inhibition of this enzyme results in a total blockade of androgen production. Thus, the enzyme is an interesting target in the treatment of prostate cancer. The development of potent enzyme inhibitors requires detailed understanding of their binding mode. The ability of ligands to bind to a receptor site is reflected by the three-dimensional (3D) shape of the ligand in the receptor-bound (bioactive) conformation.

A pharmacophore hypothesis can be derived from a series of known ligands and their common 3D-features. The pharmacophore is constructed by superimposing energetically favorable conformations of all ligands according to their consensus in structural features. The template structure for the superposition should fulfill two criteria:

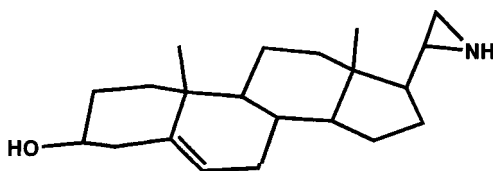
- the compound should be very potent, because then its molecular structure contains all essential features for high-affinity; and
- the structure should be rigid or semi-rigid to facilitate the search for the bioactive conformer.

The highly potent compound MH3, a semi-rigid 17 β -substituted aziridinyl-steroid was chosen as template (Fig. 9.3).

In this molecule only the bond between the steroid skeleton and the aziridine moiety can rotate freely. A molecular mechanics conformational analysis resulted in two low energy conformations (Fig. 9.4, see p. 264), both of which had almost identical potential energy, as calculated in the Tripos force field [10].

The conformers differ in the position of the nitrogen lone pair, pointing in two different directions. The lone pair is, however, essential for enzyme inhibition. The ligands interact directly with the heme iron of the enzyme, forming a coordinative bond via the nitrogen lone pair. Accordingly, orientation of the lone pair is critical for pharmacophoric superposition. Both conformers were studied in more detail by use of a quantum-mechanical approach. Applying the *ab initio* 3-21G** basis set, geometry optimization revealed

Fig. 9.3 Structure of the 17 α -hydroxylase-17,20-lyase inhibitor MH3.



a significant energy difference between the two conformers. Consequently, the resulting minimum-energy conformation was chosen as the template structure. Another structural feature to be investigated was the NH-group in the aziridine ring, for which pyramidal inversion is normally observed. This NH is, however, part of a conformationally very restricted three-membered ring with a high inversion barrier at biological temperatures. The nitrogen inversion process of strained NH aziridine rings has been the subject of many scientifically very interesting quantum chemical investigations [11].

Experimental investigation of aziridines revealed strong inversion even at low temperatures [12]. It subsequently became apparent that NH aziridines can achieve inversion by hydrogen exchange. In the presence of water the aziridine invertomers are rapidly interconverted at room temperature, by a protonation-deprotonation process. On the basis of *ab initio* calculations, it was suggested that the protonation-deprotonation is effected via a NH-bonded aziridine-2H₂O complex that promotes rapid site exchange of the geminal ring hydrogen atoms [13]. *Ab initio* optimization of the geometry of the MH3 invertomers with the 3-21G** basis set revealed no significant energy difference between the configurations. This problem could not, therefore, be solved by calculation and both invertomers of the minimum conformation had to be considered in the following steps of our study.

9.2.2

Atomic Point Charges

Accurate treatment of electron density distribution is required to determine the electrostatic properties of molecules. Thus, appropriate methods for obtaining atomic charges are of particular importance. Commercial modeling software packages offer the pos-

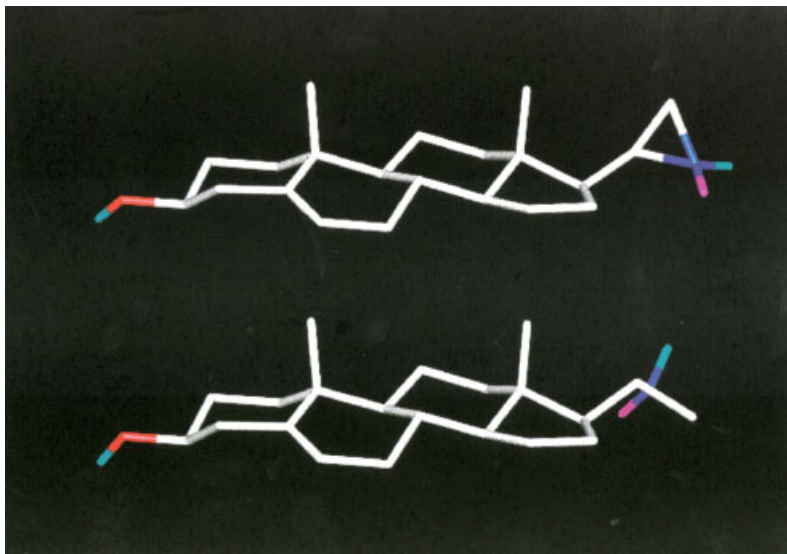
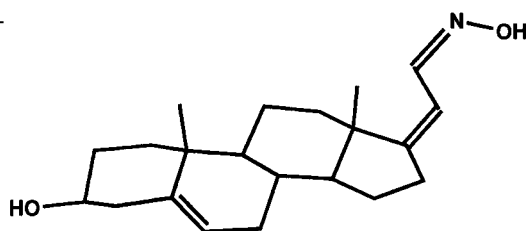


Fig. 9.4 Two energetically favorable conformations of the MH3 aziridine ring. The conformers differ in the position of the nitrogen lone pair (shown in magenta).

sibility of calculating atomic charges very easily by use of topological methods based mainly on the electronegativity of different types of atom. The algorithms used combine atomic electronegativities with experimentally derived properties of the bond types linking the atoms. The point charge of each atom is calculated, taking into account its own electronegativity and the electronegativities of the directly connected neighboring atoms and the types of their covalent bonds. Clearly, geometrical and conformational influences are neglected in these algorithms, their main advantage being that they are computationally fast. But, particularly for polar molecules, the use of topological methods to derive atomic charges can be unreliable, because of the inappropriate treatment of mesomeric or inductive effects. Where electronic effects predominate, molecular properties are best studied by use of quantum chemically-derived atomic charges.

Two general techniques are used to extract charges from wave functions – the Mulliken population analysis, based on partitioning the electron distribution, and the ESP method, based on fitting properties which depend on the electron distribution to a model which replaces this distribution by a set of atomic

Fig. 9.5 Structural formula of the 17 α -hydroxylase-17,20-lyase inhibitor MH61.



charges. Both methods can be performed with *ab initio* and semi-empirical program packages. Semi-empirical calculations are fast and user friendly and are, therefore, attractive methods for obtaining quantum mechanical results. In our experience, however, one should interpret carefully data from semi-empirical calculations involving nitrogen atoms in unsaturated systems. To compare the molecular electrostatic potentials of some 17 α -hydroxylase-17,20-lyase inhibitors (Section 9.2.2.2) we performed semi-empirical PM3 charge calculations. For ligand MH61 (Fig. 9.5) the Mulliken and the ESP method afforded rather different results for the oxime group.

The oxime contains an oxygen and a nitrogen in the same molecular segment, both of which can form a coordinative bond with the heme iron. In general, because of its greater electronegativity, oxygen is expected to be a better bonding partner than nitrogen. Nevertheless, the nitrogen in the oxime group should be able to compete as binding partner for the iron. The Mulliken method assigned a much stronger negative charge to the oxygen (-0.262) than to the nitrogen (-0.037), whereas ESP assigned a more negative value to the nitrogen atom ($O = -0.258$; $N = -0.390$). The different charge distributions will result in different pharmacophore models. Which method provides the “better” atomic charges?

Analysis of intermolecular interactions in the crystal structures of oxime molecules has been used to answer that question. In all available complex structures with one central metal ion we found no coordinative bonds from the oxime oxygen to the metal, but exclusively coordination between the nitrogen atom and the metal ion (data were retrieved from the Cambridge Crystallographic Database [14]). In a comprehensive study Böhm et al. investigated complexes of oxazoles, methoxypyridines, and oxime ethers with water [15]. On the basis of interaction energies obtained

from *ab initio* calculations they found the oxygen atoms to be rather weak hydrogen-bond acceptors when sp^2 -type nitrogen atoms were present in the same fragment. Thus, ESP-derived atomic point charges seem to be more appropriate for subsequent calculation of the molecular electrostatic potential. This example shows that charges from semi-empirical calculations should not be used without validation.

9.2.3

Molecular Electrostatic Potentials

The molecular electrostatic potential (MEP) can be used to investigate molecular interactions. If polar molecules approach each other the initial contact arises from long-range electrostatic attractions. These electrostatic interactions can be either attractive or repulsive. An electropositive part of an approaching molecule will seek to dock to an electronegative region; similarly charged parts will repel each other. Because of the charges and dipole moments in a molecule, an electrostatic field is generated in its environment. Consequently, distinct molecular electrostatic potentials exist at specific distances from the molecules. These can be represented in molecular modeling programs as interaction energy fields derived from the atomic charges and a positive point charge (a kind of proton model) which is located in a 3D grid surrounding the molecule.

We applied this method to find a biologically relevant conformation for H_2 -antagonists used for ulcer therapy. The compounds occupy the histamine binding site of the histamine H_2 -receptor, thus inhibiting histamine induced gastric acid secretion. Four main structural classes (Fig. 9.6) are used as drugs: imidazole derivatives (cimetidine), basically substituted furans (ranitidine), guanidinothiazoles (famotidine) and aminoalkylphenoxy derivatives (roxatidine). All these antagonists were supposed to occupy a common receptor binding site [16].

Because the 3D-structure of the receptor was not yet known, we wanted to derive a pharmacophore model, to prove this hypothesis. Most of the H_2 -antagonists have much conformational freedom. Evaluation of available crystal data and a systematic conformational search revealed mainly bent conformations,

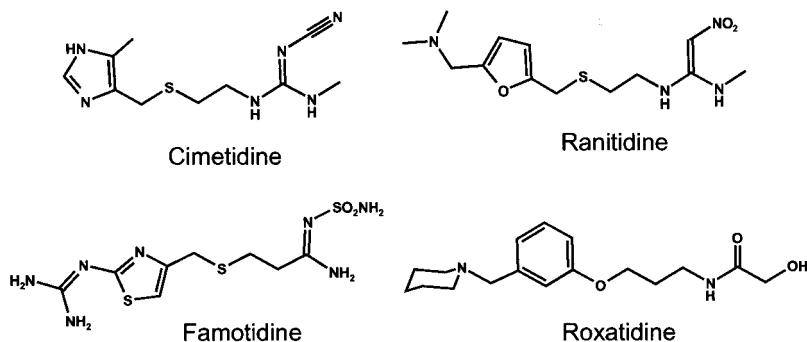


Fig. 9.6 Structural formulas of H₂-receptor antagonists used for ulcer therapy. The compounds can be divided into four different chemical classes: imidazole derivatives (cimetidine), furans (ranitidine), guanidinothiazoles (famotidine), and phenoxy derivatives (roxatidine).

partly with intramolecular hydrogen bonds, as most favorable. This result is in accordance with an investigation on the H₂-antagonist metiamide. The authors used the Hartree-Fock method with three different basis sets (3-21G*, 6-31G**, and 6-31+G**) to study the conformational properties of metiamide [17]. The calculations clearly indicate a preference for a folded conformation with a hydrogen bond between the imidazole ring and one of the NH groups, similar to the crystal structure (Fig. 9.7).

Calculations with one isolated molecule in vacuum often result in overestimation of intramolecular contacts, however, because competing interactions are absent. We started our investigation with the semi-rigid guanidinothiazole ICI27032, a potent competitive H₂-antagonist which adopts solely extended conformations, because of its aromatic ring system (Fig. 9.8).

The molecular electrostatic potential of the energetically most favorable ICI27032 conformation was calculated using AM1 ESP point charges. In the next step we searched for energetically favorable conformations of the other antagonists which yielded MEP similar to the MEP of the ICI27032 template. The results are displayed in Fig. 9.9. All the selected extended conformations have a highly similar spatial distribution of the MEP properties despite their different chemical fragments. It is therefore reasonable to suppose that all compounds could be recognized by the receptor in this extended conformation.

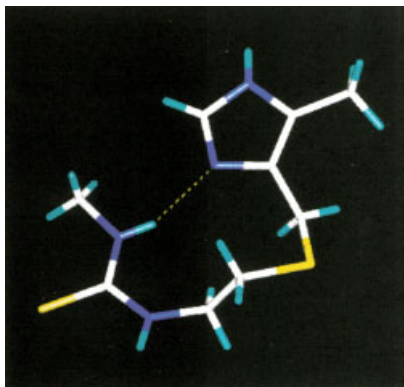


Fig. 9.7 Crystal structure of the H_2 -antagonist metiamide with an intramolecular hydrogen bond (yellow dots) between the imidazole nitrogen and an NH group of the side chain.

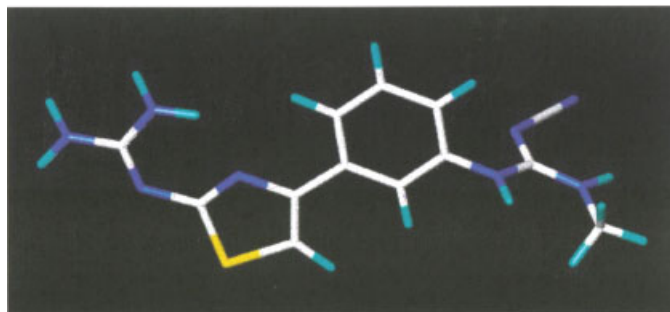
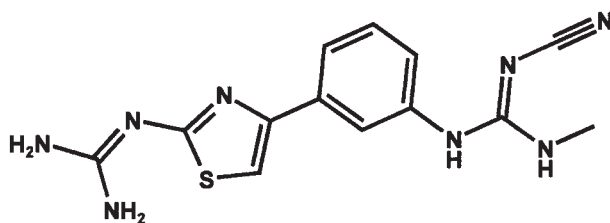


Fig. 9.8 Chemical and conformational structure of the H_2 -antagonist ICI27032.

9.2.4

Molecular Orbital Calculations

This section deals with the application of molecular orbital (MO) calculations in structure-activity relationship (SAR) analyses. Calcium channel-blocking 1,4-dihydropyridine (DHP) derivatives such as nifedipine (Fig. 9.10) are widely used in the therapy of cardiovascular disorders.

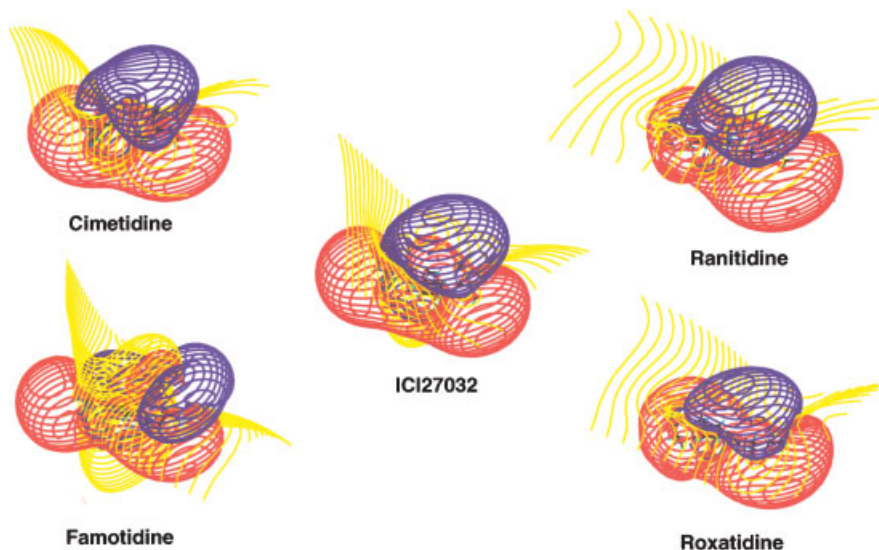


Fig. 9.9 Molecular electrostatic potentials of the H₂-antagonists. The MEP have been calculated on the basis of AM1 point charges and

are contoured at -1 kcal mol^{-1} (blue), $\pm 0 \text{ kcal mol}^{-1}$ (yellow), and 1 kcal mol^{-1} (red).

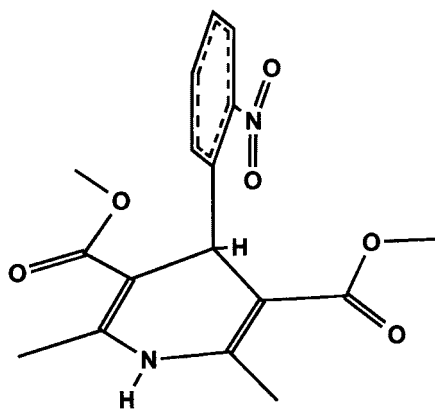
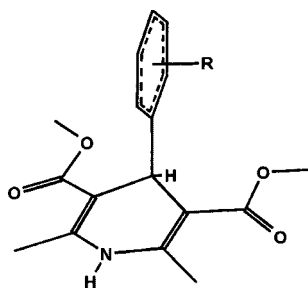


Fig. 9.10 Structure of the calcium channel-blocker nifedipine.

Although radiolabeling experiments and site-directed mutagenesis clearly reveal the α_1 -subunit of L-type calcium channels as binding site, no experimentally derived 3D coordinates of the receptor protein and its DHP binding site are available to elucidate the forces involved in specific binding. Hydrogen bond-donor properties of the NH group and at least one further hydrogen



Compound	R	ΔG
H	H	-10.7074
3Me	3'-CH ₃	-9.9299
3F	3'-F	-11.5804
3Cl	3'-Cl	-12.6852
3NO₂	3'-NO ₂	-13.5991
3CN	3'-CN	-11.8395
3N₃	3'-N ₃	-11.8259
3OMe	3'-OCH ₃	-9.9163
F₅	2',3',4',5',6'-F ₅	-14.1447

Fig. 9.11 Experimentally derived free binding energies (ΔG) of the DHP derivatives investigated [18].

bond accepted by the carbonyl groups of the ester side-chains, combined with electrostatic attractions, are indicated as the main binding forces by SAR analyses. It has, however, been demonstrated that these binding elements alone cannot account for the high affinity of some compounds. The affinity of 23 nifedipine-like DHP has been determined in radioligand binding experiments. The pK_i values range over more than five log units although the single structural change was the varied substitution pattern of the 4-phenyl ring.

It has not yet been clarified whether the ring substituents interact directly with the binding site or affect the molecular characteristics of the DHP molecules in common. A recently used atomistic pseudoreceptor model for a series of DHP indicated a putative charge-transfer interaction was stabilizing the DHP-binding site complex [19]. To prove this hypothesis qualitative and quantitative analysis of the molecular orbitals of nine DHP derivatives (Fig. 9.11) was performed [18]. Charge-transfer (or electron-donor-acceptor) interactions are indicative of electronic

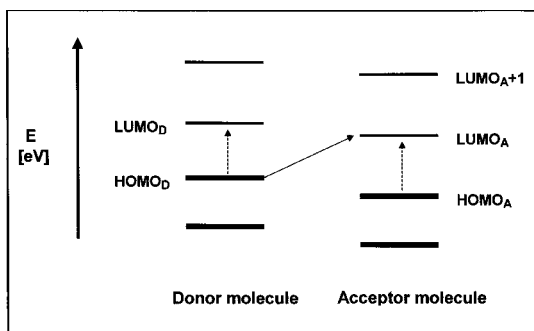


Fig. 9.12 Schematic representation of a charge-transfer interaction. The solid arrow illustrates π -electron transfer between the HOMO of the donor molecule (HOMO_D) and the LUMO of the acceptor molecule (LUMO_A). Dashed arrows indicate interactions between corresponding HOMO and LUMO of one molecule [18].

charge transfer from the HOMO of a donor molecule (HOMO_D) to the LUMO of an accepting neighboring molecule (LUMO_A). Small energy barriers between HOMO_D and LUMO_A increase the probability of charge transfer but with two further additions – the corresponding molecular orbitals must be able to overlap, and HOMO_D and LUMO_A must be energetically close (Fig. 9.12).

In a charge-transfer interaction for stabilization of the DHP-binding site complex the electron-accepting LUMO should be located on the 4-phenyl ring of the DHP, because highest binding affinities are found for derivatives with electron-withdrawing substituents at this position. Using the semi-empirical AM1 method, the molecular structures of the DHP were optimized and the molecular orbitals were computed (the author had demonstrated the reliability of AM1 by comparing results from high-level *ab initio* calculations). To assign correctly the unoccupied molecular orbitals located on the 4-phenyl ring (designated LUMO^*) all relevant unoccupied orbitals (LUMO , $\text{LUMO}+1$, $\text{LUMO}+2$) have been visualized. In nearly all instances the LUMO was positioned at the 1,4-dihydropyridine heterocycle. Only for compound 3NO_2 (3-NO_2) and F_5 ($1,2,3,4,5\text{-F}_5$) were the LUMO identical with the LUMO^* (Fig. 9.13).

All the other DHP, which lack comparably potent electron-withdrawing substituents, contain only energetically less favor-

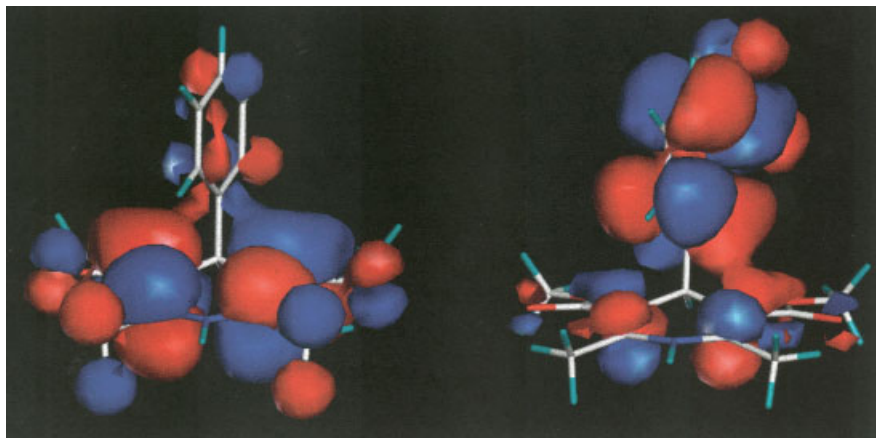


Fig. 9.13 Molecular orbital plots. The LUMO of compound **H** (left) is located on the 1,4-dihydropyridine heterocycle whereas for compound **3NO₂** it is located on the 4-phenyl ring (right).

able LUMO* (LUMO+1 and LUMO+2) at the aromatic ring system. In contrast with the situation described so far, compounds **3N₃** (3-N₃) and **3OMe** (3-OCH₃) have the HOMO located instead on the 4-phenyl ring. Electron-donating moieties apparently lead to an electron excess at the 4-phenyl ring, thereby inducing the formation of occupied MO at this position and so the existence of energetically favorable LUMO* at the 4-phenyl ring becomes more unlikely. To decide whether charge-transfer interactions might play a major role in the receptor binding of DHP, the experimentally derived free binding energies (ΔG) were correlated with the calculated LUMO* energies; a highly significant correlation coefficient of 0.91 was obtained (Fig. 9.14).

When examining these results one must bear in mind that ΔG values not only mirror charge-transfer interactions but a multitude of ligand-dependent direct (hydrogen bonding, electrostatic, van der Waals, and hydrophobic binding forces) and indirect effects (solvation energies and entropic terms). Assuming that all investigated DHP besides charge transfer exert almost identical direct interactions, one might argue that indirect factors narrow the interpretation of the data. Because of the limited structural variation of the compounds however (only a single functional group is modified), approximately the same entropic effects can be postulated. Thus, the obtained correlation should reflect the

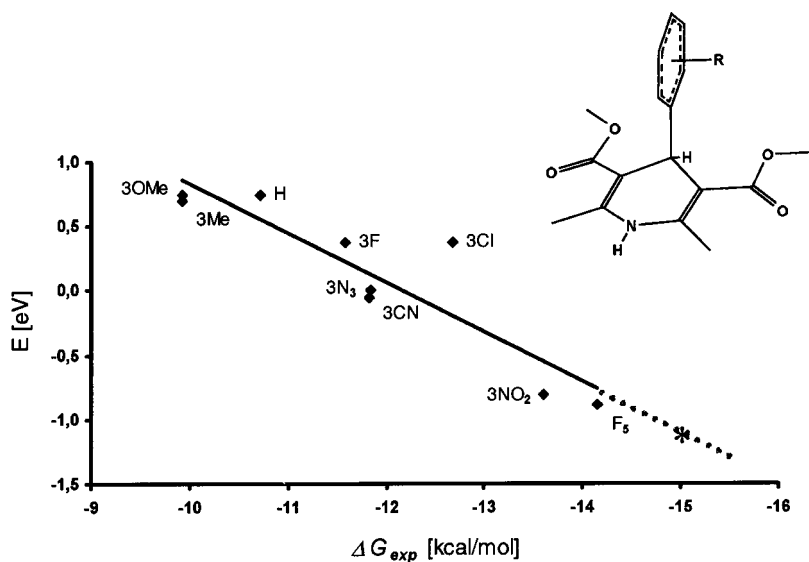


Fig. 9.14 Correlation of the calculated LUMO* energies and the experimentally derived free binding energies (ΔG) of the investigated DHP compounds. The asterisk on the extrapolated dotted line depicts the calculated LUMO* energy of the most active compound, isradipine [19].

effect of potential 4-phenyl ring charge-transfer interactions on the binding affinities of the DHP. To determine whether a reliable model had been found the binding energy of a novel DHP was predicted. Calculations were performed to determine the LUMO* of the most active DHP isradipine, a compound which has a benzoxadiazole ring instead of the 4-phenyl ring. The LUMO* energy was indicative of a very strong charge-transfer interaction and ranked the compound in accordance with the experimental data as most potent (Fig. 9.14).

9.3 Outlook

It has been shown with a few examples that the application of quantum mechanics is an essential tool in molecular modeling in medicinal chemistry. It is not, however, yet playing a role in disciplines in which the size of the molecular systems is too

large to enable study by use of *ab initio* methods. The development of massively parallel supercomputers should afford the potential to predict molecular properties relevant to a variety of problems in medicinal chemistry – e.g. penetration, diffusion, or metabolic processes. Computational calculations have a good chance of representing “reality” if they are based on quantum theory. This makes molecular modeling more reliable and offers the opportunity to complement experimental measurements.

9.4

References

- HÖLTJE, H.-D.; FOLKERS, G., *Molecular Modeling-Basic Principles and Applications in: Methods and Principles in Medicinal Chemistry*, R. MAUNHOLD, H. KUBINYI, H. TIMMERMANN (eds), Vol. 5, VCH, 1996;
- LEACH, A. R., *Molecular Modeling-Principles and Applications*, Longman, UK, 1996.
- GAEDT, K.; HÖLTJE, H.-D., *J. Comput. Chem.* **1998**, *19*, 935–946.
- Insight/Discover, Accelrys Inc. (<http://www.accelrys.com>).
- HOPFINGER, A. J.; PEARLSTEIN, R. A., *J. Comput. Chem.* **1984**, *5*, 486–499.
- LEONARD, J. M.; ASHMAN, W. P., *J. Comput. Chem.* **1990**, *11*, 952–957.
- GAEDT, K., Ph.D. Thesis, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany, 1998.
- GILLESPIE, R. G.; MORTON, M. J., *Inorg. Chem.* **1970**, *9*, 616–618.
- LUGER, P.; GRISS, G.; HURNAUS, R.; TRUMMLITZ, G., *Acta Crystallogr.* **1986**, *B42*, 478–490;
- CARANONI, P. C.; REBOUL, J. P., *Acta Crystallogr.* **1982**, *B38*, 1255–1259.
- SCHAPPACH, A.; HÖLTJE, H.-D., *Pharmazie* **2001**, *56*, 835–842.
- SYBYL 6.4, Tripos Associates (<http://www.tripos.com>).
- JENNINGS, W. B.; BOYD, D. R., *Strained Rings in: Cyclic Organonitrogen Stereodynamics*, LAMBERT, J. B.; TAKEUCHI, Y. (eds), VCH, 1992.
- SKAARUP, S., *Acta Chem. Scand.* **1972**, *10*, 4190–4192.
- CATALAN, J.; MARCIAS, A.; MO, O.; YANEZ, M., *Mol. Phys.* **1977**, *43*, 1429–1433.
- Cambridge Structural Database (<http://www.ccdc.cam.ac.uk>).
- BÖHM, H.-J.; BRODE, S.; HESSE, U.; KLEBE, G., *Chem. Eur. J.* **1996**, *2*, 1509–1513.
- HÖLTJE, H.-D.; BATZENSCHLAGER, A., *J. Comput.-Aided Mol. Design* **1990**, *4*, 391–402.
- MARTINS, J. B. L.; TAFT, C. A.; PEREZ, M. A.; STAMATO, F. M. L. G.; LONGO, E., *Int. J. Quantum Chem.* **1998**, *69*, 117–128.
- SCHLEIFER, K. J., *Pharmazie* **1999**, *54*, 804–807.
- SCHLEIFER, K.-J., *J. Med. Chem.* **1999**, *42*, 2204–2211.

Subject Index

a

ab initio 117
ab initio method 122
 – study of reaction mechanisms 157
ab initio molecular dynamics
 (AIMD) 42, 45, 93 ff.
ab initio pseudopotential (AIPP) 13,
 79
ab initio QM 157
ab initio QM/MM 184
 absorption coefficient 149
 absorption energy 35
 acetone 34 f.
 acid-base reactivity 225
 activation energy 169
 activation enthalpy 169
 active center
 – myoglobin 81
 adiabatic mapping 187
 adiabatic mapping method 185
 AFP 220
 AIP 145
 AIPP *see ab initio* pseudopotential
 AM1 123, 180
 AM1 ESP
 – point charges 267
 AMBER 18 ff.
 aniline 235, 241
 antiviral therapy 49
 applications
 – QM/MM 189, 170
 – TD-DFT 34
 atom 203
 atomic basin 203
 atomic charge 209
 atomic point charge 20, 264
 atomic properties
 – definition 208

atomic volumes 210
 atoms in molecules 201
 – theory 202
 augmented plane wave 14

b

B3LYP 119, 122
 B88 119
 back-bonding 103
 π -back-bonding 91
 basic equation
 – density-functional perturbation
 theory 21 ff.
 – density-functional theory 115
 – TD-DFT 32
 basic ideas
 – Car-Parrinello 6
 basic theory
 – QM/MM 179
 basis set superposition error 14
 BCP 204 ff., 213
 BF₃ 204
 binding affinity 99
 binding energy 84
 bioactive conformation 262
 biomimetic 77
 biomimetic complex 86
 Biot-Savart law 27
 blue shift 35
 bond
 – critical point (BCP) 204 ff., 213
 – ellipticity, ε 213
 – indices 213
 – order 207
 – path 204 f.
 bonding
 – hydrogen 214

- ionic 214
- van der Waals 214
- Born-Oppenheimer surface 11

c

- calcium channel 269
- CAMD *see* computer-aided molecular design
- carbamoyl sarcosine 27
- carbonic anhydrase 170
- carcinogenicity 239 ff.
- Car-Parrinello 5, 132
 - basic ideas 6
 - method 46
 - molecular dynamics (CP-MD) 5, 73, 93 ff.
- CASTEP 15
- CCSD(T) 122
- Cdc42 58
- charge distribution
 - transferability 208
- charge separation 246 ff.
- charge separation index, CSI 211
- charge transfer 91, 210, 238, 270
- CHARMM 100, 166, 186
- chemical descriptor 214
- chemical shielding 26 f.
- chorismate mutase 172, 180
- cimetidine 267, 269
- cisplatin 124, 126
 - activation reactions 127
 - *cis*-diaminedichloroplatinum(II) (*cis*-DDP) 124
 - interactions between DNA 134
 - modes of action 124
- Cl₂ 228
- classical molecular dynamics (classical MD) 7
- closed-shell interactions 214
- compensatory transferability 218
- computer-aided molecular design (CAMD) 41
- configurational space 94
- connection atom 182
- consistent valence force field (CVFF) 256
- Coulomb's law 233
- CP *see* critical point
- CPMD 15, 18
- critical point (CP) 205
- crown thioether 66
- CSGT 28
- CSI 211

- curvature
 - electron density $\rho(r)$ 204

d

- $\nabla^2\rho(r)$ 224
- Davidson algorithm 34
- cis*-DDP (*cis*-diaminedichloroplatinum(II)) 124
 - cisplatin 124
- definition
 - atomic properties 208
 - of the QM region
 - – QM/MM 183
- π -delocalization 255
- delocalization index
 - electron delocalization 226
- density operator 29
- density-functional perturbation theory 5, 21 ff., 25, 32
 - basic equation 22
- density-functional theory (DFT) 5 ff., 41 ff., 113 ff.
 - basic equation 115
- DFPT *see* density-functional perturbation theory
- DFTPT *see* density-functional perturbation theory
- DHP 270
- diamagnetic state 89
- cis*-diaminedichloroplatinum(II) (*cis*-DDP) 124
 - cisplatin 124
- dibenzo-*p*-dioxin 244 ff.
- diffusion constant 248
- dipolar coupling 26
- dipole 209
- dipole moment 50, 54, 120, 210, 246
- dispersion force 228
- DNA 126
- dopamine D₃-receptor agonist 256, 260
- drug design 56, 255
- dynamics 93

e

- effective fragment potential (EFP)
 - method 171
- effective Hamiltonian, H_{eff} 179
- effective potential, v_{eff} 116
- eigen-energy 117
- eigenvalue 25, 257
- electric polarizability 208, 212
- π electron 237

electron delocalization 226
 – delocalization index 226
 – quantitative measure 226
 electron density
 – topology 203
 electron density $\rho(\mathbf{r})$
 – curvature 204
 electron localization function (ELF) 62
 electron pair density 224
 electron-donor-acceptor 270
 electron-electron interaction 116
 electronic configuration 89
 electronic current density $j(\mathbf{r})$ 29
 electronic excitation spectrum 120
 electronic pair density 207
 electronic wavefunction 23, 179
 electron-withdrawing 246
 electrophile 238f.
 electrorestriction 211
 electrostatic effect 160
 electrostatic energy 50
 electrostatic field 266
 electrostatic potential 209
 electrostatic potential $V(\mathbf{r})$ 233
 embedding method 177
 empirical valence bond (EVB)
 method 159
 energy barrier 184
 energy profile 185
 enkephalin 215
 entropic effect 6
 enzymatic catalysis 57
 enzyme inhibitor 262
 epoxide 239ff.
 equations of motion (EOM) 9
 equilibrium structure 93
 ESP 264
 Euler-Lagrange equation 10, 16
 EVB method 170
 $E_{\text{xc}}[\rho]$ *see* exchange-correlation energy
 functional
 exchange-correlation energy functional
 E_{xc} 16, 23, 44, 116f.
 exchange-correlation potential 25, 45, 121
 excitation energy 34, 121
 excitation spectrum 146
 excited electronic state 21, 32
 extended system method 12
 external perturbation 33

f

famotidine 267, 269
 FeCO 92, 94ff.

FeO₂ 89
 FEP *see* free energy perturbation
 Fermi hole 225
 Fhi98md 15
 fictitious non-interacting electron 44
 finite-difference method 23
 first-principles molecular dynamics
 (*ab initio* MD) 7
 force constant 257
 force field 179
 force field parameters 256
 Fourier spectroscopy 26
 free binding energy (ΔG) 270, 273
 free energy 212
 free energy calculations
 – QM/MM 186
 free energy perturbation (FEP) 167
 frequency
 – vibrational 22
 frontier orbital 181
 frozen orbital 165
 – approach 181
 furocoumarin 144

g

G2 122
 gauge origin 28
 Gaussian 92, 257
 general interaction properties function
 (GIPF) 246, 248ff.
 generalized gradient approximation
 (GGA) 118
 generalized hybrid orbital (GHO)
 method 165
 geometry optimization 186
 GIAO 28
 global attractor 203
 glutathione S-transferase 191
 GRACE 186
 gradient corrections 117
 gradient expansion approximation
 (GEA) 118
 gradient vector field 208
 gradient-corrected density func-
 tional 118
 GROMOS 18ff.
 group additivity 208
 GTP hydrolysis 58
 guanine-cytosine basepair 207

h

H₂-antagonist 266
 halogenated olefin 239ff.

- Hamiltonian 28
 Hartree-exchange-correlation kernel 25
 HCTH 119
 heat of vaporization 248
 heat of formation 208
 Hellman-Feynman theorem 8
 heme 92
 heme-CO 102
 hemoglobin (Hb) 74
 Hessian 256
 Hessian matrix 185, 257
 Hirshfeld charge 20
 HIV 249
 – integrase 63
 – protease 194
 HK theorem *see* Hohenberg-Kohn theorem
 Hohenberg-Kohn theorem 43, 115
 HOMO 47, 271
 homogeneous electron gas 115, 117
 HOMO-LUMO 52
 human aldose reductase 189
 hybrid functional 117
 hybrid orbital approach
 – QM/MM 165
 hybrid orbital method 181
 hybridization 261
 hydride transfer 169
 hydrogen-bond
 – acidity 246
 – basicity 246
 hydrophobic character 212
 hyperpolarizability 120
- i**
- ICI27032 269
 IGLO 28
 imaginary frequency 185
 interatomic surface 206
 intrinsic physical volume 211
 intrinsic reaction coordinate (IRC) 185
 ionization energy 146
 ionization potential 143
- j**
- Jaguar program 166
 JEEP 15
- k**
- kinetic energy 9
 kinetic energy density 214
 kinetic isotope effect (KIE) 169
 Kohn-Sham energy functional 22
 Kohn-Sham equation 22, 32, 45, 116
 Kohn-Sham functional 23
 Kohn-Sham Hamiltonian (H^{KS}) 25, 46
 Kohn-Sham (KS) orbital 32, 116f.
 – one-electron 12
- l**
- Lagrangian 9, 16, 46
 Lagrangian mechanics 9
 Lagrangian multiplier 10
 Langevin force 188
 Lan12DZ 128
 Laplacian of the density, $\nabla^2\rho_b$ 213, 224
 LBHB 62
 LDA 118
 Lewis model 224
 ligand binding 99
 linear response orbital 24
 linear-response wavefunction 33
 linear-scaling algorithm 157
 link atom 164, 181
 – approach 164
 – – QM/MM 162
 local density approximation (LDA) 44, 117
 local electron-correlation method 157
 local self-consistent field (LSCF) method 165
 local spin density approximation
 LSDA 117
 localized bond orbital 181
 localized bonded 224
 localized orbitals 165, 181
 low-barrier hydrogen bond
 (LBHB) 62, 171
 LUMO 47, 271
 LYP 119
- m**
- magnetic field 25
 magnetic field B^{ext} 27
 magnetic perturbation 28
 magnetic susceptibility 208
 many-body electronic wavefunction 43
 MEP *see* molecular electrostatic potential
 meta-GGAs 120
 metallic interactions 214
 metiamide 268
 MM force field 163
 MM2 257
 MM2 force field 256
 MNDO 180

modeling enzyme reactions
 – QM/MM 182
 molecular dynamics (MD) 73, 167
 molecular dynamics simulations 93
 molecular electrostatic potential
 (MEP) 233 ff., 266
 molecular graph 205
 molecular mechanics (MM) 177
 molecular orbital calculation 179
 molecular quantum similarity measure
 (MQSM) 215
 molecular recognition 224
 molecular similarity 214 f.
 molecular mechanical methods 177
 momentum operator 28
 Monte Carlo methods 187
 MOPAC 194
 morphine 215 f.
 Mulliken charges 209
 Mulliken population analysis 264
 multipole moment 209
 myoglobin (Mb) 73 f.
 – active center 81

n

neuraminidase 193
 Newton's equations of motion 9, 47
 nifedipine 268 f.
 NMR chemical shieldings 25 ff.
 NMR chemical shifts 26
 non-bonded electron pair 224
 norm-conservation 13
 nuclear magnetic resonance (NMR)
 spectroscopy 26
 nucleophile 239
 nucleus 203
 NWCHEM 15

o

O₂ binding 90
 occupied orbital 87
 ONIOM method 180
 ONIOM-type methods 162
 open-shell singlet 87
 open-shell singlet state 88
 opioid receptor 215
 OPLS-AA force field 166
 opsochromic shift 34
 optical excitation 34
 orbital
 – linear response 24
 oxyheme 97
 oxy-myoglobin (MbO₂) 105

p

II 246
 partial molal volume \bar{v}^0 211
 particle-particle-particle mesh (P3M) 18
 partition coefficient 248
 partitioning schemes
 – QM/MM 180
 path
 – bond 204
 path integral algorithm 170
 path integral simulation 193
 pathogen eradication technology
 (PET) 143
 Pauli exclusion principle 225
 PAW 15
 PBE 119
 PBE0 119, 122
 PEO 215 f.
 periodic boundary conditions 13
 perturbation density $\rho^{(1)}$ 24
 perturbation theory 24, 28, 33, 238
 PET *see* pathogen eradication technology
 PET *see* position emission tomography
 pharmacophore 261 f.
 phosphoryl transfer 58 ff.
 photochemistry 41
 photochemotherapy 141
 physicochemical property 210
 picket-fence 86, 91 f.
 picket-oxygen 73
 pK_a 158
 pK_i values 270
 PKZB 120
 plane wave (PW) 12, 79
 PM3 123, 180, 194, 265
 point charges
 – AM1 ESP 267
 Poisson-Boltzmann equation 166
 polarizability 213
 polarizable force field 19
 polarization 238
 polarization effect 19
 polypeptide 211
 porphyrin 93
 positron emission tomography (PET) 66
 potential energy 9
 potential energy surface (PES) 7 f.
 potential of mean force 188
 pramipexol 261
 product complex 140
 protoheme complex 92
 proton affinity 158
 proton NMR chemical shieldings 31

proton tunneling 170
 protonation state 183
 pseudobond 182
 pseudopotential 13
 pseudoreceptor model 270
 psoralen 141
 PW86 119
 PW91 119, 122

q

QM/FE 167
 QM/MM 31, 34, 99 ff., 104, 158 ff.,
 161 ff.
 – applications 170, 189
 – basic theory 179
 – definition of the QM region 183
 – free energy calculations 186
 – hybrid orbital approach 165
 – link atom approach 162
 – method 177
 – modeling enzyme reactions 182
 – partitioning schemes 180
 – quantum effects 169
 – reaction pathways 185
 – selected applications 168
 – thermodynamically coupled QM/
 MM 166
 – transition structures 185
 QSAR 211
 quadrupolar polarization $Q(A)$ 216
 quadrupole 209
 quantitative measure
 – electron delocalization 226
 quantitative structure-activity relationship
 (QSAR) 211
 quantum effects 171
 – QM/MM 169
 quantum mechanical/free energy
 (QM/FE) method 160
 quantum-mechanical methods 177
 quantum mechanical/molecular mechani-
 cal (QM/MM) 5, 15 ff.
 quantum theory of atoms in molecules
 (QTAIM) 201
 quantum topological molecular similarity
 (QTMS) 214

r

radiopharmaceuticals 65
 ranitidine 267, 269
 reactant 228
 reactant complex 13, 138
 reaction barrier 169

reaction energy 139
 reaction energy surface 133
 reaction mechanism 129 f.
 reaction path 185
 – modeling 186
 – QM/MM 185
 reactive surface 228
 reactivity 234
 receptor 262
 response orbital 29
 reverse transcriptase (RT) 249
 ROKS 35
 roxatidine 267, 269

s

saddle point 186
 SAR 48, 56
 Schrödinger equation 42
 semiempirical method 187
 shared interactions 214
 shieldings 31
 side chain 212
 SIESTA 15
 singlet excitation energy 147, 149
 singlet state 88
 solvation energy 248
 solvation shell 35
 spin state 88
 state
 – excited electronic 21
 stationary point 131, 257
 stationary wavefunction 32
 steroid 262
 stochastic boundary approach 191
 stretch frequency 103
 strong hydrogen bond 184
 structure-activity relationship (SAR) 268
 structure-property relationships 212
 study of reaction mechanisms
 – *ab initio* methods 157
 surface tensions 248
 S-VWN 119
 system method
 – extended 12

t

tautomer 104
 TCDD 244
 TD-DFT
 – applications 34
 – basic equations 32
 theorem
 – Hellman-Feynman 8

- theory
 - atoms in molecules 202
 - therapy
 - antiviral 49
 - thermal motion 95
 - thermodynamically coupled QM/MM 166
 - thrombin 193
 - thymidine kinase 48
 - time-dependent density-functional response theory (TD-DFRT) 120
 - time-dependent density-functional theory (TDDFT) 5
 - introduction 32
 - time-dependent Schrödinger equation 32
 - TIP3P water 190
 - topology
 - electron density 203
 - toxicity 244f.
 - trajectory 35, 93, 203, 205
 - transferability
 - charge distribution 208
 - transition metal complexes 64
 - transition state (TS) 128, 138, 185
 - analog 57, 62
 - theory (TST) 170
 - transition structure 132
 - QM/MM 185
 - triplet state 88, 142
- u**
- ultrasoft Vanderbilt pseudopotential 13
 - umbrella sampling 188, 191, 194
 - uracil-DNA glycosylase 168
- v**
- van der Waals radii 210
 - van der Waals surface 210
 - variational principle 24
 - VASP 15
 - vector potential $A(\mathbf{r})$ 28
 - velocity Verlet 11
 - Verlet 11
 - vibrational frequency 22, 104
 - vibrational mode 22, 96
 - VIP 145
 - viscosities 248
 - VSEPR 226
 - VWN 118
- w**
- Wannier function 55
 - weighted histogram analysis 188
 - WFC 55
- x**
- xenobiotics 191
- z**
- zero-flux 207
 - surface 215
 - Z-matrix 257