

268

Topics in Current Chemistry

Editorial Board:

**V. Balzani · A. de Meijere · K. N. Houk · H. Kessler · J.-M. Lehn
S. V. Ley · S. L. Schreiber · J. Thiem · B. M. Trost · F. Vögtle
H. Yamamoto**

Topics in Current Chemistry

Recently Published and Forthcoming Volumes

Bioactive Confirmation I

Volume Editor: Peters, T.
Vol. 272, 2007

Biom mineralization II

Mineralization Using Synthetic Polymers and
Templates
Volume Editor: Naka, K.
Vol. 271, 2007

Biom mineralization I

Crystallization and Self-Organization Process
Volume Editor: Naka, K.
Vol. 270, 2007

Novel Optical Resolution Technologies

Volume Editors:
Sakai, K., Hirayama, N., Tamura, R.
Vol. 269, 2007

Atomistic Approaches in Modern Biology

From Quantum Chemistry
to Molecular Simulations
Volume Editor: Reiher, M.
Vol. 268, 2006

Glycopeptides and Glycoproteins

Synthesis, Structure, and Application
Volume Editor: Wittmann, V.
Vol. 267, 2006

Microwave Methods in Organic Synthesis

Volume Editors: Larhed, M., Olofsson, K.
Vol. 266, 2006

Supramolecular Chirality

Volume Editors: Crego-Calama, M.,
Reinhoudt, D. N.
Vol. 265, 2006

Radicals in Synthesis II

Complex Molecules
Volume Editor: Gansäuer, A.
Vol. 264, 2006

Radicals in Synthesis I

Methods and Mechanisms
Volume Editor: Gansäuer, A.
Vol. 263, 2006

Molecular Machines

Volume Editor: Kelly, T. R.
Vol. 262, 2006

Immobilisation of DNA on Chips II

Volume Editor: Wittmann, C.
Vol. 261, 2005

Immobilisation of DNA on Chips I

Volume Editor: Wittmann, C.
Vol. 260, 2005

Prebiotic Chemistry

From Simple Amphiphiles to Protocell Models
Volume Editor: Walde, P.
Vol. 259, 2005

Supramolecular Dye Chemistry

Volume Editor: Würthner, F.
Vol. 258, 2005

Molecular Wires

From Design to Properties
Volume Editor: De Cola, L.
Vol. 257, 2005

Low Molecular Mass Gelators

Design, Self-Assembly, Function
Volume Editor: Fages, F.
Vol. 256, 2005

Anion Sensing

Volume Editor: Stibor, I.
Vol. 255, 2005

Organic Solid State Reactions

Volume Editor: Toda, F.
Vol. 254, 2005

Atomistic Approaches in Modern Biology

From Quantum Chemistry to Molecular Simulations

Volume Editor: Markus Reiher

With contributions by

L. Bertini · P. G. Bolhuis · M. Bruschi · C. Dellago
M. Dittrich · P. Fantucci · L. de Gioia · C. Greco · C. Herrmann
B. Kirchner · F. Neese · W. Reckien · M. Reiher · K. Schulten
H. M. Senn · S. Sinnecker · J. Thar · W. Thiel · J. Yu
G. Zampella

The series *Topics in Current Chemistry* presents critical reviews of the present and future trends in modern chemical research. The scope of coverage includes all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science. The goal of each thematic volume is to give the nonspecialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to a larger scientific audience.

As a rule, contributions are specially commissioned. The editors and publishers will, however, always be pleased to receive suggestions and supplementary information. Papers are accepted for *Topics in Current Chemistry* in English.

In references *Topics in Current Chemistry* is abbreviated Top Curr Chem and is cited as a journal.

Visit the TCC content at springerlink.com

ISSN 0340-1022

ISBN-10 3-540-38082-5 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-38082-5 Springer Berlin Heidelberg New York

DOI 10.1007/978-3-540-38085-6

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMXDesign GmbH, Heidelberg

Typesetting and Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig

Printed on acid-free paper 02/3100 YL – 5 4 3 2 1 0

Volume Editor

Prof. Dr. Markus Reiher

Laboratorium für Physikalische Chemie
ETH Zürich
Hoenggerberg Campus, Wolfgang-Pauli-Str.10
8093 Zürich, Switzerland
markus.reiher@phys.chem.ethz.ch

Editorial Board

Prof. Vincenzo Balzani

Dipartimento di Chimica „G. Ciamician“
University of Bologna
via Selmi 2
40126 Bologna, Italy
vincenzo.balzani@unibo.it

Prof. Dr. Armin de Meijere

Institut für Organische Chemie
der Georg-August-Universität
Tammanstr. 2
37077 Göttingen, Germany
ameijer1@uni-goettingen.de

Prof. Dr. Kendall N. Houk

University of California
Department of Chemistry and
Biochemistry
405 Hilgard Avenue
Los Angeles, CA 90024-1589
USA
houk@chem.ucla.edu

Prof. Dr. Horst Kessler

Institut für Organische Chemie
TU München
Lichtenbergstraße 4
86747 Garching, Germany
kessler@ch.tum.de

Prof. Jean-Marie Lehn

ISIS
8, allée Gaspard Monge
BP 70028
67083 Strasbourg Cedex, France
lehn@isis.u-strasbg.fr

Prof. Steven V. Ley

University Chemical Laboratory
Lensfield Road
Cambridge CB2 1EW
Great Britain
Svl1000@cus.cam.ac.uk

Prof. Stuart L. Schreiber

Chemical Laboratories
Harvard University
12 Oxford Street
Cambridge, MA 02138-2902
USA
sls@slsiris.harvard.edu

Prof. Dr. Joachim Thiem

Institut für Organische Chemie
Universität Hamburg
Martin-Luther-King-Platz 6
20146 Hamburg, Germany
thiem@chemie.uni-hamburg.de

Prof. Barry M. Trost

Department of Chemistry
Stanford University
Stanford, CA 94305-5080
USA
bmtrost@leland.stanford.edu

Prof. Dr. Hisashi Yamamoto

Department of Chemistry
The University of Chicago
5735 South Ellis Avenue
Chicago, IL 60637
USA
yamamoto@uchicago.edu

Prof. Dr. F. Vögtle

Kekulé-Institut für Organische Chemie
und Biochemie
der Universität Bonn
Gerhard-Domagk-Str. 1
53121 Bonn, Germany
voegtle@uni-bonn.de

Topics in Current Chemistry **Also Available Electronically**

For all customers who have a standing order to Topics in Current Chemistry, we offer the electronic version via SpringerLink free of charge. Please contact your librarian who can receive a password or free access to the full articles by registering at:

springerlink.com

If you do not have a subscription, you can still view the tables of contents of the volumes and the abstract of each article by going to the SpringerLink Homepage, clicking on "Browse by Online Libraries", then "Chemical Sciences", and finally choose Topics in Current Chemistry.

You will find information about the

- Editorial Board
- Aims and Scope
- Instructions for Authors
- Sample Contribution

at springer.com using the search function.

Preface

This volume of *Topics in Current Chemistry* presents an overview of atomistic theoretical methods applied to molecular biological systems. It thus represents a bottom-up view of chemistry on biology from a theoretical perspective. The chapters are arranged such that important issues are considered starting from a quantum mechanical perspective and proceeding to a molecular mechanics and molecular dynamics description of the motion of the elementary particles involved (i.e., of electrons and atomic nuclei and then of atoms and molecules), which are responsible for the properties and function of biomolecules.

Depending on the length and time scales relevant for a given phenomenon to be investigated, tailored theoretical methods are required to account for these. If one is interested in large scale motions of molecules, a molecular-mechanics-based description will be appropriate. If, however, chemical reactions at local active sites within extended supramolecular ensembles such as metalloproteins shall be described, a quantum chemical descriptions of electrons and atomic nuclei is most appropriate as such a type of modelling does automatically adjust to any given chemical situation.

While the emphasis in each chapter is on the applicability and usefulness of the various theoretical approaches for a given biological system, the authors made an effort to also provide brief introductions to the foundations of these approaches. It is hoped that each introduction helps to understand the origin of the particular method under consideration. For instance, the first chapter provides a brief introduction to density functional theory and time-dependent density functional theory in the appendix.

This volume is organized in a bottom-up spirit and starts with a review by De Gioia *et al.* on how chemical reactions at the active site of the hydrogenase metalloenzyme can be analyzed with quantum chemical methods, i.e., with those based on the fundamental laws of quantum mechanics.

Quantum chemistry can also provide detailed knowledge on molecular properties as probed by various spectroscopic techniques. These options are discussed in the following two chapters. The first of these by Sinnacker and Neese deals with theoretical prediction and interpretation of resonance spectra obtained, for instance, from electron spin resonance (ESR), nuclear magnetic resonance (NMR), electron–nuclear double resonance (ENDOR) as employed in studies on active sites of metalloenzymes.

The second one by Herrmann and Reiher discusses techniques of vibrational spectroscopy with a focus on how information on local structures within large aggregates of molecules such as proteins can be extracted selectively based on local vibrations or through the selective process of intensity uptake in different vibrational spectroscopy techniques.

While the first three chapters deal with the basic quantities provided by quantum chemistry, namely (stationary) structures, properties and energetics in the electronic ground and excited states, the fourth chapter by Kirchner and collaborators introduces the reader to concepts of how to add the time-dimension to quantum chemical studies. *First-principles* dynamics is introduced in order to describe the Newtonian motion of atomic nuclei on a potential energy surface, which is available on the fly from the electronic structure.

The coupling of quantum chemical and Newtonian mechanics methods is subject of chapter 5 by Senn and Thiel. These so-called QM/MM methods aim at a combination of the best of both worlds. Of course, since the elements of both theories are by and large incompatible (compare the notions of orbitals and positions) a coupling of both can be achieved via the general concept of energy defined in both worlds.

A notorious problem of molecular dynamics is, however, the fact that the simulation time of a single reactive aggregate is usually much too small to observe spontaneous events that require a nonnegligible activation barrier. To overcome this problem, simulation techniques for rare events have been developed. The sixth chapter by Dellago and Bulhuis concentrates on one of these techniques called transition path sampling.

Then, chapter 7 by Schulten and collaborators focuses on molecular mechanics only. Here, the entities described are atoms, united atoms or functional groups. This reduces the number of degrees of freedom and allows one to study the dynamical behaviour of huge protein complexes like ATPase.

Finally, I would like to thank R. Mastalerz for his help concerning formatting issues of this volume.

Zürich, July 2006

M. Reiher

Contents

Quantum Chemical Investigations of Reaction Paths of Metalloenzymes and Biomimetic Models – The Hydrogenase Example

L. Bertini · M. Bruschi · L. de Gioia · P. Fantucci · C. Greco · G. Zampella 1

Theoretical Bioinorganic Spectroscopy

S. Sinnecker · F. Neese 47

First-Principles Approach to Vibrational Spectroscopy of Biomolecules

C. Herrmann · M. Reiher 85

Car–Parrinello Molecular Dynamics Simulations and Biological Systems

J. Thar · W. Reckien · B. Kirchner 133

QM/MM Methods for Biological Systems

H. M. Senn · W. Thiel 173

Transition Path Sampling Simulations of Biological Systems

C. Dellago · P. G. Bolhuis 291

PcrA Helicase, a Molecular Motor Studied from the Electronic to the Functional Level

M. Dittrich · J. Yu · K. Schulten 319

Author Index Volumes 251–271 349

Subject Index 359

Quantum Chemical Investigations of Reaction Paths of Metalloenzymes and Biomimetic Models – The Hydrogenase Example

Luca Bertini¹ · Maurizio Bruschi² · Luca de Gioia¹ (✉) · Piercarlo Fantucci¹ · Claudio Greco¹ · Giuseppe Zampella¹

¹Department of Biotechnology and Biosciences,
 Università degli Studi di Milano-Bicocca, Piazza della Scienza, 2, 20126 Milan, Italy
Luca.degioia@unimib.it

²Department of Environmental Sciences, Università degli Studi di Milano-Bicocca,
 Piazza della Scienza, 2, 20126 Milan, Italy

1	Investigation of Metalloenzymes with Quantum Chemical Methods	3
2	The Hydrogenase Example	5
2.1	[NiFe] and [Fe] Hydrogenases	5
2.2	Biomimetic Models	7
3	Modeling Reactivity	9
3.1	Optimization of Molecular Structures	9
3.2	Investigation of Electronic Properties and Relative Stability of Active Site Models	10
3.3	Theoretical Investigations Related to the Reactivity of Bio-Inspired Coordination Compounds	12
3.4	Theoretical Investigations of the Enzyme Catalytic Cycle	15
4	Modeling the Full H-Cluster	17
4.1	The Broken Symmetry Approach	17
4.2	Application of the Broken Symmetry Approach to the Entire H-Cluster . .	21
5	Modeling Photochemical Reaction Paths	26
5.1	Introduction to Time-Dependent Density Functional Theory (TDDFT) . .	26
5.2	TDDFT Modeling of CO Photolysis in the Carbon Monoxide Inhibited Form of [Fe] Hydrogenase	28
6	Conclusions and Perspectives	32
	Appendix	33
A	Outline of Density Functional Theory	33
B	Outline of Time-Dependent Density Functional Theory	38
	References	41

Abstract Quantum chemical methods allow one to investigate chemical aspects that are often difficult to evaluate using only experimental approaches. In particular, the continuous increase in reliability and speed of quantum chemical methods has recently allowed

the investigation of very complex molecular systems, such as biological macromolecules. In this contribution, we present applications of quantum chemical methods to the investigation of reaction paths of metalloenzymes and related biomimetic models, using hydrogenase models as a reference case. In particular, we discuss several examples from the literature, emphasizing the possibilities (and limitations) offered by present theoretical approaches to study structures, electronic properties and reactivity of metalloenzyme models. Some relevant aspects which have not yet been fully explored using theoretical methods, such as the role of antiferromagnetic coupling and photochemical reactions in [Fe] hydrogenases, are treated in more detail, with presentation and discussion of original data recently obtained in our laboratory.

Keywords Coordination compounds · DFT · Hydrogenases · Metalloenzymes · Quantum chemistry

Abbreviations

B3LYP	Becke3-Lee-Yang-Parr DFT functional
BP86	Becke-Perdew 1986 DFT functional
CI	Configuration interaction
CIS	Configuration interaction singles
DFT	Density functional theory
dppe	1,2-bis(diphenylphosphino)ethane
DTMA	Di(thiomethyl)amine
EDT	1,2 Ethanedithiolate
EPR	Electron paramagnetic resonance
Fe _p	Iron atom of the binuclear cluster proximal to the [Fe ₄ S ₄] cluster in [Fe] hydrogenases
Fe _d	Iron atom of the binuclear cluster distal to the [Fe ₄ S ₄] cluster in [Fe] hydrogenases
G2	Gaussian-2 molecules set
GGA	Generalized gradient approximation
HF	Hartree-Fock
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
KS	Kohn-Sham
LDA	Local density approximation
MP2	Møller-Plesset second order perturbation method
MOs	Molecular orbitals
CT	Charge transfer
o-xyltdt	Orto-xylenedithiolate
PBE	Perdew-Burke-Ernzerhof DFT functional
PDT	1,3-Propanedithiolate
RI	Resolution of identity
TDA	Tamm-Dancoff approximation
TDDFT	Time-Dependent density functional theory
TDHF	Time-dependent Hartree-Fock
TZVP	A triple-zeta basis set
VWN	Vosko-Wilk-Nusair DFT functional
ZORA	Zero-order regular approximation
BS	Broken symmetry

HS High spin
COSMO Conductor-like screening model

1

Investigation of Metalloenzymes with Quantum Chemical Methods

Enzymes are interesting molecules not only due to their crucial biochemical and medical relevance, but also because the elucidation of the molecular properties responsible of their often remarkable catalytic activity can drive the design and synthesis of bio-inspired catalysts with potential technological applications. Several enzymes are known to bind one or more transition metal ions, which often play a key role in the catalytic mechanism. The biological role of transition metals has, in turn, stimulated the synthesis and characterization of coordination compounds featuring structural and/or functional features related to metalloenzymes and several examples are known in which the investigation of biomimetic complexes has complemented the elucidation of the properties of the corresponding metalloenzymes [1].

The tremendous increase in computer power, coupled to the refinement of theories and algorithms, has allowed in recent years the application of quantum chemical methods to the investigation of complex molecular systems containing atoms from almost all the periodic table, thereby disclosing the possibility to investigate the reactivity of metalloenzymes and biomimetic models using theoretical methods [2–7].

Present quantum chemical methods are suited to compute structures and relative energies of reactants, products and intermediate species, as well as transition states, allowing the dissection of reaction paths. Moreover, the possibility to characterize structural and electronic properties of both ground and excited states gives the opportunity to investigate photochemical reactions and spectroscopic properties, which can often be directly compared with experimental data.

In this contribution, the possibilities offered by quantum chemical methods to the investigation of reaction paths of metalloenzymes and related biomimetic models are presented using hydrogenases as reference example. In fact, almost all theoretical investigations of coordination compounds related to the active site of hydrogenases have been carried out using quantum chemical approaches based on Density Functional Theory (DFT) [8–11], which is based on the fundamental theorem that proves the existence of a functional of the electron density that contains all energy contributions, including the so-called correlation contributions, which are related to the mutual interaction of groups of electrons [12]. For a thorough description of DFT, the reader is referred to the seminal book of Parr and Yang [13], whereas a short primer can be found here in Appendix I.

The different description of correlation effects is one of the fundamental differences between DFT and the conventional Hartree–Fock (HF) theory. In fact, the HF method neglects correlation effects because each electron is supposed to move in the mean field provided by all other $N-1$ electrons, even though the HF energy expression includes an exact term for the exchange energy, which could be considered as a sort of correlation for electrons of equal spin. In the past, several methods have been proposed and used to introduce correlation effects via post-HF corrections, all of them including to different extent the configuration interaction (CI) [14], which, in turn, is based on the idea of approximating the exact wavefunction (in the limit of the adopted basis set) building up a combination of N -electron configuration state functions.

The HF method, among the so-called “*ab initio*” approaches, provides the cheapest way to obtain a molecular wave function, whereas the CI expansion (in all possible variants) is usually extremely time consuming. On the contrary, the computational time required by DFT calculations is only slightly longer than for HF, giving at the same time a one-electron structure similar to HF, thus offering a straightforward interpretation of the results while incorporating the correlation effects. Moreover, the description of open-shell systems is more balanced in DFT than in the HF, which generally enhances the importance of high-spin configurations (that are characterized by the highest exchange contributions). Finally, when open-shell systems are treated at the “unrestricted” level (different MOs for different spin) the DFT solution is generally less contaminated by high multiplicity contributions than the HF one. DFT methods can also give excellent results for molecular systems characterized by near-degenerate states, which are generally not adequately described by *ab initio* mono-determinantal perturbation theories (for example MP2). However, limitations affect also DFT methods. The accuracy of DFT results cannot be increased systematically, as it occurs with *ab initio* approaches based on variational principle, which ensures that large CI expansions on large basis sets of N -electron state functions can be arbitrarily close to the (non-relativistic) exact solution. Other limitations of the DFT approach are related to the adopted exchange-correlation functional, for which the exact form is unknown. As a consequence, functional reliability can be only established in a heuristic way by comparing DFT results with experiments or extremely accurate results obtained from highly correlated *ab initio* methods. In this context, coordination compounds represent a challenging case because very accurate *ab initio* results are often unavailable and also experimental results may be scarce. In spite of the above limitations, DFT has become a very useful tool for the investigation of models of metal-containing proteins. In fact, functionals based on GGA and so called hybrid methods, in which a portion of Hartree–Fock exchange energy is added to the exchange-correlation energy, give accuracies similar or higher than *ab initio* MP2 methods. In add-

ition, efficient approaches developed to calculate the Coulomb energy, such as the RI approximation [15–18], have further improved the computational efficiency.

The hybrid three parameter B3LYP and the pure BP86 functionals have been widely used in computational studies of models of the active site of enzymes. The hybrid B3LYP functional [19–22], which was originally calibrated on the G2 database of organic molecules, is now extensively used to study also metal-containing molecules, due to increasing evidence that B3LYP can predict accurately properties like bond dissociation energies and molecular geometries when coupled to an appropriate basis set. Similar considerations hold true for BP86 [23, 24]. Reaction energies can be generally evaluated with a reasonable accuracy, even if they can be affected by the choice of the exchange-correlation functional and basis sets [2, 25]. Other relevant experimental observables such as spin densities, EPR hyperfine coupling constants, g-tensors and vibrational frequencies can be presently computed within the DFT framework with a sufficient accuracy to allow comparison with the corresponding experimental data [2–6].

2

The Hydrogenase Example

Hydrogenases, which are widespread in prokaryotes and found also in some unicellular eukaryotes, are enzymes that catalyze the reversible oxidation of dihydrogen and therefore play a fundamental role in energy metabolism [26]. Most hydrogenases can catalyze the reaction in both directions *in vitro*, although *in vivo* they are usually involved either in H₂ uptake or H₂ evolution, depending on the metabolic state of the organism. Studies of isotope exchange led to the conclusion that dihydrogen cleavage catalyzed by hydrogenases follows a heterolytic reaction path, which implies formation of intermediate hydride species [27].

Experimental and theoretical studies of hydrogenases are not only driven by the biological relevance of these enzymes [28, 29], but also by the possibility to cheaply produce large amount of hydrogen gas, which is a very promising future energy carrier [30].

2.1

[NiFe] and [Fe] Hydrogenases

The two hydrogenase families more thoroughly investigated so far are [NiFe] and [Fe] hydrogenases, which are phylogenetically distinct and named according to the metal ions bound in their active site [31–39]. Remarkably, another evolutionary distinct family of hydrogenases has been recently found to contain an iron cofactor of functional importance [40].

The structural organization of [NiFe] and [Fe] hydrogenases, as well as several peculiar features of the metal cofactors involved in H_2 activation, have been disclosed by X-ray diffraction and spectroscopic studies [41–48]. In the active site of [NiFe] hydrogenases a nickel ion is bound to the protein by four cysteine residues, two of which are also coordinated to an iron ion (Fig. 1). IR spectroscopy studies revealed that the iron ion binds also two CN^- and one CO ligands [49]. In the catalytically inactive form of the enzyme an additional oxygen-containing ligand, which has been recently characterized as a hydroperoxide moiety, bridges the two metal ions [50]. Several redox states for [NiFe] hydrogenases have been characterized by kinetic and spectroscopic studies. Four correspond to paramagnetic forms (usually referred to as Ni-A, Ni-B, Ni-C, Ni-L). The Ni-A form is slowly activated in the presence of H_2 , whereas Ni-B undergoes quick activation. The Ni-C form, which is two electrons more reduced than Ni-B, is catalytically active and upon illumination forms the Ni-L species, which is stable at temperature below 100 K. EPR silent forms (Ni-SU, Ni-SI and Ni-R) have also been identified and might be involved in the catalytic cycle [51].

An unusual $[Fe_6S_6]$ cluster, referred to as H-cluster and formed by a classical $[Fe_4S_4]$ cluster bridged by a cysteine residue to a binuclear subcluster ($[2Fe]_H$), is bound in the active site of [Fe] hydrogenases (Fig. 2).

The two iron atoms of the $[2Fe]_H$ subcluster are coordinated by CO and CN^- ligands and by a chelating $S-X_3-S$ moiety, which has been proposed to be either di(thiomethyl)amine (DTMA) or 1,3-propanedithiolate (PDT). Spectroscopic investigations of [Fe] hydrogenases are consistent with a +2 oxidation state for the $[Fe_4S_4]$ cluster, both in the oxidized and reduced forms

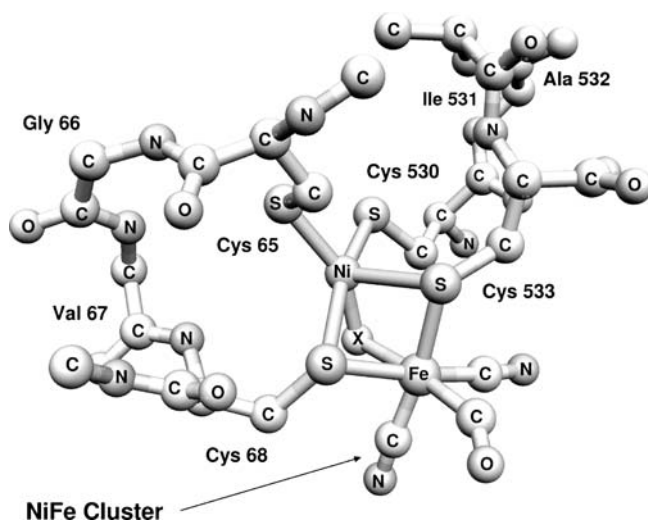


Fig. 1 Structure of the bimetallic cluster found in the active site of [NiFe] hydrogenases

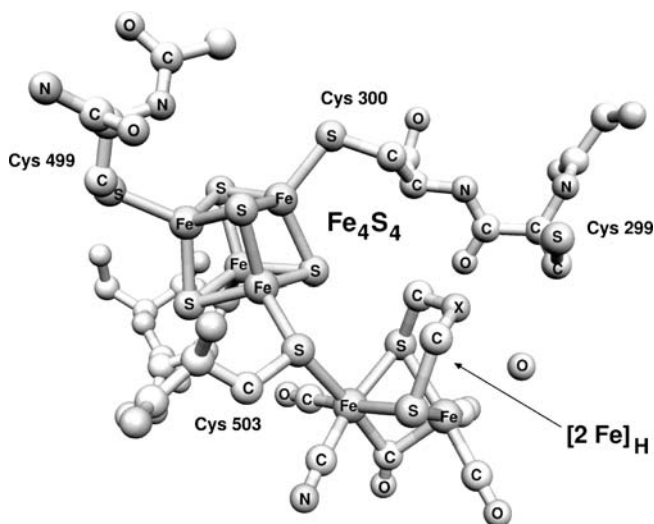


Fig. 2 Structure of the H-cluster found in [Fe] hydrogenases

of the H-cluster [52, 53]. Three redox states for the $[2\text{Fe}]_{\text{H}}$ subcluster have been characterized spectroscopically [54–56]. In the fully oxidized and fully reduced forms the binuclear cluster is EPR silent, whereas the partially oxidized form is paramagnetic. Moreover, in the oxidized inactive form of the enzyme a CO group bridges the iron atoms of the $[2\text{Fe}]_{\text{H}}$ cluster and a coordination site on the iron atom (Fe_{d}) distal relative to the cysteine residue is occupied by an oxygen containing species which is displaced by CO in the CO-inhibited form of the enzyme [57]. In addition, spectroscopic data are consistent with movement of the μ -CO group from bridging to terminal position upon reduction of the enzyme [58].

2.2

Biomimetic Models

The possibility of using molecular hydrogen as a clean energy carrier in the next future is stimulating the development of novel, cheap and eco-compatible processes for H_2 production and oxidation. In this context, insight into electronic and structural features of [NiFe] and [Fe] hydrogenases active sites has driven the synthesis of coordination compounds that could mimic the catalytic properties of the enzymes [59–63] (Fig. 3).

The coordination compound $[\text{Fe}_2(\text{S}_2\text{C}_3\text{H}_6)(\text{CO})_6]$ is the simplest synthetic model of the $[2\text{Fe}]_{\text{H}}$ binuclear subsite found in [Fe] hydrogenases [64–66]. Further synthetic efforts led to model complexes characterized by cyanide and/or phosphine ligands [62, 67, 68] (Fig. 4). Recently Pickett and collaborators reported the synthesis of a exa-carbonyl Fe_6S_6 complex closely resem-

bling the structure of the entire H-cluster [69], while other research groups obtained dinuclear model containing isonitriles [70], carbenes [71] and new DTMA derivatives [72]. Notably, most of the above-mentioned model compounds proved to be able to catalyze the electrochemical reduction of protons [73, 74].

Complexes resembling the cofactor of [NiFe] hydrogenases were initially synthesized by Darensbourg [61] and Pohl [75]. NiFe complexes resembling the enzyme active site were also reported by Evans and coworkers [76, 77], by the Schroder's group [78–80] and by Sellmann and collaborators [81] (Fig. 5).

It is important to underline that the characterization of the chemical properties of synthetic models of hydrogenases has greatly benefited by the application of quantum chemical methods (see below). In this context, theoretical studies have been mainly used to characterize transiently formed species, evaluate the relative stability of isomers, dissect reaction mechanisms and propose the synthesis of novel coordination compounds that are expected to show catalytic activity.

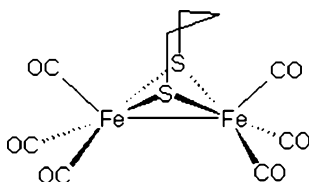


Fig. 3 Structure of the simplest synthetic model of the $[2\text{Fe}]_{\text{H}}$ cluster

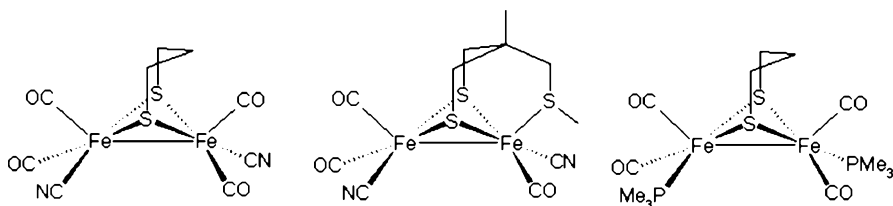


Fig. 4 Structures of recently reported synthetic models of the $[2\text{Fe}]_{\text{H}}$ cluster

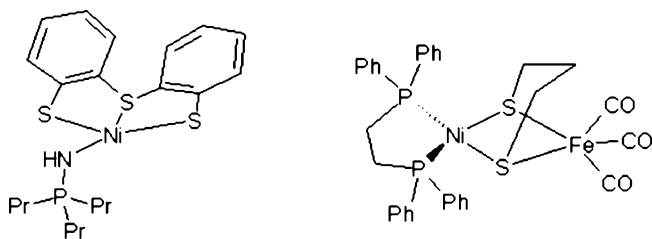


Fig. 5 Structures of some functional and structural synthetic models of the [NiFe] hydrogenase bimetallic cluster

3 Modeling Reactivity

3.1 Optimization of Molecular Structures

The ability of quantum chemical methods to accurately reproduce the experimental geometry of coordination compounds is a fundamental prerequisite for the investigation of reactivity. While all quantum chemical methods generally perform very well when used to predict the structure of organic molecules, their performance can be less satisfactory when dealing with coordination compounds [7]. It is known that HF calculations generally lead to too long metal–ligand bonds, whereas MP2 calculations generally overcorrects for the HF error when considering coordination compounds. Therefore, CCSD(T) calculations are often necessary to obtain accurate M–L distances. The comparison of experimental and computed M–L bond distances reveals that, among DFT methods, those adopting the GGA scheme generally lead to results in very close agreement with experiment. However, it has been noted that GGA methods tend to overestimate M–L bond lengths when L = halide or chalcogenide. Bond distances obtained with the hybrid functional B3LYP generally compare well with those obtained with GGA methods [7].

As a consequence of the complexity of hydrogenases and experimental limitations, there has been great interest in applying theoretical methods, particularly density functional theory, to investigate structures of models of the different redox states of hydrogenases. Most calculations reported so far have been carried out to investigate models inspired to the active site of [Fe] hydrogenases, for which more experimental data to be used as reference are available. The theoretical investigation of the structural features of coordination compounds related to [Fe] hydrogenases has been carried out almost exclusively using DFT methods. In particular, most results have been obtained using the pure BP86 and the hybrid B3LYP functionals. Bond distances are generally reproduced accurately by both functionals. However, Ni–S distances are systematically overestimated by about 0.05 Å by B3LYP [82], whereas BP86 structures are affected by slightly smaller deviations when compared to experimental data [83].

Structures of $[\text{Fe}(\text{CN})_x(\text{CO})_y]$ complexes have also been recently investigated at the BP86 level of theory, showing that computed Fe–CO bond distances are in very close agreement with X-ray data, whereas Fe–CN bond lengths are generally slightly overestimated [84].

The capability of B3LYP/TZVP and BP86/TZVP to accurately reproduce experimental structures has been compared also for the coordination compound $[\text{Ni}(\text{NHP}n\text{Pr}_3)(\text{S}_3)]$, which is related to the active site of [NiFe] hydrogenases. Notably, both functionals reproduce accurately the peculiar flattened tetrahedral geometry of the complex [85].

FT-IR spectra are extremely useful for the characterization of intermediate species and redox states of [NiFe] and [Fe] hydrogenases, as well as of related model complexes, due to the peculiar vibrational frequencies of the CN^- and CO groups. In particular, the calculation of vibrational frequencies and their comparison with experimental data have been used extensively to validate theoretical protocols and to predict the structural features of intermediate species relevant to [NiFe] and [Fe] hydrogenase chemistry (see below). Note that the computation of vibrational frequencies is also necessary to check the curvature of the potential energy surface around the stationary point, in order to exactly locate a true minimum (where all vibrational frequencies are positive) or a transition state (one imaginary vibrational frequency), and to calculate the zero point energy correction to the pure electronic energy in order to obtain thermodynamic data. As for the accuracy of computed vibrational frequencies, it is well known that, due to the harmonic approximation adopted in the calculation of force constants, *ab initio* computed vibrational frequencies are systematically higher than the experimental ones [25]. Therefore, empirical correction factors are often used to improve the agreement with the experiment. DFT vibrational frequencies are generally more accurate than HF ones and similar to those obtained with correlated *ab initio* methods. In the specific case of CO and CN^- groups in [NiFe] and [Fe] hydrogenase models, wave numbers obtained from BP86 calculations with harmonic approximation compare extremely well with experimental data. Such an excellent agreement may be partially due to error cancellation effects [86]. In fact, several groups have evaluated the relationship between computed and experimental CO and CN^- vibrational frequencies in models related to the hydrogenases active sites, observing excellent correlation between calculated frequencies and experimental spectra [87, 88].

3.2

Investigation of Electronic Properties and Relative Stability of Active Site Models

The evaluation of the relative stability of isomers, as well as insight into reaction pathways leading to their inter-conversion, is often crucial to address chemically relevant problems in hydrogenase chemistry.

In general, computed free energy values are obtained from the electronic SCF energy considering translational, rotational and vibrational contributions to the total partition function, under the assumption that the partition function can be written as the product of such terms. To evaluate enthalpy and entropy contributions, the values for temperature and pressure have to be set according to the conditions that have to be modeled. Moreover, a scaling factor for the SCF wavenumbers is usually employed. Rotations are usually treated classically and vibrational modes are generally described according to harmonic approximations.

In the field of coordination compounds related to [Fe] hydrogenases, the systematic investigation of $[\text{Fe}(\text{CN})(\text{PMe}_3)(\text{CO})]_2(\mu\text{-CO})(\text{S}_2\text{C}_n\text{H}_{2n})$ isomers was undertaken using DFT to rationalize experimental data [89]. Interestingly, it turned out that the most stable isomer does not correspond to the crystallographic structure. The cause for the different stability ranking between experiment and DFT is to be ascribed to crystal packing forces, which could not be taken into account in DFT calculations. In fact, four isomers of $[\text{Fe}(\text{CN})(\text{PMe}_3)(\text{CO})]_2(\mu\text{-CO})(\text{S}_2\text{C}_n\text{H}_{2n})$ were characterized in solution by spectroscopic techniques and the properties of the most abundant isomer well compare with those computed for the most stable isomer predicted by DFT calculations.

DFT was also used to predict the regiochemistry of protonation in models of the active site of [Fe] hydrogenases, evaluating the energy difference among the possible isomers of $[\text{HFe}_2(\text{S}_2\text{C}_3\text{H}_6)(\text{CN})(\text{CO})_4(\text{PMe}_3)]$ [90].

The comparison between the $[2\text{Fe}]_{\text{H}}$ cluster in the active site of [Fe] hydrogenases and related synthetic model compounds has stimulated several theoretical investigations aimed at disclosing the properties of the active site responsible for the peculiar structure of the enzymatic cluster. In fact, even though several synthetic models reproduce the main structural features of the $[2\text{Fe}]_{\text{H}}$ cluster well, one striking difference is evident comparing $\text{Fe}(\text{I})\text{Fe}(\text{I})$ synthetic models and the $[2\text{Fe}]_{\text{H}}$ cluster. The former are characterized by a vacant coordination site along the Fe–Fe axis, whereas in the enzyme a CO group coordinated to Fe_{d} approaches the other iron ion and a vacant coordination site appears on Fe_{d} . To shed light on this issue, Darensbourg, Hall and coworkers investigated factors affecting the rotation of the $\text{Fe}(\text{CO})_3$ group in the series of binuclear complexes $[\text{Fe}_2(\mu\text{-SRS})(\text{CO})_6]$ in which R = EDT, PDT or o-xyldt (orto-xylenedithiolate) [82]. The computation of stable species and transition states showed that $\text{Fe}(\text{CO})_3$ rotation leading to $\mu\text{-CO}$ species is easy, leading to the suggestion that the protein environment could fine tune the geometry of the binuclear cluster. Electronic factors affecting the structure of the $[2\text{Fe}]_{\text{H}}$ cluster were investigated, using DFT calculations, also by Bruschi et al., who studied a series of $\text{Fe}(\text{I})\text{Fe}(\text{I})$ complexes differing for the nature of the ligand corresponding to the cysteine residue coordinated to the $[2\text{Fe}]_{\text{H}}$ cluster in the protein [91]. The results highlight that increased rotation of the $\text{Fe}_{\text{d}}(\text{CO})_2(\text{CN})$ group was observed going from soft to hard ligands, leading to structures that closely resembled the $[2\text{Fe}]_{\text{H}}$ cluster when the ligand is a strong electron-donor group. These observations suggested that the proximal $[\text{Fe}_4\text{S}_4]$ cluster could play a role in modulating the structural and electronic properties of the $[2\text{Fe}]_{\text{H}}$ cluster in the enzyme.

The above-mentioned results nicely show how the investigation of very simple models is not always a shortcoming of quantum chemical investigations. In fact, the possibility to neglect environmental effects and to investigate series of related compounds (which could be very difficult to obtain

experimentally) can allow one to highlight relationships that otherwise could be difficult to unravel.

The investigation of structural and electronic properties of Fe(I)Fe(I) models related to the reduced form of the $[2\text{Fe}]_{\text{H}}$ cluster has also allowed the disclosure of key factors affecting H^+ binding to the metal cofactor. Darensbourg, Hall and collaborators have used molecular orbital analysis to show that, as a consequence of rotation of the $\text{Fe}(\text{CO})_3$ group, the highest occupied molecular orbital (HOMO) of the $[2\text{Fe}]_{\text{H}}$ cluster changes drastically, resulting in partial disruption of the Fe – Fe bond density and development of a charge dipole along the Fe – Fe axis. Moreover, the iron center corresponding to Fe_d in the $[2\text{Fe}]_{\text{H}}$ subcluster becomes more electrophilic [82].

Starting from the observation that most synthetic Fe(I)Fe(I) complexes fail to reproduce the orientation of diatomic ligands around the iron ions that is observed in the enzyme active site, and assuming that this difference could be crucial to explain the different catalytic activity of synthetic and enzymatic systems, Darensbourg, Hall and collaborators have used DFT to design coordination compounds that should be characterized by structural features more closely resembling the enzyme active site [92]. To this end they explored structures, electronic properties, and relative stabilities of a broad series of binuclear complexes differing for the nature of the dithiolate linker and of the other ligands.

As for $[\text{NiFe}]$ hydrogenase models, DFT was recently used to evaluate the possible formation of a Ni – Fe bond in the diamagnetic complex $[(\text{dppe})\text{Ni}(\mu\text{-PDT})\text{Fe}(\text{CO})_3]$, keeping into account the possibility that Ni and Fe fragments could have open-shell character. With this aim, Schroder, Neese and colleagues carried out geometry optimizations both for singlet and triplet states, starting from the assumption that geometry optimization for ferromagnetic alignment is generally a good approximation to the geometry of antiferromagnetically coupled centers. It turned out that the geometry of the closed-shell singlet state is in better agreement with experimental data, providing strong evidence against open-shell character in $[(\text{dppe})\text{Ni}(\mu\text{-PDT})\text{Fe}(\text{CO})_3]$ and suggesting that a strong two-center, two-electron bond is formed between Ni and Fe ions [79].

3.3

Theoretical Investigations Related to the Reactivity of Bio-Inspired Coordination Compounds

Quantum chemical investigations have been widely used to dissect reaction pathways leading to bio-inspired synthetic models related to the active site of hydrogenases. Indeed, theoretical approaches can well complement experimental investigations in this field because it is often difficult to characterize the structure and electronic properties of short-lived intermediates. In addition, quantum chemical methods may allow one to disclose the steric and

electronic factors at the basis of peculiar reactivity properties, as well as to explore the effect of chemical modifications not yet sampled experimentally. In this context, it is also important to underline that the theoretical investigation of a reaction path not only implies the characterization of intermediate species but also of transition state structures. The optimization of transition state structures by quantum chemical methods still depends very much on the chemical intuition of the computational chemist. Several strategies have been proposed to efficiently locate transition states and the reader is referred to specific texts for a thorough discussion of the topic [2, 25, 93]. In our laboratory, the optimization of transition state structures is generally carried out according to a procedure based on a pseudo Newton–Raphson method. Initially, geometry optimization of a guessed transition state structure is carried out constraining the distances corresponding to the reaction coordinate. Then, vibrational analysis of the constrained minimum energy structure is carried out and, if one negative eigenmode corresponding to the reaction coordinate is found, the curvature determined at such a point is used as starting point in the transition state search. Localization of the transition state is carried out using an eigenvector-following search: the eigenvectors in the Hessian are sorted in ascending order, the first one being that associated with the negative eigenvalue. However, after the first step the search is performed by choosing the critical eigenvector with a maximum overlap criterion, which is based on the dot product with the eigenvector followed in the previous step.

Hall, Darensbourg and collaborators have characterized reaction paths related to substitution at di-nuclear carbonyl species, shedding light onto the role of μ -CO and neighboring groups [82]. In particular, the cyanation of the complex $[\text{Fe}_2(\mu\text{-PDT})(\text{CO})_6]$ was investigated by DFT, showing that the CN/CO substitution leading to $[\text{Fe}_2(\text{SCH}_2\text{CH}_2\text{CH}_2\text{S})(\text{CO})_4(\text{CN})_2]^{2-}$ takes place in two steps characterized by associative mechanism. In particular, DFT calculations allowed to show that the CN^- group initially reacts with the less sterically hindered iron ion, leading to a μ -CO intermediate species. The first cyanation step is followed by flip of the PDT linker, leading to a species in which the $\text{Fe}(\text{CO})_3$ moiety is sterically unhindered. Then, another CN^- group reacts with the $\text{Fe}(\text{CO})_3$ unit, going through a transition state characterized by a semibridging CO.

Zampella et al. [83] have used DFT to investigate the mechanism of cyanation of the complex $[\text{Fe}_2(\text{CO})_5(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)]$, which is the first example of a coordination compound characterized by the $[\text{Fe}_2\text{S}_3]$ ligation observed in the $[2\text{Fe}]_{\text{H}}$ cluster [94]. According to experimental results, the reaction of $[\text{Fe}_2(\text{CO})_5(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)]$ with cyanide takes place at the Fe atom distal to the thioether ligand, leading to $[\text{Fe}_2(\text{CO})_5(\text{CN})(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)]$, which spontaneously loses CO. The resulting complex can react again with CN^- , leading to a species which was assigned to $[\text{Fe}_2(\text{CO})_4(\mu\text{-CO})(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)(\text{CN})_2]^{2-}$. How-

ever, peculiar features of the reaction pathways could not be fully addressed on the basis of experimental results. In particular, the first cyanation step was about 10 000-times faster than the corresponding process for $[\text{Fe}_2\text{S}_2]$ species. In addition, the second cyanation step on the $[\text{Fe}_2\text{S}_3]$ species was kinetically hindered and spectroscopic data suggested the formation of long-lived μ -CO species. The dissection of the reaction path within the DFT approach allowed to show that cyanation of $[\text{Fe}_2(\text{CO})_5(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)]$, which implies a transition state structure featuring a μ -CO group, is characterized by a low energy barrier due to stabilization of the μ -CO group by the thioether ligand. In addition, the second cyanation step is kinetically hindered due to deactivation of the Fe ion and hindered rotation of the reacting FeL_3 group. Finally, close correspondence between computed and experimental FT-IR data supported the assignment of $[\text{Fe}_2(\text{CO})_4(\mu\text{-CO})(\text{MeSCH}_2\text{C}(\text{Me})(\text{CH}_2\text{S})_2)(\text{CN})_2]^{2-}$ to a μ -CO species.

DFT calculations have been used to complement experimental results also by Rauchfuss, Benard and coworkers, which have shown that the peculiar structural properties of the coordination compound $[\text{Fe}_2(\text{SCH}_2\text{NMeCH}_2\text{S})(\text{CO})_6]$ are a consequence of the balance between two different factors: the anomeric effect, which favors the axial disposition of the methyl group, and steric repulsion between the axial methyl and one carbonyl ligand [95]. DFT was also used to elucidate factors affecting the regiochemistry of protonation in bimetallic carbonyl thiolates, showing that protonation at the Fe – Fe bond and at CN[−] ligands can be modulated by the nature of coligands, occurring more easily at the Fe – Fe bond in $[\text{Fe}_2(\text{PDT})(\text{CN})(\text{CO})_4(\text{PMe}_3)]^-$ than in $[\text{Fe}_2(\text{PDT})(\text{CN})(\text{CO})_4(\text{PH}_3)]^-$ or $[\text{Fe}_2(\text{PDT})(\text{CN})(\text{CO})_4(\text{P}(\text{OMe})_3)]^-$ [90].

DFT calculations have been also carried out to elucidate the catalytic properties of models of [NiFe]-hydrogenase active site. As an example, Zampella et al. have recently used DFT to investigate the reaction between H_2 and models of the complex $[\text{Ni}(\text{NHP}n\text{Pr}_3)(\text{'S3'})]$ (where 'S3' = bis(2-sulfanyl-phenyl)sulfide^{2−}), with the aim of clarifying the most important factors contributing to its remarkable catalytic activity [85]. In fact, $[\text{Ni}(\text{NHP}n\text{Pr}_3)(\text{'S3'})]$ is one of the few known coordination compounds featuring a Ni thiolate core resembling the enzyme active site and capable of catalyzing H_2 cleavage. The DFT characterization of transient intermediate species and transition states revealed that formation of a non-classical adduct between $[\text{Ni}(\text{NHP}n\text{Pr}_3)(\text{'S3'})]$ and H_2 implies rearrangement of the Ni coordination geometry, which can take place due to the peculiar strongly flattened tetrahedral geometry of the parent complex. Moreover, the theoretical investigation showed that H_2 cleavage is achieved according to an oxidative addition step followed by hydrogen migration to one of the sulfur ligands.

3.4

Theoretical Investigations of the Enzyme Catalytic Cycle

The most widely studied issue related to hydrogenase chemistry concerns the investigation of the mechanism of H_2 cleavage/production in the active site of [NiFe] and [Fe] hydrogenases. However, the elucidation of structure-function relationship in the protein poses formidable problems to the theoretical community. In spite of continuous increase in computing power and development of efficient computer codes, the quantum chemical investigation of entire proteins is still a long-term goal. Nevertheless, theoretical investigation of simpler models of the hydrogenase active site has allowed to shed light on several key factors relevant to understand the enzyme catalytic cycle.

Binding and activation of H_2 on models of the $[2Fe]_H$ cluster found in [Fe] hydrogenases has been investigated by several groups. Prompted by the observation that Fe_d has been suggested to be the site of H_2 binding and activation, DFT was used to study H_2 heterolytic cleavage taking place on the model $[(H_2)(CO)_2(CN)Fe(\mu-PDT)Fe(CO)(CN)(SMe)]^-$ [88, 96–99]. Theoretical characterization of several plausible intermediate species, as well as comparison between computed and experimental IR frequencies, led to the conclusion that heterolytic cleavage of H_2 on a $Fe(II)Fe(II)$ redox form with concomitant protonation of a S atom or a CN^- group is kinetically or thermodynamically unfavorable. Notably, when PDT was replaced by DTMA, cleavage of H_2 , mediated by the N atom of DTMA, became kinetically and thermodynamically favorable. Consequently, on the ground of DFT results it was proposed a catalytic cycle in which binding of H_2 takes place at the vacant coordination site of a $\mu-CO$ $Fe(II)Fe(I)$ species. Subsequent one-electron oxidation leads to a $Fe(II)Fe(II)$ species which is capable to activate H_2 , leading to a formal $Fe(II)Fe(II)$ intermediate characterized by protonated DTMA and terminal hydride coordination to Fe_d . Release of $2H^+$ and $2e^-$ closes the catalytic cycle.

DFT was also used to explore other plausible routes of H_2 activation catalyzed by models of the $[2Fe]_H$ cluster, with the aim of better understanding the chemistry of synthetic coordination compounds and possibly drive the design of new catalysts inspired to the active site of [Fe] hydrogenase. In particular, starting from theoretical [100, 101] and experimental [58] observations indicating that $[2Fe]_H$ models characterized by bridged and terminal CO can easily interconvert, Bruschi et al. have investigated H_2 cleavage catalyzed by $[(CO)_2(CN)Fe(\mu-PDT)Fe(CO)(CN)(CH_3S)]^-$ model complex, according to a path that implies formation of $\mu-H$ intermediates. Results showed that H_2 heterolytic cleavage can take place on Fe_p , leading to protonation of a S atom of PDT and concomitant coordination of H^- to both metal centers. Zhou et al. have reached analogous conclusions [102, 103], showing also that a $\mu-H$ $Fe(II)Fe(II)$ intermediate cannot be easily reduced because all the low-lying $3d$ bonding orbitals of the iron ions are occupied and the LUMO is characterized by high energy.

Theoretical investigations based on quantum chemical approaches have been aimed at investigating crucial aspects related to the catalytic cycle of [NiFe] hydrogenases. Even though most quantum chemical studies published so far were carried out using DFT, different computational strategies were used to gain insight into key features of intermediate species formed in the catalytic cycle [8–11]. Several groups have evaluated relative stabilities of possible intermediate species, as well as energy barriers associated to heterolytic cleavage of dihydrogen, to propose plausible catalytic cycles [96, 99, 104–106]. Others have compared experimental and computed spectroscopic properties, such as g tensors and hyperfine tensors, to discern among different structures of the NiFe cluster [107–114]. In particular, calculations have been carried out within the ZORA [115, 116], using both local VWN [117] and gradient corrected (BP86) functionals. Even though further improvements in the accuracy of computed magnetic properties is needed and might be expected to arise when simultaneously considering spin-polarization and spin-orbit coupling [118], theoretical results generally agree reasonably well with experimental data.

Quantum chemical results obtained studying [NiFe] hydrogenase models converge on many details, such as the proposal that μ -H species should be formed along the catalytic cycle. However, many issues, such as the site of H₂ activation, as well as the redox state of the metal ions in the intermediate species remain unclear. Another relevant and still controversial issue of [NiFe] hydrogenase chemistry concerns the nature of the ground state of Ni(II) in Ni-SI and Ni-R redox states, which have been proposed to be involved in the catalytic cycle. In fact, several spectroscopic data are compatible with a diamagnetic ($S = 0$) ground state for the Ni(II) ion [119, 120]. However, L-edge soft X-ray data have pointed to possible formation of high-spin ($S = 1$) Ni(II) species in the catalytic cycle [121, 122]. With the aim of shedding light on this crucial aspect, Hall and coworkers have carried out DFT calculations, using the B3LYP functional, showing that the Ni coordination geometry in high spin (triplet) models fits well to the distorted tetrahedral structure of the bimetallic cluster observed in the enzyme [99]. However, other aspects related to the relative stability of low and high spin states of the Ni(II) ion in the catalytic cycle are less easily addressable by DFT calculations. In fact, it has been recently shown that the both hybrid and non-hybrid functionals can predict the wrong multiplicity of the ground state for some transition metal complexes [123–126]. In the hydrogenase field, Bruschi et al. have recently compared the performance of three functionals (BP86, B3LYP and B3LYP* [86, 127]) in reproducing the structural and electronic properties of well-characterized high- and low-spin [Ni(II)S₄] complexes [91]. Results led to the conclusion that BP86 is the most suited functional to reproduce the structures of [Ni(II)S₄] complexes, whereas the prediction of the ground state, as well as the relative stability of high and low spin states, is more problematic and may be dependent on the adopted functional. It

was also concluded that both Ni-SI and Ni-R are potential spin-crossover species.

4

Modeling the Full H-Cluster

4.1

The Broken Symmetry Approach

As discussed in the previous sections, much of the theoretical work carried out in order to understand the mechanism for hydrogen production in [Fe] hydrogenases has been focused on the $[2\text{Fe}]_{\text{H}}$ cluster and related models. However, in the enzyme the $[2\text{Fe}]_{\text{H}}$ unit is bridged by a cysteine residue to a $[\text{Fe}_4\text{S}_4]$ cluster (Fig. 6), which could play an important role in the modulation of the stereo-electronic properties of the $[2\text{Fe}]_{\text{H}}$ cluster, and consequently, on its catalytic features.

Mössbauer studies carried out on the isolated forms of [Fe] hydrogenases have shown that the $[\text{Fe}_4\text{S}_4]$ cluster is always in the formally diamagnetic oxidation state +2, indicating that the changes in the oxidation state occur only at the $[2\text{Fe}]_{\text{H}}$ cluster [52, 53]. However, Fe atoms of the $[\text{Fe}_4\text{S}_4]$ cluster acquire a paramagnetic character due to the exchange coupling between the two sub-units of the H-cluster. These experimental observations further support the hypothesis that the $[\text{Fe}_4\text{S}_4]$ cluster could play an important role in modulating redox and magnetic properties of the active site. Furthermore, as known from the chemistry of other proteins, $[\text{Fe}_4\text{S}_4]$ clusters are often involved in electron transfer. Therefore, transient species in which the $[\text{Fe}_4\text{S}_4]$ unit is reduced to the oxidation state +1 should also be considered in order to model the reactivity of the full H-cluster.

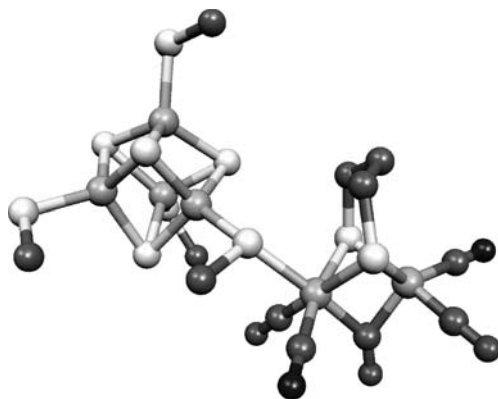


Fig. 6 Schematic representation of the entire H-cluster

The theoretical modeling of the entire H cluster represents a difficult challenge, not only for the size of the system, but also because the $[\text{Fe}_4\text{S}_4]$ cluster is composed by two layers of high spin Fe atoms coupled antiferromagnetically to give a global low spin ground state. For example, in the oxidized $[\text{Fe}_4\text{S}_4]^{2+}$ cluster the two layers, composed by high spin Fe^{2+} ($S_1 = 2$) and high spin Fe^{3+} ($S_2 = 5/2$) atoms with aligned spins (ferromagnetic interaction) ($S_{\text{tot}} = S_1 + S_2 = 9/2$), are coupled antiferromagnetically to give a diamagnetic ($S = 0$) ground state, while in the reduced $[\text{Fe}_4\text{S}_4]^+$ cluster there is one $\text{Fe}^{2+} - \text{Fe}^{3+}$ ($S_1 = 9/2$) and one $\text{Fe}^{2+} - \text{Fe}^{2+}$ layer ($S_2 = 4$) coupled antiferromagnetically to give the $S = 1/2$ ground state. Such spin-coupled systems are difficult to treat within the DFT approach because their ground state wavefunctions typically correspond to linear combinations of multiple determinants.

To discuss energies and properties of $[\text{Fe}_4\text{S}_4]$ clusters, it is important to describe the weak antiferromagnetic coupling at the same level of theory of strong metal-metal and metal-ligand bonds. In the framework of the unrestricted formalism, these exchange interactions can be treated reasonably well by employing the broken symmetry (BS) approach developed by Noodleman et al. [128, 129]. The BS approach consists in the localization of the opposite spins on different parts of the molecule to give a mono-determinant representation of the spin exchange interactions within the molecule. For example, in $[\text{Fe}_4\text{S}_4]$ clusters a BS wavefunction can be built by localizing the electrons on one of the dimeric layers of Fe atoms with spin up and on the other one featuring spin down. In order to localize the spins on different centers, space and spin wavefunctions are relaxed by reducing their symmetry with respect to that of the nuclear framework. The BS wavefunction is not a pure spin state (it is an eigenstate of S_z , but not of S^2). In this context, the BS wavefunction can be considered as a low spin wavefunction affected by a large spin contamination. In spite of this restriction, Noodleman [128, 129] showed that the BS state is a weighted average of pure-spin wavefunctions (where the weights are the Clebsch–Gordan coefficients relating the spin quantum numbers for oppositely aligned spin vectors of the subunits) and that correspondingly the BS state energy is a weighted average of pure spin state energies. The BS state energy can then be correlated to the spin state energies derived from the phenomenological Heisenberg Hamiltonian, which for two spin centers takes the form:

$$H = J (\mathbf{S}_A \cdot \mathbf{S}_B) , \quad (1)$$

where $\mathbf{S}_A$ and $\mathbf{S}_B$ are the spin operators on center A and B, respectively, and J is the exchange coupling constant, which represents the strength of the exchange interaction between the two spin centers. In the following, it will be assumed that J has negative values for antiferromagnetic interactions, and positive values for ferromagnetic interactions. In this case, the Heisenberg

Hamiltonian gives the spin state energies

$$E(S_{\text{tot}}) = 1/2J(S_{\text{tot}}(S_{\text{tot}} + 1)) , \quad (2)$$

and the difference between the energies of two successive spin states equals to

$$E(S_{\text{tot}}) - E(S_{\text{tot}} - 1) = JS_{\text{tot}} , \quad (3)$$

where S_{tot} is the total spin ranging from $|S_A - S_B|$ to $|S_A + S_B|$. The determination of J from Eq. 2 allows one to construct the full ladder of pure spin states energies. Notably, the J value can be determined from the BS wavefunction and the pure high spin wavefunction which is easily calculated by aligning ferromagnetically the spins of the iron sites. In fact, Noodleman showed that the expectation values of the spin operator product defined in Eq. 1 can be derived for the HS and BS states as [129]:

$$\langle \mathbf{S}_A \cdot \mathbf{S}_B \rangle_{\text{HS}} = + S_A S_B \quad (4)$$

$$\langle \mathbf{S}_A \cdot \mathbf{S}_B \rangle_{\text{BS}} = - S_A S_B , \quad (5)$$

from which the difference between HS and BS state energies can be written as:

$$E(\text{HS}; S_{\text{max}} = S_A + S_B) - E(\text{BS}; M_S = |S_A - S_B|) = 2JS_A S_B . \quad (6)$$

Equation 6 can, therefore, be used to derive J , which in turn can then be used in projecting the BS state energy on any pure spin state energy. In particular, the energy difference between the high spin and the low spin ground state energy ($E(S_{\text{max}}) - E(S_{\text{min}})$) can be derived from Eq. 2 as:

$$E(S_{\text{max}}) - E(S_{\text{min}}) = 1/2J(S_{\text{max}}(S_{\text{max}} + 1)) - 1/2J(S_{\text{min}}(S_{\text{min}} + 1)) , \quad (7)$$

and the difference between this equation and Eq. 6 gives $\Delta E_{\text{spin}} = E(\text{BS}) - E(S_{\text{min}})$, which represents the energy correction of the BS state energy to the pure spin ground state energy.

It should be noted that the use of the Heisenberg Hamiltonian is restricted to weakly interacting dimers, for which unpaired electrons are well localized onto the two metal centers. In the case of strongly interacting dimers the unpaired electrons are largely delocalized onto the ligands and $E(\text{BS}) \approx E(S_{\text{min}})$.

The description of the BS approach given above for a two spin centers system can be easily generalized to a system with more than two spin centers by introducing the spin Hamiltonian operator

$$H = \sum_{ij} J_{ij} (\mathbf{S}_i \cdot \mathbf{S}_j) , \quad (8)$$

where the summation is extended over all pairs of spin centers i and j bearing spin S_i and S_j , and J_{ij} is the exchange coupling constant between them. In these multiple spin centers systems several BS states can be calculated by aligning ferromagnetically or antiferromagnetically the spins of different centers. For N spin centers, $2^N/2$ non-equivalent combinations of spins can be defined to obtain BS states. The J_{ij} parameters can then be determined from

the HS state and a consistent number of BS states by applying the spin algebra of Eqs. 4 and 5. In order to reduce the number of coupling constants to be determined, the approximation of restricting the summation in Eq. 8 to adjacent spin centers is generally assumed. Other simplifying assumptions, such as the dimer approximation [130], have been proposed.

The electronic structure of the $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]^x$ cluster, going from the all ferric ($x = 0$) to the all ferrous redox states ($x = -4$), have been extensively investigated by Noodleman et al. [131–135]. Spectroscopic studies on $[\text{Fe}_4\text{S}_4]$ clusters in synthetic complexes and in proteins [136–140] have shown that the four Fe atoms generally occur in equivalent pairs, thus supporting the use of a spin Hamiltonian defined in terms of pairs of centers. For the highest ($x = 0$) and lowest oxidation states ($x = -4$), the two $\text{Fe}^{2+} - \text{Fe}^{2+}$ or $\text{Fe}^{3+} - \text{Fe}^{3+}$ layers coupled antiferromagnetically are treated as single spin centers with $S = 4$ and $S = 5$, respectively, and the correction for the energy of the diamagnetic pure spin state is readily found by applying the equations given above. In the case of the $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]^{-2}$ cluster, which is composed by the two mixed valence $\text{Fe}^{3+} - \text{Fe}^{2+}$ dimers, a further complication arises. Mössbauer studies have shown that the four Fe atoms are equivalent, indicating delocalization of the valence electron distribution to give a proper $\text{Fe}^{2.5+} - \text{Fe}^{2.5+}$ oxidation state for each layer. This intralayer delocalization can be described by introducing a non-Heisenberg term into the spin Hamiltonian operator which produces a splitting of the spin states [141–144]. The spin Hamiltonian operator takes the form

$$H = J (\mathbf{S}_A \cdot \mathbf{S}_B) \pm B (S_{12} + 1/2) \pm B (S_{34} + 1/2) , \quad (9)$$

where B is the resonance delocalization parameter and S_{12} and S_{34} are the total spin of the two mixed-valence layers. The spin state energies obtained from this Hamiltonian are given by

$$E(S) = 1/2J (S(S+1)) \pm B (S_{12} + 1/2) \pm B (S_{34} + 1/2) . \quad (10)$$

It is interesting to note that the delocalization energy depends on the spin quantum number (S_{ij}) of the mixed valence pair and that it increases with the pair spin. Several authors have shown that the ground spin state of the system is generally the result of the competition between the Heisenberg and resonance delocalization of the Hamiltonian [142, 145, 146].

The $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]$ clusters in the oxidation states -1 and -3 are composed by a mixed-valence $\text{Fe}^{2.5+} - \text{Fe}^{2.5+}$ layer and by a $\text{Fe}^{3+} - \text{Fe}^{3+}$ or $\text{Fe}^{2+} - \text{Fe}^{2+}$ layer, respectively. In these cases a third parameter, which describes the possible interlayer resonance delocalization between the ferrous or ferric layers and the mixed-valence layer, has to be added to the spin Hamiltonian operator. This term has the same form of that used for the intralayer delocalization and leads to the expression of the operator

$$H = J (\mathbf{S}_A \cdot \mathbf{S}_B) \pm B (S_{12} + 1/2) \pm B' (S + 1/2) , \quad (11)$$

and to the corresponding energies

$$E(S) = 1/2J (S(S+1)) \pm B (S_{12} + 1/2) \pm B' (S + 1/2) , \quad (12)$$

where B' is the interlayer resonance delocalization parameter and S is the total spin of the cluster. The values of J calculated for $[\text{Fe}_4\text{S}_4]$ clusters in different redox states ranges from about 100 cm^{-1} in the all-ferrous $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]^{-4}$ cluster to about 900 cm^{-1} in the all-ferric $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]^0$ cluster [133] while the calculated values of B are in the range $700\text{--}900 \text{ cm}^{-1}$ [132]. For a detailed explanation of the approaches used for the calculation of the exchange parameters the reader is referred to the extensive work of Noodleman et al. In particular, analytical expressions for the energy correction from the BS to the pure spin state for the $[\text{Fe}_4\text{S}_4(\text{SCH}_3)_4]$ cluster in different redox states are reported in [132, 133].

4.2

Application of the Broken Symmetry Approach to the Entire H-Cluster

The application of the BS approach to the H-cluster should account not only for exchange interactions within the $[\text{Fe}_4\text{S}_4]$ cluster but also for the exchange interaction between the $[\text{Fe}_4\text{S}_4]$ and $[2\text{Fe}]_{\text{H}}$ clusters, a requisite that further complicates the analysis of this large system. Therefore, it is not surprising that only one theoretical work on the entire H-cluster in combination with the BS approach has been reported so far [147]. This study was focused on the active form of the H-cluster $[(\text{CH}_3\text{S})_4(\text{Fe}_4\text{S}_4)(\mu\text{-PDT})(\text{Fe}_2)(\text{CO})_3(\text{CN})_2]^{3-}$ (**3a**) derived from the X-ray structure of *Clostridium pasteurianum* (CpI) in which a CO ligand is bridged between the Fe_p and Fe_d atoms [48] (hereafter the labels will indicate the total charge of the complex and the isomer form of the $[2\text{Fe}]_{\text{H}}$ moiety; **a** for the isomer with a bridged CO ligand and **b** for the isomer with all terminal CO ligands. The two isomers will be also referred as bridged and open, respectively). In **3a** the $[\text{Fe}_4\text{S}_4]$ cluster was found to feature the formal $\text{Fe(II)Fe(II)Fe(III)Fe(III)}$ redox state and the $[2\text{Fe}]_{\text{H}}$ cluster the formal Fe(I) – Fe(II) redox state. Brunold et al. [147] also investigated the molecular and electronic properties of the CO-inhibited form of the H cluster $[(\text{CH}_3\text{S})_4(\text{Fe}_4\text{S}_4)(\mu\text{-PDT})(\text{Fe}_2)(\text{CO})_4(\text{CN})_2]^{3-}$ (**3a-CO**) and of the transient state generated by irradiation of the CO-inhibited form at cryogenic temperatures (**3b**). This unstable state results from the photodissociation of the bridging CO, leading to an isomer of **3a** featuring all terminal COs.

As discussed above, the exchange interactions in this paramagnetic active form of the H cluster can be treated with a three spin model Hamiltonian, which may be written in the general form as:

$$H = J_{12} (\mathbf{S}_1 \cdot \mathbf{S}_2) + J_{13} (\mathbf{S}_1 \cdot \mathbf{S}_3) + J_{23} (\mathbf{S}_2 \cdot \mathbf{S}_3) . \quad (13)$$

As suggested by Brunold et al. [147] the interaction between the non-adjacent spin centers 1 and 3 is negligible. This simplify the spin model Hamiltonian to

the form:

$$\mathbf{H} = J_{\text{cube}} (\mathbf{S}_A \cdot \mathbf{S}_B) + J_H (\mathbf{S}_A \cdot \mathbf{S}_H) , \quad (14)$$

where J_{cube} represents the exchange interaction between the two Fe dimers of the diamagnetic $[\text{Fe}_4\text{S}_4]$ cluster with $S_A = S_B = 9/2$ and J_H corresponds to the exchange interaction between the vicinal mixed-valence Fe dimer (A) and the $[2\text{Fe}]_H$ cluster with $S = 1/2$ (Fig. 7). In this model, three different spin configurations of the $[\text{Fe}_4\text{S}_4]$ cluster have to be considered, depending on which Fe atoms are coupled to form the two mixed-valence dimers [147]. In addition, for each of these configurations the vicinal mixed-valence Fe dimer (A) can be coupled to the unpaired electron of the $[2\text{Fe}]_H$ cluster in a ferromagnetic (F) or antiferromagnetic (AF) way. This leads to two different BS solutions: one with the unpaired electron of the $[2\text{Fe}]_H$ aligned with those of the vicinal mixed-valence Fe dimer (F), and one with opposite spins (AF). Thus, a total of six different BS configurations have been calculated and the resulting value for J_H has been obtained from (Eq. 6):

$$E(\text{F}; M_{\text{BS}} = 5) - E(\text{AF}; M_{\text{BS}} = 4) = 2J_H (9/2) (1/2) = (9/2) J_H . \quad (15)$$

Because of the three different spin configurations of the $[\text{Fe}_4\text{S}_4]$ cluster, three distinct J_H values can be obtained. A fourth value may be calculated by the ferromagnetic and antiferromagnetic coupling of the unpaired electron of the $[2\text{Fe}]_H$ cluster with the high spin state of the $[\text{Fe}_4\text{S}_4]$ cubane according to:

$$E(\text{F}; M_{\text{BS}} = 19/2) - E(\text{AF}; M_{\text{BS}} = 17/2) = 2J_H (9) (1/2) = (9) J_H . \quad (16)$$

The average value of J_H for the complex **3a** is equal to $+15 \pm 10 \text{ cm}^{-1}$, in very good agreement with the experimental value of 20 cm^{-1} [52, 53]. Notably, the calculated value of J_H in the CO-inhibited form of the H-cluster **3a-CO** ($150 \pm 50 \text{ cm}^{-1}$) is significantly larger than the value calculated for **3a**, in satisfactory agreement with the experimental value of about 100 cm^{-1} . The agreement between exchange parameters derived from theory and experiment shows that the BS approach within the DFT method is a suitable procedure for complex magnetic systems like those examined here. The computed molecular and electronic structure of the elusive species **3b** is consistent with that suggested on the basis of the few structural and spectroscopic experimental data available. The structure of the $[2\text{Fe}]_H$ cluster is, in fact, characterized by all terminal CO ligands, with the unpaired electron remaining on this moiety, and the $[\text{Fe}_4\text{S}_4]$ maintaining the formal $+2$ oxidation state (Fig. 7b).

The results so far discussed assign to the $[\text{Fe}_4\text{S}_4]$ cluster a marked influence on the electronic and molecular properties of the $[2\text{Fe}]_H$ cluster. In particular, the unpaired electron is almost fully localized on the Fe_d atom in **3a**, while the addition of CO to the Fe_d atom significantly alters the spin density within the $[2\text{Fe}]_H$ cluster, leading to a more extended spin delocalization between the two Fe centers. This electronic structure description accounts for the different values of the exchange parameter J_H . In fact, the

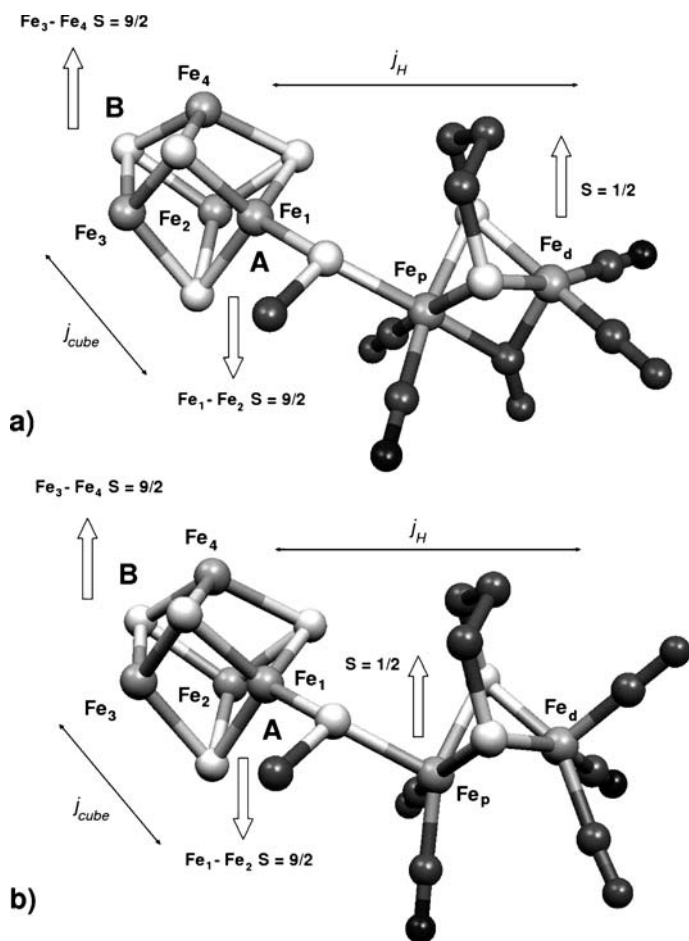


Fig. 7 Schematic diagram of the three spin model for the bridged (a) and open (b) forms of H-cluster. It should be noted that the unpaired electron in the $[2Fe]_H$ moiety is localized on the Fe_d and Fe_p atom for the bridged and open form, respectively

large amount of spin density on the Fe_p atom in **3a-CO** should facilitate the exchange interaction with the mixed-valence Fe dimer in the $[Fe_4S_4]$ cluster inducing more paramagnetism into the latter, in agreement with Mössbauer investigations [52, 53]. Notably, in the case of **3b** the unpaired electron is almost entirely localized on the Fe_p atom indicating that transition between the bridged (**3a**) and open (**3b**) isomers provides a transfer of the spin from the Fe_d to the Fe_p atom.

Recently, Bruschi et al. have undertaken the study of the H-cluster as a function of environment and redox state [148], investigating the -5 , -4 and -3 redox states and considering both open and bridged isomers of the $[2Fe]_H$ cluster. In order to explore the effects of the environment on the

electronic and molecular properties of the H-cluster model, geometry optimization and electronic structure calculations have also been carried out on models soaked in the COSMO polarizable continuum medium [149] characterized by dielectric constants ϵ equal to 4 and 40 (from hereafter in the labels of complexes a subscript is used to indicate the ϵ value of the environment), which are values commonly used to describe protein environment [150–152]. Geometry optimisations have been carried out with the TURBOMOLE suite of programs [153] in connection with the resolution of the identity technique (RI) [17, 18], using the BP86 functional [23, 24] and an all-electron valence triple- ζ basis set with polarization functions on all atoms [154]. The starting structure was taken from the X-ray geometry of the H-cluster of *CpI* [Fe] hydrogenase [48]. Some relevant computed structural parameters are reported in Table 1.

In the more reduced complex $[(\text{CH}_3\text{S})_4(\text{Fe}_4\text{S}_4)(\mu\text{-PDT})(\text{Fe}_2)(\text{CO})_3(\text{CN})_2]^{-5}$ **5b**, the unpaired electron is localized on the $[\text{Fe}_4\text{S}_4]$ cluster (Table 2) suggesting the formal redox state $\text{Fe(III)Fe(II)Fe(II)Fe(II)}$ and correspondingly, the Fe(I)Fe(I) redox state for the $[\text{2Fe}]_{\text{H}}$ cluster. In the open isomer the $\text{Fe}_d(\text{CO})_2\text{CN}$ group is partially rotated, with a CO ligand oriented toward the Fe_p atom. The $\text{Fe} - \text{S} - \text{Fe}_p$ angle is equal to 159 degrees, a value significantly larger than that observed in the X-ray diffraction structure (Table 1). This could be an indication that the environment has a marked influence on **5a**; in particular, the $\text{Fe} - \text{S} - \text{Fe}_p$ angle is drastically reduced increasing the dielectric constant of the environment to 4. In **5b**₄₀ the effect of the environment is even more evident and the stable form corresponds to a struc-

Table 1 Relevant geometrical parameters of complexes **5a–3a** and **5b–3b** computed at BP86/def-TZVP level of theory. The three values from the left refer to the parameters calculated in gas-phase, and in the polarizable continuum medium with $\epsilon = 4$ and 40, respectively

	$\text{Fe} - \text{S} - \text{Fe}_p^a$			$\text{Fe}_p - \text{Fe}_d^b$		
5a	—,	142.1,	125.0	—,	2.601,	2.571
5b	159.2,	142.7,	—	2.732,	2.596,	—
4a	154.4,	141.3,	130.4	2.621,	2.599,	2.580
4b	153.4,	140.7,	129.1	2.662,	2.593,	2.562
3a	134.3,	126.2,	121.7	2.556,	2.551,	2.542
3b	146.8,	133.4,	121.7	2.614,	2.588,	2.569
Expt.	118.0 ^c , 116.4 ^d , 119.5 ^e			2.617 ^c , 3.555 ^d , 2.601 ^e		

^a in degrees

^b in Å

^c taken from [48]

^d taken from [43]

^e taken from [58]

Table 2 Atomic spin densities of the Fe atoms in the H-cluster of the [Fe] hydrogenase computed at BP86/def-TZVP level of theory. Fe_{1,2} and Fe_{3,4} refer to the Fe atoms in the [Fe₄S₄] cluster of the two layers coupled antiferromagnetically; Fe_p and Fe_d indicate the proximal and distal Fe atoms of the [2Fe]_H cluster, respectively. The *asterisks* specify the atomic spin density of the Fe center coordinated to the bridging S atom

	5b ₁	5a ₁	4b ₁	4a ₁	3b ₁	3a ₁
Fe _{1,2}	3.24*, 3.22	3.24*, 3.21	3.04*, 3.07	3.04, 3.06	3.07*, 3.07	3.10, 3.07
Fe _{3,4}	-2.93, -2.89	-2.89, -2.91	-3.21, -3.15	-3.14, -3.13	-3.09, -3.06	-3.08, -3.07
Fe _p	0.03	0.01	0.21	0.04	1.21	0.17
Fe _d	0.01	0.03	0.01	0.1	-0.09	0.88

ture characterized by a μ -CO group. The gas-phase geometry optimization of the [(CH₃S)₄(Fe₄S₄)(PDT)(Fe₂)(μ -CO)(CO)₂(CN)₂]⁻⁵ isomer **5a**₁, reveals that this species is unstable. However, the same complex is stable when the optimization is carried out in presence of a polarizable continuum medium **5a**₄ and **5a**₄₀. In this case the electronic structure of the [Fe₄S₄] cluster is very similar to that calculated for **5b**₄, being still compatible with a formal Fe(III)Fe(II)Fe(II)Fe(II) redox state and with the Fe(I)Fe(I) redox state of the [2Fe]_H cluster.

After mono-electron oxidation of the fully reduced species to give the - 4 redox state of the H-cluster, both the open and bridged isomers have been characterized in gas-phase, as well as in the polarizable continuum medium. In both isomers the electron is removed from the [Fe₄S₄] cluster. This leads to an electronic structure which is compatible with the Fe(II)Fe(II)Fe(III)Fe(III) redox state of the [Fe₄S₄] cluster and with the Fe(I)Fe(I) redox state of the [2Fe]_H cluster. However, in the gas-phase a significant spin density is localized on the Fe_p and Fe_d atoms in the open and bridged forms, respectively, indicating electron delocalization between the [Fe₄S₄] and the [2Fe]_H clusters. This delocalization is reduced when the H-cluster is soaked in the polarizable continuum medium. The environment significantly affects also the geometry, as observed in the case of the - 5 redox state. In particular the Fe - S - Fe_p angle is reduced from about 150 degrees to about 130 degrees, when moving from the gas-phase to the $\epsilon = 40$. The molecular geometries of the [2Fe]_H cluster in the open and bridged isomers are very similar to those calculated for the - 5 redox state, indicating that the oxidation state of the [Fe₄S₄] cluster has only a minor effect on the geometry of the [2Fe]_H cluster.

A further mono-electron oxidation of the [(CH₃S)₄(Fe₄S₄)(μ -PDT)(Fe₂)(CO)₃(CN)₂]⁻⁴ complex leads to the - 3 redox state, which corresponds to the form investigated by Brunold and collaborators. In this case, the influence of the environment is less significant than in the case of **5** and **4** complexes, as expected due to the reduction of the total charge. However, also in this case

the $\text{Fe}-\text{S}-\text{Fe}_\text{p}$ angle decreases from 146.8 to 121.7 degrees when moving from the gas-phase to the polarizable continuum medium with $\epsilon = 40$.

Several interesting observation can be pointed out in the light of the presented results. The large negative charge in the gas phase models leads to a structure of the $[\text{2Fe}]_\text{H}$ subcluster significantly different from that observed experimentally in the protein. In this regard, the better agreement of the molecular structures of the solvated complexes may suggest that the H-cluster in the protein environment is characterized by a lower effective charge. This observation is intriguing since a detailed analysis of the protein structure shows that there are only a few charged residues near the H-cluster, which are, however, not sufficient to neutralize its negative charge. Furthermore, this investigation shows that the protein environment can influence the electronic and molecular characteristic of the binuclear cluster. In particular, the modulation of the dielectric constant of the protein environment affects the stability of the bridged and open isomers, providing a possible mechanism of interconversion between the two forms.

5

Modeling Photochemical Reaction Paths

5.1

Introduction to Time-Dependent Density Functional Theory (TDDFT)

TDDFT has become the reference method for studying the excited state of medium and large size molecular systems [155]. This approach has the same advantages and limitations that DFT has in comparison with RHF and correlated methods for the ground state. TDDFT includes electronic correlation effects through i) the approximated exchange-correlation functionals and ii) the contribution of double excited configurations, which are not accounted in the Configuration-Interaction Singles (CIS) method. Nevertheless, TDDFT is a single-reference method, and therefore can not achieve high accuracy as high-level correlated methods like multi-reference configuration-interaction or equation-of-motion coupled-cluster. Finally, TDDFT computational cost is of the same magnitude as ground state DFT, and when RI approximation is applied with non-hybrid functionals, one can save extra CPU time up to 10–100 times. Appendix II describes shortly the basic theory of TDDFT.

Regarding the vertical excitation energies, as far as low-lying excited states with negligible double-excitation character are considered, the error that affects TDDFT results is around 0.1 to 0.5 eV, remembering that the chemical accuracy is 0.05 eV (around 1 Kcal/mol). TDDFT results can be very poor for excited state of molecular system with extended π conjugated bonds [156], as well as for Rydberg and charge-transfer (CT) excited states [155]. These facts are mainly related to an unsatisfactory description of virtual orbitals because

of the non-correct long-range behavior of the exchange-correlation potentials [157]. The effects of the type of exchange-correlation functional and the basis set are also a crucial issue in the TDDFT level of theory. As commented above, the accuracy of TDDFT depends strongly on the type of excited state considered.

Since we are dealing with biomimetic models of metalloenzymes, it is important to discuss the accuracy of TDDFT for transition metal complexes. Like ground state DFT, also for TDDFT such systems pose more problems than organic molecules, because of the multireference character of the wavefunction of the former. Gisbergen et al. [158] present a small benchmark of excitation energies of transition metal (TM) compounds, from which it results a TDDFT non-uniform accuracy with respect to the high-level methods. As far as the type of functional is concerned, Hummel et al. [159] claim that B3LYP plus the inclusion of solvation effects by means of a continuum model yield accurate excitation energies for $M(CO)_n$ carbonyl complexes. A recent benchmark by Petit et al. [160] clearly shows that excitation energies do not depend on the basis set and pseudopotential quality, and that the accuracy of the PBE0 hybrid functional [161] is close to that of high-level methods.

Excited state energy gradients are crucial properties in studying photochemical reaction path, since they allow one to explore excited state potential energy surfaces, going beyond applications related to the evaluation of the vertical excitations. Recent laser spectroscopy developments have increased our knowledge of the excited state properties (normal modes, dipole moment, etc). With ultra-fast spectroscopy, it is possible in some simple cases to follow whole photochemical reaction paths [162].

The development of the TDDFT gradient of the excited state energy is recent. In 1999 Van Caillie and Amos used the so-called Z-vector method previously developed by Handy and Schaefer [163], to implement LDA [164] and GGA [165] excited state energy first derivatives (CADPAC). A few years later, Furche and Ahlrichs [166] implemented the gradients in the TURBOMOLE program suite [153], using the Lagrangian method [167]. Finally, within a plane-wave basis set framework, Hutter [168] implemented the TDA gradient using the Lagrangian methods, while Doltsinis and Kosov [169] used an implicit differentiation scheme.

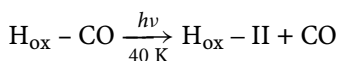
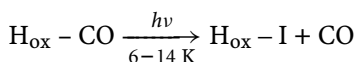
The reliability and effectiveness of excited state geometry optimization depends on many factors, as the size and the symmetry of the investigated system and the topology of the energy surfaces (excited state intersections, etc). In general, the computational cost of this optimization is similar to that in the ground state, but unfavorable situations like fluctuations of the excited state energies and gradients or singlet-triplet instability [170] can occur. The quality of the TDDFT excited state optimized geometry is promising. In the case of small systems [166, 171] adiabatic excitation energies and bond distances are in good agreement with experimental or high-level calculated values.

5.2

TDDFT Modeling of CO Photolysis in the Carbon Monoxide Inhibited Form of [Fe] Hydrogenase

The theoretical investigations of photochemical processes involving transition metals at TDDFT level are mainly devoted to the study of the vertical excitations, with the assignment of the excitation levels, the simulation of the electronic spectra, etc. The exploration of the excited state potential energy surfaces (characterization of the stationary points as local/global minima and first-order saddle points) is still lacking, since TDDFT gradients have been only recently developed and implemented in some quantum chemistry suite programs. Moreover, excited states wave functions of coordination compounds often possesses multireference character, which is unavoidably poorly described by a single reference methods as TDDFT. Nevertheless, in the recent literature there are some examples of TDDFT study of photochemical processes in which bio-inorganic systems are involved. For example, in the case of the CO photodissociation of CO-ligated hemoglobin [172], the authors have characterized some potential energy surfaces by single-point excitation energy computations along the Fe – CO bond length reaction coordinate. As the authors pointed out, it is reasonable to expect that this picture would be close to the one obtained from the true minimum-energy path of hemoglobin. Another interesting example is the study of the photochemical activation of N₂ by dinuclear Ru and Fe complexes [173].

[Fe] hydrogenase gives an intriguing example of a photochemical reaction path which can be investigated by means of TDDFT. When the oxidized form of the enzyme is exposed to CO, the [2Fe]_H cluster binds a further CO molecule which substitutes the H₂O molecule on the distal Fe atom [174–177]. In this form the enzyme (H_{ox} – CO hereafter) is inactive, but the active form (H_{ox} hereafter) can be restored by illumination with white light at cryogenic temperature. This process gives two different light-induced species as a function of temperature:



IR spectra [178] have highlighted that H_{ox} – I is the result of the photolysis of the exogenous CO, while H_{ox} – II is characterized by the loss of the CO group bridging the iron centres.

The theoretical modeling of this photolysis process is still lacking, but some interesting results have been recently reported in two papers by Fiedler et al. [147, 179]. In the first paper the valence excited states of two biomimetic model of [2Fe]_H, namely [Fe₂(S₂C₃H₆)(CO)₆] and [Fe₂(S₂C₃H₆)(CO)₄(CN)₂]²⁻ have been studied. Beside electronic and res-

onance Raman spectra simulations, the authors estimate the elongation of Fe – Fe bond length in the first and third excited state of $[\text{Fe}_2(\text{S}_2\text{C}_3\text{H}_6)(\text{CO})_6]$. The second paper is devoted to the calculation of different properties on a model of the full H-cluster in which the $[\text{Fe}_4\text{S}_4]$ cluster is explicitly considered. In this paper two models for $\text{H}_{\text{ox}}\text{-I}$ and $\text{H}_{\text{ox}}\text{-II}$ based on the comparison between calculated and experimental [174, 177, 180] EPR parameters are also proposed, corroborating the hypothesis made by Chen et al. [178].

The following section focuses on the possibilities supplied by TDDFT exploration of the excited state potential energy surfaces in studying photochemical reaction paths, using CO photolysis of $\text{H}_{\text{ox}}\text{-CO}$ as an example. The attempt is to show which kind of information can be obtained from a TDDFT study on a model system, using geometry optimization of the excited state geometry structure. Computations are performed using TURBOMOLE with the BP86 functional and TZVP basis set. RI approximation is used with auxiliary basis set of the same TZ quality [181]. Excited state structures are optimized until the excited state energy gradient norm reaches values at lower than $0.01 \text{ hartree } \text{\AA}^{-1}$.

The modeling of the CO photolysis mechanism is carried out using the CO inhibited $[\text{2Fe}]_{\text{H}}$ cluster biomimetic system $[\text{Fe}_2(\mu\text{-PDT})(\mu\text{-CO})(\text{CO})_6]$ (hereafter **a-CO**). This binuclear complex has two $\text{Fe}^{\text{I}}\text{Fe}^{\text{I}}$ atoms and terminal CO groups can be distinguished between the two CO trans to the $\mu\text{-CO}$ (one *anti* and one *syn* with respect to the β carbon of CH_2 group of the $\mu\text{-PDT}$) and the four CO trans to the $\mu\text{-PDT}$. When **a-CO** loses a CO molecule one obtains $[\text{Fe}_2(\mu\text{-PDT})(\text{CO})_6]$ (hereafter **a**). **a-CO** and **a** have $\text{S}_0 \text{ } ^1\text{A}'$ ground state with global C_s minimum geometry (Fig. 8).

The analysis of the optimized geometries for **a-CO** and **a** reveals some important differences between the two structures. The Fe-Fe internuclear distance increases by $0.448 \text{ \AA}$ going from **a-CO** to **a**. Fe – C distances relative to the two C atoms of *syn* and *anti* CO increase by $0.067 \text{ \AA}$ and $0.062 \text{ \AA}$ re-

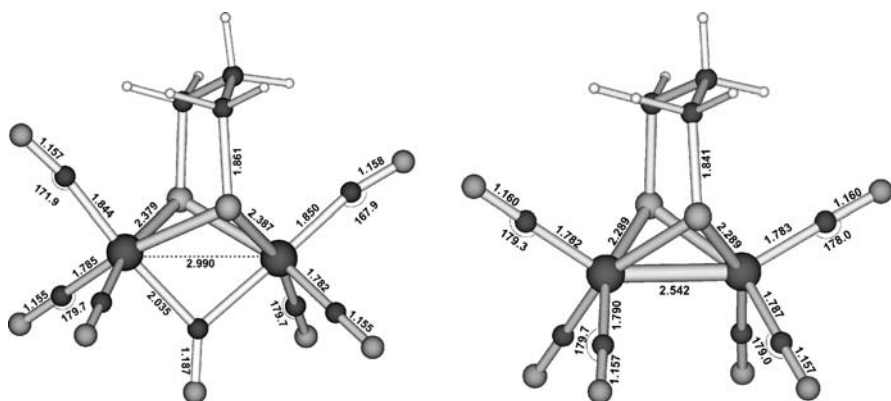
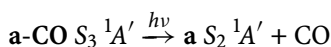
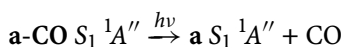


Fig. 8 Structure of **a** and **a-CO** ground states. All distances and angles in $\text{\AA}$ and degree

spectively, while the Fe – C – O bond angles decrease by 10.1 and 7.4 degrees. These effects can be rationalized in light of the analysis of the highest occupied molecular orbital (HOMO) of the two model complexes [179]. The first bond length increase can be accounted for by observing that in **a**-CO a *d*-orbitals Fe – Fe antibonding combination is found, which becomes bonding in **a**. Regarding the Fe – C bond length increase, the HOMO analysis does not give a clear explanation. In both **a**-CO and **a** the HOMO has a *d*-Fe-*p*-C orbital combination with antibonding character, which is stronger for **a**-CO, according to the overlap population. This could be one of the reasons of the observed structural changing.

In the investigation of the valence excited states, we consider the two single **a**-CO excited states $S_1\ ^1A''$ and $S_3\ ^1A'$, along with the corresponding states of **a** $S_1\ ^1A''$ and $S_2\ ^1A'$. The **a** excited states have the same electronic state of the **a**-CO ones, and the processes



can be considered as two of the possible paths of the hypothetical CO photolysis of **a**-CO. In Table 3 the main features of the two excited states under investigation for both systems are reported. The $S_1\ ^1A''$ and $S_3\ ^1A'$ **d**-CO excited states are characterized by the one-electron excitation HOMO $\rightarrow$ LUMO ($37a''$) and HOMO $\rightarrow$ LUMO+1 ($68a'$), respectively. These virtual MOs possess antibonding combinations with respect to both Fe – C bonds of two CO trans to the μ -CO. In the case of the LUMO + 1 the combination in which

Table 3 Computed energies for **a**-CO and **a** (in hartree). E_{vertical} is the energy of the excited state at the geometry of the optimized ground state; for the S_0 ground state E_{vertical} is the energy at the optimized minimum geometry; the leading one-electron excitation refers to the single-excitation contribution $a \rightarrow i$; E_{opt} is the excited state energy at its own optimized geometry; Gradient corresponds to the norm of the excited state energy gradient in hartree $\text{\AA}^{-1}$

d -CO state	E_{vertical} (hartree)	leading one-electron excitations	E_{opt} (hartree)	Gradient
$S_0\ ^1A'$	-4236.3114			
$S_1\ ^1A''$	-4236.2521	67 $a' \rightarrow 37a''$	-4236.2716	0.010
$S_3\ ^1A'$	-4236.2280	67 $a' \rightarrow 68a'$	-4236.2425	0.016
d state				
$S_0\ ^1A'$	-4122.9559			
$S_1\ ^1A''$	-4122.8510	35 $a'' \rightarrow 62a'$	-4122.8737	0.0010
$S_2\ ^1A'$	-4122.8501	60 $a', 61a' \rightarrow 62a'$	-4122.8704	0.0011

the *anti* CO in trans has an higher weight, while the opposite happens for the LUMO.

The optimized **a**-CO excited states structures are reported in Fig. 9. The final norm of the gradient is different from zero in a non-negligible way (Table 3) and fluctuation of energies and geometries are observed during optimization. Furthermore, we found for both optimized structures a critically low value (around 0.01) of the lowest singlet instability matrix eigenvalue. However, it is reasonable to consider these structures as two points along the CO photolysis reaction path that belong to S_1 and S_3 potential energy surfaces. Within this assumption, we can characterize the reaction paths through the analysis of the optimized structure.

In both **a**-CO optimized excited state structures, the μ -CO loses its bridging character, becoming a terminal CO. The labilized Fe–C bond for S_1 is the one in trans position to the *anti* terminal CO, while for S_3 the one in trans

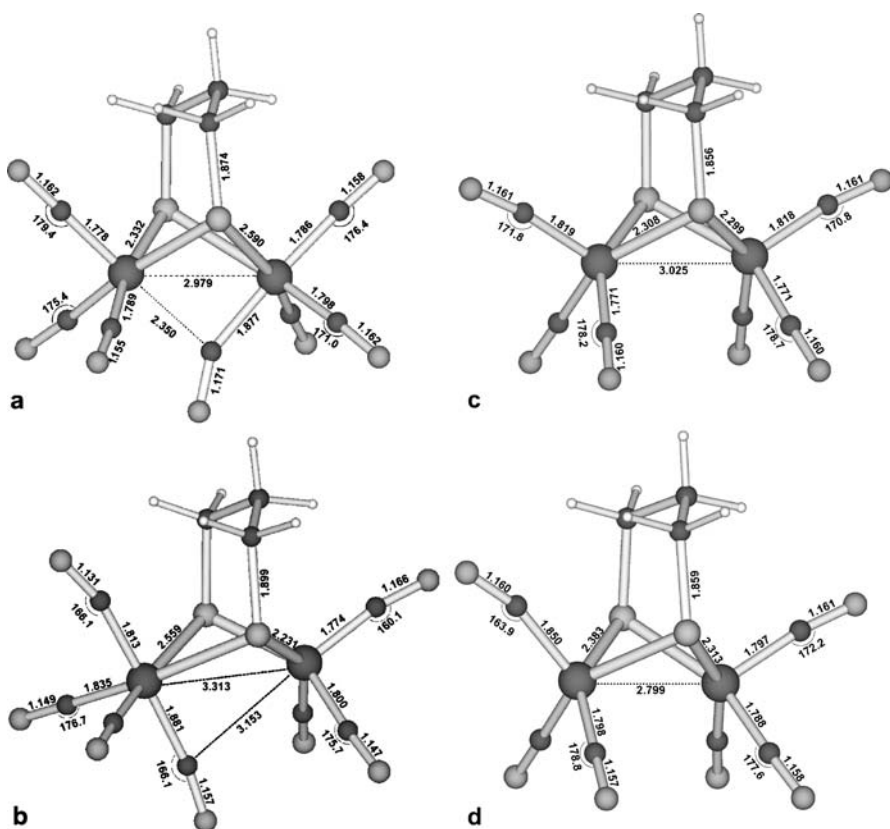


Fig. 9 Structure of **a**-CO and **a** excited states. All distances and angles in Å and degree. **a,b** **a**-CO S_1 $^1A''$ and S_3 $^1A'$ excited state structures, respectively; **c,d** **a** S_1 $^1A''$ and S_2 $^1A'$ excited state structures, respectively

position to the *syn* terminal CO. Therefore we have characterized two channels in S_1 and S_3 potential energy surfaces along which the photolysis brings to the loss of the μ -CO group, although it is not possible to surely assess which is exactly the leaving CO group. Nevertheless for S_1 and S_3 the leaving CO group is one of those bound to the Fe *syn* or *anti* to the β carbon of μ -PDT, respectively.

The excited state optimized structures for **a** are reported in Fig. 9. The gradient norms of optimized excited state energies are around $0.001 \text{ hartree } \text{\AA}^{-1}$, and therefore these are likely stationary structures. The Fe – Fe internuclear distances are remarkably increased with respect to the ground state, because in both cases molecular orbitals with Fe – Fe antibonding combinations are populated.

The final considerations we can make in light of the results obtained are that **a**-CO S_1 and S_3 are involved in the photolysis of μ -CO group. Since **a**-CO is a model for the CO inhibited H-cluster, we can argue that similar excited states might be involved in the formation of H_{ox} -II, in the approximation of considering the effect of the $[\text{Fe}_4\text{S}_4]$ cluster negligible.

6

Conclusions and Perspectives

The application of quantum chemical methods to the investigation of hydrogenases and related biomimetic models has nicely complemented experimental studies, allowing us to disclose key factors related to the catalytic properties of the enzymes. In particular, the possibility to use theoretical methods to study intermediate and transition state structures, reaction energies and activation barriers, as well as spectroscopic properties, has allowed us to dissect the molecular mechanism of several reaction paths relevant to hydrogenase chemistry. Nevertheless, several relevant issues remain unsolved. For example, further experimental and theoretical work is necessary to corroborate the formation of high-spin species in the catalytic cycle of $[\text{NiFe}]$ hydrogenases. This will also need further studies to evaluate and improve the reliability of DFT functionals in properly describing relative stabilities of low and high spin species. As for $[\text{Fe}]$ hydrogenases, the investigation of the role of the $[\text{Fe}_4\text{S}_4]$ moiety on the catalytic properties of the $[\text{2Fe}]_H$ cluster has started to be addressed only very recently and is still largely incomplete.

The examples discussed in this review also show that both the investigation of small and large models is necessary to obtain a clear picture of the chemistry of complex molecular systems such as hydrogenases. In general, the investigation of simple models, which are computationally inexpensive, allows one to use sophisticated quantum chemical approaches and to compare properties of several models. However, the investigation of more complex

models is mandatory when the effect of neighboring groups has to be taken into account. Indeed, the adopted computational models are often a compromise between a realistic representation of the enzyme active site and the necessity to use small models to maintain computational costs to a reasonable limit. In fact, even though calculations carried out on model systems including also a few aminoacids not directly bound to the metal clusters have been reported [112, 113], most theoretical investigations of hydrogenases active site have been carried out using relatively small models in which the cysteine residues coordinated to the metal ions have been modeled by H or CH₃ groups.

The extension of small models to include effects due to steric constraints and electrostatic effects imposed by the protein environment is expected to be crucial to clarify key issues in hydrogenase chemistry. As an example, aminoacids in the nearby of the metallic clusters of hydrogenases, such as cysteine, aspartate, glutamate and histidine might be directly involved in H₂ activation. To the best of our knowledge, only one theoretical study related to hydrogenases and keeping explicitly into account the protein environment of the active site has been reported so far [182]. Amara et al. employed a hybrid QM/MM potential method for the investigation of some [NiFe] hydrogenase redox states, reaching the conclusion that one of the cysteine residues is involved in the heterolytic cleavage of dihydrogen and that μ -H species can be formed in the catalytic cycle. However, as emphasized by the authors, owing to the expense of calculations an exhaustive investigation of all plausible structures and redox states could not be carried out, underlining some of the problems that are still associated with the proper modeling of large protein portions.

Appendix

A

Outline of Density Functional Theory

In density-functional theory [13] (DFT) methods, the problem of finding the electronic structure of an atomic or molecular system is solved on the basis of its three-dimensional electron density $\rho(\mathbf{r})$ as a basic variable. This approach represents a simplification with respect to common methods of quantum chemistry based on the $3N$ -dimensional (or $4N$ when spin variables are included) wave function, where N is the number of the electron involved. The Hohenberg–Kohn theorem [12] provides a rigorous framework for this approach. The first theorem assesses that a one-to-one map between the ground state wave function and the electron density exists through the external po-

tential

$$v(\mathbf{r}) = - \sum_{\alpha} \frac{Z_{\alpha}}{|\mathbf{r} - \mathbf{R}_{\alpha}|}, \quad (17)$$

where α runs over the number of nuclei of charges Z_{α} with coordinates $\mathbf{R}_{\alpha}$. The total energy is written as a functional of the electron density $\rho(\mathbf{r})$

$$E[\rho(\mathbf{r})] = F[\rho(\mathbf{r})] + \int v(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r}, \quad (18)$$

where $F[\rho(\mathbf{r})]$ is the unknown functional of the density. Since $F[\rho(\mathbf{r})]$ does not depend on the $v(\mathbf{r})$, it must be an universal functional of the electron density. The second theorem defines a variational principle for the electron density: for a given trial density $\rho_t(\mathbf{r})$ we have that $E_0 \leq E[\rho_t(\mathbf{r})]$. The energy functional is minimized according to the variational principle with the constraint of the N -representability $\int \rho(\mathbf{r}) d\mathbf{r} = N$. In the language of the functional derivatives this means that the differential $\delta E[\rho(\mathbf{r})] = 0$

$$\frac{\delta}{\delta \rho(\mathbf{r})} \left\{ E[\rho(\mathbf{r})] - \mu \left[\int \rho(\mathbf{r}) d\mathbf{r} - N \right] \right\} = 0, \quad (19)$$

where μ is the undetermined Lagrange multiplier. From the definition of the differential of a functionals, it turns out that

$$\mu = \frac{\delta F[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} + v(\mathbf{r}). \quad (20)$$

All current DFT implementations in quantum chemistry are based on the Kohn–Sham (KS) approach which introduces in an ingenious way the molecular orbitals machinery to overcome the problems of the early Thomas-Fermi model in computing the kinetic energy [183]. The basic idea is the introduction of the so-called non-interacting reference system, obtained by turning off the electron-electron repulsion in the real interacting system as follow. Let us define the parameter dependent Hamiltonian operator $\hat{H}_{\lambda}$ as

$$\hat{H}_{\lambda} = \hat{T} + \hat{V}_{\lambda} + \lambda \hat{V}_{ee}, \quad (21)$$

where $0 \leq \lambda \leq 1$ is called interelectronic coupling-strength parameter. For $\lambda = 1$ we have the real interacting system, and $\hat{V}_{\lambda=1}$ is the usual electron-nucleus potential such that $\hat{V} = \sum_i^N v(\mathbf{r}_i)$. The $\lambda = 0$ case corresponds to the non-interacting system in which $\hat{V}_{\lambda=0}$ is such that the electron density $\rho(\mathbf{r})_{\lambda=0} = \rho(\mathbf{r})_{\lambda=1}$.

The exact ground state kinetic energy $T[\rho(\mathbf{r})]$ of the real interacting system ($\lambda = 1$) can be expressed as a natural orbitals (NO) expansion, but it can be approximated with the exact ground state kinetic energy $T_S[\rho(\mathbf{r})]$ of the

non-interacting system:

$$T[\rho(\mathbf{r})] = \sum_{i=1}^{\infty} n_i \left\langle \phi_i^{\text{NO}} \left| -\frac{\nabla^2}{2} \right| \phi_i^{\text{NO}} \right\rangle \approx \sum_{i=1}^N \left\langle \phi_i \left| -\frac{\nabla^2}{2} \right| \phi_i \right\rangle = T_S[\rho(\mathbf{r})], \quad (22)$$

where the natural orbitals ϕ_i^{NO} are eigenvectors of the one-electron density matrix, with the orbital occupation numbers $0 \leq n_i \leq 1$ and with the total number of electron $N = \sum_i^{\infty} n_i$. The exact electron density $\rho(\mathbf{r})$ is then approximate as follows:

$$\rho(\mathbf{r}) = \sum_{i=1}^{\infty} n_i |\phi_i^{\text{NO}}|^2 \approx \sum_{i=1}^N |\phi_i|^2, \quad (23)$$

where ϕ_i are the KS molecular orbitals. The exchange-correlation energy functional is defined as the difference between the kinetic and electron-electron interaction energy of the real and reference systems:

$$E_{\text{xc}}[\rho(\mathbf{r})] = (T[\rho(\mathbf{r})] - T_S[\rho(\mathbf{r})]) + (V_{\text{ee}}[\rho(\mathbf{r})] - J[\rho(\mathbf{r})]), \quad (24)$$

where $V_{\text{ee}}[\rho(\mathbf{r})]$ is the electron-electron interaction energy whose classical part $J[\rho(\mathbf{r})]$ is given by

$$J[\rho(\mathbf{r})] = \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (25)$$

The E_{xc} contains the non-classical contributions to the kinetic and electron-electron interaction energies, but it is mainly dominated by the exchange contribution, since $T_S[\rho(\mathbf{r})]$ is the dominant part of the true kinetic energy. Finally the $v_{\text{KS}}(\mathbf{r})$ is

$$v_{\text{KS}}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}), \quad (26)$$

where $v_{\text{xc}}(\mathbf{r})$ is the exchange-correlation potential given by the functional derivatives of E_{xc} with respect to $\rho(\mathbf{r})$. The problem of finding the KS molecular orbitals ϕ_i is then solved using the self-consistent-field (SCF) machinery: the $\phi_i(\mathbf{r})$ orbitals are the lowest energy eigenfunctions of the KS one-electron operator $\hat{h}_{\text{KS}}$

$$\hat{h}_{\text{KS}}\phi_i(\mathbf{r}) = \left[-\frac{\nabla^2}{2} + v_{\text{KS}}(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \quad (27)$$

where ϵ_i are the KS orbital energies. The KS total energy is

$$E_{\text{KS}} = \sum_{i=1}^N \left\langle \phi_i \left| -\frac{\nabla^2}{2} \right| \phi_i \right\rangle + \int v(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r} + J[\rho(\mathbf{r})] + E_{\text{xc}}[\rho(\mathbf{r})]. \quad (28)$$

From a computational point of view, the ϕ_i KS MOs are conveniently expanded in a fixed basis set of functions (usually called “atomic” orbitals) φ such that $\phi_i = \sum_p c_{pi} \varphi_p$. In turn, φ_p can be represented by a given number of

linearly combined functions (gaussian). The number of φ_p basis functions considered determines the variational flexibility of the basis set and therefore the quality of the results. For instance, for atoms having occupied orbitals $1s$, $2s$ and $2p$, a basis set composed by 4φ of s type and 2φ of p type is called “double zeta” basis. Additional functions can be added to represents diffuse states (negatively charged) or polarization effects (high angular momentum functions).

Expanding the KS molecular orbitals in the atomic basis sets φ (LCAO approach) the total KS electronic energy becomes [184]

$$E_{\text{KS}} = \sum_{pq}^{N_{\text{basis}}} P_{pq} h_{pq} + \frac{1}{2} \sum_{pqrs}^{N_{\text{basis}}} P_{pq} P_{rs} (pq|rs) + E_{\text{xc}}(P), \quad (29)$$

where the indices p, q, r and s run over the number of atomic basis functions. P_{pq} are the elements of the one-electron density matrix in the basis of the atomic orbitals such that the $\rho(\mathbf{r})$

$$P_{pq} = \sum_{i=1}^N c_{pi}^* c_{qi} \quad (30)$$

$$\rho(\mathbf{r}) = \sum_i^N \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) = \sum_{pq}^{N_{\text{basis}}} P_{pq} \varphi_p^*(\mathbf{r}) \varphi_q(\mathbf{r}). \quad (31)$$

Finally h_{pq} and $(pq|rs)$ are the one- and two-electron integrals, and $E_{\text{xc}}(P)$ underlines the P density matrix dependence of the exchange-correlation energy, since $\rho(\mathbf{r})$ is computed according to Eq. 31. The KS orbital equation is solved iteratively since $v_{\text{KS}}(\mathbf{r})$ is a functional of the $\rho(\mathbf{r})$ but $\rho(\mathbf{r})$ is determined by $v_{\text{KS}}(\mathbf{r})$. A guess electron density is computed according to Eq. 31 and the E_{xc} functional is calculated to generate the v_{xc} potential. Then one optimizes the electron density by minimizing the KS energy.

The form of the exchange-correlation energy (E_{xc}) within the Kohn–Sham framework is the following

$$E_{\text{xc}}[\rho(\mathbf{r})] = \int f(\rho(\mathbf{r}), \nabla \rho(\mathbf{r}), \nabla^2 \rho(\mathbf{r}), \dots) d\mathbf{r} \quad (32)$$

and the v_{xc} potential of the Eq. 26 is the functional derivative of E_{xc} . For instance if f depends only on $\rho(\mathbf{r})$ and its gradient, v_{xc} is

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} = \frac{\partial f}{\partial \rho(\mathbf{r})} - \frac{\partial f}{\partial \rho_x(\mathbf{r})} - \frac{\partial f}{\partial \rho_y(\mathbf{r})} - \frac{\partial f}{\partial \rho_z(\mathbf{r})}, \quad (33)$$

where $\rho_x(\mathbf{r}) = \partial \rho(\mathbf{r}) / \partial x$. The f function in Eq. 32 is generally written as follows

$$f = \rho(\mathbf{r}) \varepsilon_{\text{xc}}(\rho(\mathbf{r}), \nabla \rho(\mathbf{r}), \nabla^2 \rho(\mathbf{r}), \dots). \quad (34)$$

The exchange-correlation energy per unit volume ε_{xc} is an analytic function with some parameters that can be fitted to some experimental data or determined according to some model systems. The form of ε_{xc} represents the main feature of a given DFT functional and therefore also its reliability and limitations. When ε_{xc} is that of the homogeneous electron gas, one obtains the simplest approximation to the exact E_{xc} called local-density-approximation (LDA) functional. In LDA the ε_{xc} is a function only of the density, but this approximation can be improved by including the explicit dependence on the $\nabla\rho(\mathbf{r})$ (generalized gradient approximation, GGA) or on the kinetic energy (meta-GGA functionals [185]). Among GGA functionals, the Becke–Perdew (BP86) exchange-correlation functional is one of the most used. This functional is composed by the Becke88 exchange [23] and the Perdew GGA correlation [24] terms.

Another important approach to build DFT functionals is the so-called hybrid method based on the adiabatic connection formula [186]. According to the latter, the exact E_{xc} functional is given by the following integral:

$$E_{xc} = \int_0^1 \langle \psi_\lambda | v_{xc}(\lambda) | \psi_\lambda \rangle d\lambda, \quad (35)$$

where λ is the parameter defined in the Hamiltonian operator in Eq. 21, while ψ_λ is the antisymmetric wave function that minimizes the energy functional. To cite Axel Becke from one of his seminal papers on hybrid functionals [20]: “This formula connects the non-interacting Kohn–Sham reference systems to the fully interacting real system through a continuum of partially interacting systems all which share a common density ρ (the density of the real, fully interaction system.” The approximate E_{xc} functional can be obtained by numerical integration. The simplest case is the two end-points approximation in which the integrand E_{xc} is evaluated at $\lambda = 0$ and $\lambda = 1$. The $\langle \psi_0 | v_{xc}(0) | \psi_0 \rangle$ is reduced to the HF exchange, since we are dealing with the non-interacting system. The $\langle \psi_1 | v_{xc}(1) | \psi_1 \rangle$ remains unknown, but it can be approximated using others DFT functionals. The functionals obtained in this simple way are called E_{xc}^{H+H} half-and-half functionals:

$$E_{xc}^{H+H} = \frac{1}{2}E_x^{HF} + \frac{1}{2}E_{xc}^{DFT}. \quad (36)$$

Becke proposed an improvement of the E_{xc}^{H+H} , which is known in the quantum chemistry literature as Becke-3 parameters hybrid exchange-correlation functional (B3):

$$E_{xc}^{B3} = E_{xc}^{LSDA} + a(E_x^{HF} + E_x^{LSDA}) + b\Delta E_x^{B88} + c\Delta E_c^{PW91}. \quad (37)$$

The empirical parameters a , b and c are determined by fitting experimental data. In Eq. 37 LSDA refers to the local-spin-density approximation and ΔE_x^{B88} [23] and ΔE_c^{PW91} [22] are the gradient corrections to the

LSDA exchange and correlation. The most popular hybrid functional is the B3LYP [21], which is similar to B3:

$$E_{xc}^{B3LYP} = E_x^{LSDA} + a (E_x^{HF} + E_x^{LSDA}) + b \Delta E_x^{B88} + E_c^{VWN} + c (\Delta E_c^{LYP} - E_c^{VWN}), \quad (38)$$

where LYP and VWN refer to the Lee-Yang-Parr [22] and Vosko-Wilk-Nussair [187] correlation functionals.

GGA functionals correct some of the LDA shortcomings, meta-GGA ones improve certain properties with respect to GGA, and finally hybrid functionals give significant improvement over GGA for several molecular properties. This seems to represent a constant improvement of the functionals but, as a matter of fact, there is not a unique recipe to accurately reproduce molecular properties using DFT, and the reliability of this approach critically depends on the type of molecule considered and on the properties of interest.

The evaluation of the analytic DFT energy gradients is similar to that of HF [184]:

$$\frac{\partial E_{tot}}{\partial \lambda} = \sum_{pq} \left[\frac{\partial h_{pq}}{\partial \lambda} P_{pq} - \frac{\partial S_{pq}}{\partial \lambda} W_{pq} + \frac{\partial E_{pq}^{xc}}{\partial \lambda} \right] + \sum_{pqrs} \left[\frac{\partial (pq|rs)}{\partial \lambda} P_{pq} P_{rs} \right], \quad (39)$$

where S_{pq} are the overlap integrals and $W_{pq} = \sum_{i=1}^N \varepsilon_i c_{pi} c_{qi}$ is the energy-weighted density matrix. Analytic second derivatives are more complicated since they involve the derivatives of the MO coefficients, which are computed using the coupled-perturbed SCF equations.

B

Outline of Time-Dependent Density Functional Theory

TDDFT is based on the linear response theory [188]. The formal framework of this approach was given by Runge and Gross who have extended the first Hohenberg-Kohn theorem to the time-dependent case [189]. The linear response electron density of a given excited state is [190–193]

$$\rho(\mathbf{r}, \omega) = \sum_{ia} [X_{ia}(\omega) \phi_a^*(\mathbf{r}) \phi_i(\mathbf{r}) + Y_{ai}(\omega) \phi_a(\mathbf{r}) \phi_i^*(\mathbf{r})], \quad (40)$$

where indices i and a run over the occupied and virtual KS ground state MOs, respectively. The frequency dependent coefficients $X_{ia}(\omega)$ are associated with the single excitations $\phi_i \rightarrow \phi_a$, while $Y_{ai}(\omega)$ are related to non-physical de-excitations $\phi_a \rightarrow \phi_i$ which introduce the correlation effects of double-excitations in $\rho(\mathbf{r}, \omega)$. X and Y matrices represent the linear response of the ground state KS MOs to the time-dependent external perturbation $v(t)$ intro-

duced in the time-dependent KS equation with potential $v_{\text{KS}}(\mathbf{r}, t)$ given by

$$v_{\text{KS}}(\mathbf{r}, t) = v(t) + \int \frac{\rho(\mathbf{r}, t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}, t), \quad (41)$$

$\rho(\mathbf{r}, t)$ is the Fourier transform of $\rho(\mathbf{r}, \omega)$, $v(t)$ is the applied field, and when it is turned on very slowly (adiabatic approximation) the $v_{\text{xc}}(\mathbf{r}, t)$ potential can be evaluated as the functional derivative of the static $E_{\text{xc}}[\rho(\mathbf{r})_t]$ with respect to $\rho_t(\mathbf{r})$, where the time-dependent density is evaluated at a fixed time t .

To evaluate X and Y coefficients, one invokes the dependence of the variation of $\rho(\mathbf{r}, \omega)$ on the variation of the KS potential through the KS kernel $\chi_{\text{KS}}(\mathbf{r}, \mathbf{r}', \omega)$, which represents the linear response of the KS non-interacting system:

$$\rho(\mathbf{r}, \omega) = \int \chi_{\text{KS}}(\mathbf{r}, \mathbf{r}', \omega) v_{\text{xc}}(\mathbf{r}', \omega) d\mathbf{r} d\mathbf{r}' \quad (42)$$

$$\chi_{\text{KS}}(\mathbf{r}, \mathbf{r}', \omega) = \sum_{ia} \left[\frac{\phi_i^*(\mathbf{r}) \phi_a(\mathbf{r}) \phi_i(\mathbf{r}') \phi_a^*(\mathbf{r}')}{\omega - (\epsilon_a - \epsilon_i)} + \frac{\phi_i(\mathbf{r}) \phi_a^*(\mathbf{r}) \phi_i^*(\mathbf{r}') \phi_a(\mathbf{r}')}{\omega + (\epsilon_a - \epsilon_i)} \right]. \quad (43)$$

Calculation of the excitation energy requires to find the poles of $\chi_{\text{KS}}(\mathbf{r}, \mathbf{r}', \omega)$. This can be done solving the following non-hermitian eigenvalues-eigenvectors problem

$$\begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \quad (44)$$

$$A_{ai,bj} = \delta_{ab} \delta_{ij} (\epsilon_a - \epsilon_i) + K_{ai,bj} \quad (45)$$

$$B_{ai,bj} = K_{ai,jb} \quad (46)$$

$$K_{ai,bj} = \int \phi_a^*(\mathbf{r}) \phi_i(\mathbf{r}) \left[\frac{1}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta^2 E_{\text{xc}}}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')} \right] \phi_b^*(\mathbf{r}') \phi_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \quad (47)$$

The second functional derivatives of the E_{xc} with respect to $\rho(\mathbf{r}, \omega)$ represents the exchange-correlation kernel $f_{\text{xc}}(\mathbf{r}, \mathbf{r}', \omega)$: the adiabatic approximation can be seen as the static limit $f_{\text{xc}}(\mathbf{r}, \mathbf{r}', \omega = 0)$, which is evaluated as the functional derivative of the stationary potential $\delta v_{\text{xc}} / \delta \rho$. Since the frequency dependence of f_{xc} is neglected, TDDFT cannot give accurate excitation energies when excited states have strong double-excitation character (i.e., for organic polyenes [194]). In fact it can be shown that exact f_{xc} is strongly frequency-dependent near the double excitation [195, 196].

It is worth mentioning the so-called Casida's formulation of the TDDFT [197] in which the problem 21 is written in the Hermitian form $\Omega F_i = \omega_i^2 F_i$. Here the Ω matrix elements are written as a function of the KS orbital energies and the $K_{ai,bj}$ integrals, and F_i eigenvector easily allow the excited state term assignment. Indeed a CIS wave function that corresponds to

the $\rho(r, \omega)$ response density can be calculated from the F_i , which permits one to determine the main excitations involved in a given excited state.

In order to understand the TDDFT level of theory with respect to other methods, one has to point out that when the E_{xc} functional derivative is neglected and $B = 0$, then TDDFT is reduced to CIS. Therefore A elements introduce the correlation effects of the single excited configurations. If one neglects E_{xc} functional derivative in Eq. 46, TDDFT is reduced to time-dependent Hartree–Fock (TDHF). Within TDHF, double-excited configurations are introduced through the B elements since they involve $\phi_i \rightarrow \phi_a$ excitation and $\phi_b \rightarrow \phi_j$ de-excitation at the same time. This is mathematically equivalent to compute ground state/double-excited configuration matrix elements, $\langle \psi_0 | \hat{H}^{i \rightarrow a}_{j \rightarrow b} | \psi_0 \rangle$ where ψ_0 is the single-determinant wave function that corresponds to KS ground state density [198]. Finally if one neglects the $Y_{ia}(\omega)$ elements, the Tamm–Dancoff approximation (TDA) formulation of the TDDFT is obtained [199]. This approach often gives results of the same quality of the full TDDFT approach and it is easier to implement. Besides, in TDDFT one introduces additional correlation effects through the functional derivative of E_{xc} . It could be argued that the TDDFT level of theory is higher than ground state DFT because of the effects of single/double excitations introduced through A and B matrix elements. On the other hand, the inclusion of single and double configurations tries to account for electronic correlation effects as E_{xc} does, and therefore a problem of overestimation of the correlation energy could arise.

Excited state energy first derivatives with respect to the nuclear positions can be derived in an advantageous way through the definition of the variational Lagrangian of the excited state energy $L[X, Y, C_0, \Omega, W, Z]$ [166], where C_0 are the KS ground state MOs coefficients and Ω is the excitation energy. The two extra parameters W and Z constrain the MOs to be orthonormal solutions of the ground state KS equation. When a stationary conditions of L with respect to all parameters is reached, the final expression of the first derivative with respect to a generic nuclear position λ is given by

$$\begin{aligned} \frac{\partial \Omega}{\partial \lambda} = & \sum_{pq} \left[\frac{\partial h_{pq}}{\partial \lambda} P_{pq} - \frac{\partial S_{pq}}{\partial \lambda} W_{pq} + \frac{\partial v_{pq}^{xc}}{\partial \lambda} P_{pq} \right] \\ & + \sum_{pqrs} \left[\frac{\partial (pq|rs)}{\partial \lambda} \Gamma_{pqrs} + \frac{\partial f_{pqrs}^{xc}}{\partial \lambda} (X + Y)_{pq} (X + Y)_{rs} \right] \end{aligned} \quad (48)$$

where i) f_{pqrs}^{xc} are second functional derivatives of E_{xc} integrals which are found in Eq. 7; ii) P and Γ are the relaxed one and two electron density matrices [164] and, together with the W energy weight relaxed density matrix, represents the derivative of the excitation energy with respect to one-electron, two-electron and overlap integrals (h_{pq} , $(pq|rs)$ and S_{pq} , respectively). In Eq. 24 no derivatives of the C_0 coefficients are needed. Moreover, it depends

on h_{pq} , $(pq|rs)$ and S_{pq} derivatives built on the atomic orbital basis set, which are already implemented in the ground state gradients computation. For these reasons, the computational cost of excited state and ground state gradients is similar. Furthermore the recent application of the RI approximation to both excited state energy and gradients calculations, allow the computation on large system up to 4000 basis functions [200].

The TDDFT optimization of the geometrical structure of a given excited state is carried out as follows. Using the KS ground state MOs, one calculates the excited state energy and response density. Afterwards the W and Z elements are computed in order to evaluate the nuclear forces. Finally, the nuclear coordinates are updated, and KS ground state MOs newly calculated.

References

1. Holm RH, Solomon EI (2004) Thematic issue: Biomimetic inorganic chemistry. *Chem Rev* 104:347
2. Koch W, Holthausen MC (2002) A Chemist's guide to density functional theory. Wiley, second edition
3. Niu S, Hall MB (2000) *Chem Rev* 100:353
4. Siegbahn PEM, Blomberg MRA (2000) *Chem Rev* 100:421
5. Lovell T, Himo F, Han W-G, Noodleman L (2003) *Coord Chem Rev* 238-239:211
6. Friesner RA, Baik M-H, Ghernan BF, Guallar V, Wirstam M, Murphy RB, Lippard SJ (2003) *Coord Chem Rev* 238-239:267
7. Ziegler T, Autschbach J (2005) *Chem Rev* 105:2695
8. Siegbahn PEM, Blomberg MRA, Wirstam Pavlov M, Crabtree RH (2001) *J Biol Inorg Chem* 6:460
9. Fan H-J, Hall MB (2001) *J Biol Inorg Chem* 6:467
10. Stein M, Lubitz W (2002) *Curr Opin Chem Biol* 6:243
11. Bruschi M, Zampella G, Fantucci P, De Gioia L (2005) *Coord Chem Rev* 249:1620
12. Hohenberg P, Kohn W (1964) *Phys Rev* 136:B864
13. Parr RG, Yang W (1989) *Density-Functional Theory of Atoms, Molecules*. Oxford University Press, Oxford
14. Levine I (1999) *Quantum Chemistry*. Prentice Hall, fifth edition
15. Dunlap BI, Connolly JWD, Sabin JR (1979) *J Chem Phys* 71:3396
16. Baerends EJ, Ellis DE, Ros P (1973) *Chem Phys* 2:41
17. Eichkorn K, Treutler O, Ohm H, Haser M, Ahlrichs R (1995) *Chem Phys Lett* 240:283
18. Eichkorn K, Wigend F, Treutler O, Ahlrichs R (1997) *Theor Chem Acc* 97:119
19. Becke AD (1992) *J Chem Phys* 96:2155
20. Becke AD (1993) *J Chem Phys* 98:5648
21. Stevens PJ, Devlin FJ, Chablowski CF, Frisch MJ (1994) *J Phys Chem* 98:11623
22. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785
23. Becke AD (1988) *Phys Rev A* 38:3098
24. Perdew JP (1986) *Phys Rev B* 33:8822
25. Cramer CJ (2002) *Essentials of computational chemistry – Theories and models*. Wiley
26. Vignais P (2005) *Coord Chem Rev* 249:1677
27. Krasna AJ, Rittenberg D (1954) *J Am Chem Soc* 76:3015

28. Vignais PM, Billoud B, Meyer J (2001) *FEMS Microbiol Rev* 25:455
29. Lenz O, Friedrich B (2001) *Biospektrum* 6:515
30. Cammack R, Robson R, Frey M (1997) *Hydrogen as a fuel learning from nature*. Taylor and Francis, London UK
31. Albracht SP (1994) *Biochim Biophys Acta* 167:1188
32. Graf EG, Thauer RK (1981) *FEBS Lett* 136:165
33. Adams MWW (1990) *Biochim Biophys Acta* 1020:115
34. Cammack R (1999) *Nature* 397:214
35. Nicolet Y, Lemon BJ, Fontecilla-Camps JC, Peters JW (2000) *Trends Biochem Sci* 25:138
36. Peters JW (1999) *Curr Opin Struct Biol* 9:670
37. Horner DS, Heil B, Happe T, Embley TM (2002) *Trends Biochem Sci* 27:148
38. Nicolet Y, Cavazza C, Fontecilla-Camps JC (2002) *J Inorg Biochem* 91:1
39. Armstrong FA (2004) *Curr Op Chem Biol* 8:133
40. Lyon EJ, Shima S, Buurman G, Chowdhuri S, Batschauer A, Steinbach K, Thauer RK (2004) *Eur J Biochem* 271:195
41. Volbeda A, Charon MH, Piras C, Hatchikian EC, Frey M, Fontecilla-Camps JC (1995) *Nature* 373:580
42. Volbeda A, Garcin E, Pieras C, De ALLacey, Fernandez VM, Hatchikian EC, Frey M, Fontecilla-Camps JC (1996) *J Am Chem Soc* 118:12989
43. Garcin E, Vernede X, Hatchikian EC, Volbeda A, Frey M, Fontecilla-Camps JC (1999) *Structure* 7:557
44. Higuchi Y, Yagi T, Yasuoka N (1997) *Structure* 5:1671
45. Higuchi Y, Ogata H, Miki K, Yasuoka N, Yagi T (1999) *Structure* 7:549
46. Ogata H, Mizoguchi Y, Mizuno N, Miki K, Adachi S, Yasuoka N, Yagi T, Yamauchi O, Hirota S, Higuchi Y (2002) *J Am Chem Soc* 124:11628
47. Nicolet Y, Piras C, Legrand P, Hatchikian EC, Fontecilla-Camps JC (1999) *Structure* 7:13
48. Peters JW, Lanzilotta WN, Lemon BJ, Seefeldt LC (1998) *Science* 282:1853
49. Bagley KA, Duin EC, Roseboom W, Albracht SP, Woodruff WH (1995) *Biochemistry* 34:5527
50. Volbeda A, Martin L, Cavazza C, Matho M, Faber BW, Roseboom W, Albracht SP, Garcin E, Rousset M, Fontecilla-Camps JC (2005) *J Biol Inorg Chem* 10:239
51. Volbeda A, Fontecilla-Camps JC (2005) *Coord Chem Rev* 249:1609
52. Pereira AS, Tavares P, Moura I, Moura JGG, Huynh BH (2001) *J Am Chem Soc* 123:2771
53. Popescu CV, Munck E (1999) *J Am Chem Soc* 121:7877
54. De Lacey A, Stadler C, Cavazza C, Hatchikian EC, Fernandez VM (2000) *J Am Chem Soc* 122:11232
55. Pereira AS, Tavares P, Moura I, Moura JJ, Huynh BH (2001) *J Am Chem Soc* 123:2771
56. Bennet B, Lemon BJ, Peters JW (2000) *Biochemistry* 39:7455
57. Lemon BJ, Peters JW (1999) *Biochemistry* 38:12969
58. Nicolet Y, DeLacey AL, Vernede X, Fernandez VM, Hatchikian EC, Fontecilla-Camps JC (2001) *J Am Chem Soc* 123:1596
59. Evans DJ, Pickett CJ (2003) *Chem Soc Rev* 35:268
60. Rauchfuss TB (2004) *Inorg Chem* 43:14
61. Darensbourg MY, Lyon EJ, Smee JJ (2000) *Coord Chem Rev* 206–207:533
62. Liu X, Ibrahim SK, Tard C, Pickett CJ (2005) *Coord Chem Rev* 249:1641
63. Bouwman E, Reedijk J (2005) *Coord Chem Rev* 249:1555
64. Lyon EJ, Georgakaki IP, Reibenspies JH, Darensbourg MY (1999) *Angew Chem, Int Ed* 38:3178

65. Le Cloirec A, Best SP, Borg S, Davies SC, Evans DJ, Hughes DL, Pickett CJ (1999) *Chem Comm*, p 2285
66. Schmidt M, Contakes SM, Rauchfuss TB (1999) *J Am Chem Soc* 121:9736
67. Dong W, Wang M, Liu X, Jin K, Li G, Wang F, Sun L (2006) *Chem Comm*, p 305
68. Schwartz L, Eilers G, Eriksson L, Gogoll A, Lomoth R, Ott S (2006) *Chem Comm*, p 520
69. Tard C, Liu X, Ibrahim SK, Bruschi M, De Gioia L, Davies SC, Yang X, Wang L-S, Sawers G, Pickett CJ (2005) *Nature* 433:610
70. Nehring JL, Heinekey DM (2003) *Inorg Chem* 42:4288
71. Tye JW, Lee J, Wang H-W, Mejia-Rodriguez R, Reibenspies JH, Hall MB, Darensbourg MY (2005) *Inorg Chem* 44:5550
72. Gao W, Liu J, Ma C, Weng L, Jin K, Chen C, Akermark B, Sun L (2006) *Inorg Chim Acta* 359:1071
73. Artero V, Fontecave M (2005) *Coord Chem Rev* 249:1518
74. Best SP (2005) *Coord Chem Rev* 249:1536
75. Osterloh F, Saak W, Hasse D, Pohl S (1997) *Chem Comm*, p 979
76. Davies SC, Evans DJ, Hughes DL, Longhurst S, Sanders JR (1999) *Chem Comm*, p 1935
77. Smith MC, Barclay JE, Cramer SP, Davies SC, Gu WW, Hughes DL, Longhurst S, Evans DJ (2002) *J Chem Soc, Dalton Trans*, p 2641
78. Marr AC, Spencer DJE, Schroder M (2000) *Coord Chem Rev* 206–207:1055
79. Zhu W, Marr AC, Wang Q, Neese F, Spencer DJE, Blake AJ, Cooke PA, Wilson C, Schroder M (2005) *Proc Natl Acad Sci* 102:18280
80. Stenson PA, Marin-Becerra A, Wilson C, Blake AJ, Mc Master J, Schroder M (2006) *Chem Comm*, p 317
81. Sellman D, Geipel F, Lauderbach F, Heinemann FW (2002) *Angew Chem, Int Ed* 41:632
82. Georgakaki IP, Thomson LM, Lyon EJ, Hall MB, Darensbourg MY (2003) *Coord Chem Rev* 238–239:255
83. Zampella G, Bruschi M, Fantucci P, Razavet M, Pickett CJ, De Gioia L (2005) *Chem Eur J*, 11:509
84. Loschen C, Frenking G (2004) *Inorg Chem* 43:778
85. Zampella G, Bruschi M, Fantucci P, De Gioia L (2005) *J Am Chem Soc* 127:13180
86. Neugebauer J, Reiher M, Kind C, Hess BA (2002) *J Comput Chem* 23:895
87. Cao Z, Hall MB (2001) *J Am Chem Soc* 123:3734
88. Liu Z-P, Hu P (2002) *J Am Chem Soc* 124:5175
89. van der Vlugt JI, Rauchfuss TB, Whaley CM, Wilson SR (2005) *J Am Chem Soc* 127:1601
90. Gloaguen F, Lawrence JD, Rauchfuss TB, Benard M, Rohmer MM (2002) *Inorg Chem* 41:6573
91. Bruschi M, De Gioia L, Zampella G, Reiher M, Fantucci P, Stein M (2004) *J Biol Inorg Chem* 9:873
92. Tye JW, Darensbourg MY, Hall MB (2006) *Inorg Chem* 45:1552
93. Jensen F (1998) *Introduction to computational chemistry*. Wiley
94. van der Vlugt JI, Rauchfuss TB, Whaley CM, Wilson SR (2002) *Chem Eur J* 8:4037
95. Lawrence JD, Li H, Rauchfuss TB, Benard M, Rohmer MM (2001) *Angew Chem Int Ed* 40:1768
96. Niu S, Thomson LM, Hall MB (1999) *J Am Chem Soc* 121:4000
97. Fan H-J, Hall MB (2001) *J Am Chem Soc* 123:3828
98. Niu S, Hall MB (2001) *Inorg Chem* 40:6201

99. Fan H-J, Hall MB (2002) *J Am Chem Soc* 124:394
100. Bruschi M, Fantucci P, De Gioia L (2002) *Inorg Chem* 41:1421
101. Bruschi M, Fantucci P, De Gioia L (2003) *Inorg Chem* 42:4773
102. Zhou T, Mo Y, Liu A, Zhou Z, Tsai KR (2004) *Inorg Chem* 43:923
103. Zhou T, Mo Y, Zhou Z, Tsai K (2005) *Inorg Chem* 44:4941
104. Pavlov M, Siegbahn PEM, Blomberg MRA, Crabtree RH (1998) *J Am Chem Soc* 129:548
105. De Gioia L, Fantucci P, Guigliarelli B, Bertrand P (1999) *Inorg Chem* 38:2658
106. De Gioia L, Fantucci P, Guigliarelli B, Bertrand P (1999) *Int J Quantum Chem* 73:187
107. Stein M, van Lenthe E, Baerends EJ, Lubitz W (2001) *J Am Chem Soc* 123:5839
108. Stein M, Lubitz W (2001) *Phys Chem Chem Phys* 3:5115
109. Stein M, Lubitz W (2001) *Phys Chem Chem Phys* 3:2668
110. Foerster S, Stein M, Brecht M, Ogata H, Higuchi Y, Lubitz W (2003) *J Am Chem Soc* 125:83
111. van Gastel M, Stein M, Brecht M, Schroder A, Lendzian F, Bittl R, Ogata H, Higuchi Y, Lubitz W (2006) *J Biol Inorg Chem* 11:41
112. Stadler C, de Lacey AL, Hernandez B, Fernandez VM, Conesa JC (2002) *Inorg Chem* 41:4417
113. Stadler C, de Lacey AL, Montet Y, Volbeda A, Fontecilla-Camps JC, Conesa JC, Fernandez VM (2002) *Inorg Chem* 41:4424
114. Neese F (2003) *Curr Opin Chem Biol* 7:125
115. van Lenthe E, Baerends EJ, Snijders JG (1993) *J Chem Phys* 99:4597
116. Sadlej AJ, Snijders JG, van Lenthe E, Baerends EJ (1995) *J Chem Phys* 102:1758
117. Vosko SH, Wilk L, Nusair M (1980) *Can J Phys* 58:1200
118. Stein M, Lubitz W (2004) *J Inorg Biochem* 98:862
119. Dole F, Fournel A, Magro V, Hatchikian EC, Bertrand P, Guigliarelli B (1997) *Biochemistry* 36:7847
120. Wang CP, Franco R, Moura JJG, Moura I, Day EP (1992) *J Biol Chem* 267:7378
121. Wang H, Ralston CY, Patil DS, Jones RM, Gu W, Verhagen M, Adams M, Ge P, Riordan C, Marganian CA, Mascharak P, Kovacs J, Miller CG, Collins TJ, Brooker S, Croucher PD, Wang K, Stiefel EI, Cramer SP (2000) *J Am Chem Soc* 122:10544
122. Gu W, Jacquamet L, Patil DS, Wang H-X, Evans DJ, Smith MC, Millar M, Koch S, Eichhorn DM, Latimer M, Cramer SP (2003) *J Inorg Biochem* 93:41
123. Reiher M, Salomon O, Sellmann D, Hess BA (2001) *Chem Eur J* 7:5195
124. Cavallo L, Jacobsen H (2003) *J Phys Chem A* 107:5466
125. Paulsen H, Duelund L, Winkler H, Toftlund HH, Trautwein AX (2001) *Inorg Chem* 40:2201
126. Poli R, Harvey JN (2003) *Chem Soc Rev* 32:1
127. Salomon O, Reiher M, Hess BA (2002) *J Chem Phys* 117:4729
128. Noodleman L, Norman JG Jr (1979) *J Chem Phys* 70:4903
129. Noodleman L (1981) *J Chem Phys* 74:5737
130. Xu Z, Thompson LK, Miller LK, Ruiz DO, Alvarez S (2003) *Inorg Chem* 42:1107
131. Noodleman L, Peng CY, Case DA, Mouesca J-M (1995) *Coord Chem Rev* 144:199
132. Mouesca J-M, Chen JL, Noodleman L, Bashford D, Case DA (1994) *J Am Chem Soc* 116:11898
133. Torres RA, Lovell T, Noodleman L, Case DA (2003) *J Am Chem Soc* 125:1923
134. Noodleman L, Norman JG Jr, Osborne JH, Aizman A, Case DA (1985) *J Am Chem Soc* 107:3418
135. Aizman A, Case DA (1982) *J Am Chem Soc* 104:3269
136. Papaefthymiou V, Millar MM, Munck E (1986) *Inorg Chem* 25:3010

137. Carney MJ, Papaefthymiou GC, Spertalian K, Frankel RB, Holm RH (1988) *J Am Chem Soc* 110:6084
138. Auric P, Gaillard J, Meyer J, Moulis JM (1987) *Biochem J* 242:525
139. Antanaitis BC, Moss TH (1975) *Biochim Biophys Acta* 405:262
140. Middleton P, Dickson DPE, Johnson CE, Rush JD (1988) *Eur J Biochem* 88:135
141. Anderson PW, Hasegawa H (1955) *Phys Rev* 100:675
142. Blondin G, Girerd JJ (1990) *Chem Rev* 90:1359
143. Noodleman L, Baerends EJ (1984) *J Am Chem Soc* 106:2316
144. Papaefthymiou V, Girerd JJ, Moura I, Moura JGG, Munck E (1987) *J Am Chem Soc* 109:4703
145. Noodleman L, Case DA (1992) *Adv Inorg Chem* 38:423
146. Borchch SA, Bominaar EL, Blondin B, Girerd JJ (1993) *J Am Chem Soc* 115:5155
147. Fiedler AT, Brunold TC (2005) *Inorg Chem* 44:9322
148. Bruschi M, Greco C, Fantucci P, De Gioia L, manuscript in preparation
149. Klamt A, Schürmann G (1993) *J Chem Soc Perkin Trans 2* 5:799
150. Klamt A (1995) *J Phys Chem* 99:2224
151. Warshel A, Naray-Szabo G, Sussman F, Hwang K (1989) *Biochemistry* 28:3629
152. Bottoni A, Lanza CZ, Miscione GP, Spinelli D (2004) *J Am Chem Soc* 126:1542
153. Ahlrichs R, Bar M, Haser M, Horn H, Kolmel C (1989) *Chem Phys Lett* 62:165
154. Schafer A, Huber C, Ahlrichs R (1994) *J Chem Phys* 100:5829
155. Dreuw A, Head-Gordon M (2005) *Chem Rev* 105:4009
156. Cai ZL, Sendt K, Reimers JR (2002) *J Chem Phys* 117:5543
157. Tozer DJ, Handy NC (1998) *J Chem Phys* 109:10180
158. Van Gisbergen SJA, Groeneveld JA, Rosa A, Snijders JG, Baerends EJ (1999) *J Phys Chem* 103:6835
159. Hummel P, Oxgaard J, Goddard WA III, Gray HB (2005) *Inorg Chem* 44:2454
160. Petit L, Maldivi P, Adamo C (2006) *J Chem Theory and Comput* 1:953
161. Adamo C, Scuseria G, Barone V (1999) *J Chem Phys* 111:2889
162. Hyotcherl H, Cao J, Zewail AH (2001) *Angew Chem, Int Ed* 20:1532
163. Handy NC, Schaefer HF (1984) *J Chem Phys* 81:5031
164. Van Caillie C, Amos RD (1999) *Chem Phys Lett* 308:249
165. Van Caillie C, Amos RD (2000) *Chem Phys Lett* 317:159
166. Furche F, Ahlrichs R (2002) *J Chem Phys* 117:7433
167. Helgaker T, Jorgensen P (1989) *Theor Chim Acta* 75:111
168. Hutter J (2003) *J Chem Phys* 118:3928
169. Doltsinis NL, Kosov DS (2005) *J Chem Phys* 122:144101
170. Weiss H, Ahlrichs R, Haser M (1993) *J Chem Phys* 99:1262
171. Burcl R, Amos RD, Handy NH (2002) *Chem Phys Lett* 335:8
172. Dreuw A, Dunietz BD, Head-Gordon M (2002) *J Am Chem Soc* 124:12070
173. Reiher M, Kirchner B, Hutter J, Sellmann D, Hess BA (2004) *Chem Eur J* 10:4443
174. Adams MWW (1987) *J Biol Chem* 262:15054
175. Lemon BJ, Peters JW (1999) *Biochemistry* 38:12969
176. Lemon BJ, Peters JW (2000) *J Am Chem Soc* 122:3793
177. Bennett B, Lemon BJ, Peters JW (2000) *Biochemistry* 39:7455
178. Chen Z, Lemon BJ, Huang S, Swartz JD, Peters JW, Bagley KA (2002) *Biochemistry* 41:2036
179. Fiedler AT, Brunold TC (2005) *Inorg Chem* 44:1794
180. Albracht SPJ, Roseboom W, Hatchikian EC (2006) *J Biol Inorg Chem* 11:88
181. Bertini L, Greco C, Fantucci P, De Gioia L, manuscript in preparation
182. Amara P, Volbeda A, Fontecilla-Camps JC (1999) *J Am Chem Soc* 121:4468

183. Kohn W, Sham LJ (1965) *Phys Rev* 140:A1133
184. Johnson BG, Fisch MJ (1994) *J Chem Phys* 100:7429
185. Perdew JP, Kurth S, Zupan A, Blaha P (1999) *Phys Rev Lett* 82:2544
186. Harris J (1984) *Phys Rev A* 29:1648
187. Vosko S, Wilk L, Nussair M (1980) *Can J Phys* 58:1200
188. McWeeny R (1989) *Methods of Molecular Quantum Mechanics*, 2. edn. Academic Press, London
189. Runge E, Gross EKV (1984) *Phys Rev Lett* 52:997
190. Gross EKV, Kohn W (1985) *Phys Rev Lett* 55:2850
191. Petersilka M, Gossmann UJ, Gross EKV (1996) *Phys Rev Lett* 76:1212
192. Marques MAL, Gross EKV (2004) *Ann Rev Phys Chem* 55:427
193. Bauernschmitt R, Ahlrichs R (1996) *Chem Phys Lett* 256:454
194. Hsu C-P, Haed-Gordon M (2001) *J Phys Chem A* 105:451
195. Gorling A (1998) *Phys Rev A* 57:3433
196. Aryasetiawan F, Gunnarsson O (2002) *Phys Rev B* 66:16–5119
197. Casida M (1995) *Recent Advances in Density Functional Methods*, chapter Time-Dependent Density Functional Response Theory for Molecules. World Scientific, Singapore, p 220
198. Stratmann RE, Scuseria GE, Frisch MJ (1998) *J Chem Phys* 109:8218
199. Hirata S, Head-Gordon M (1999) *Chem Phys Lett* 314:291
200. Rappoport D, Furche F (2005) *J Chem Phys* 122:1277

Theoretical Bioinorganic Spectroscopy

Sebastian Sinnecker · Frank Neese (✉)

Institut für Physikalische und Theoretische Chemie, Universität Bonn, Wegelerstr. 12,
 53115 Bonn, Germany
theochem@thch.uni-bonn.de

1	Introduction	48
2	Theoretical Methods for Bioinorganic Spectroscopy	50
2.1	Magnetic Spectroscopy of the Ground State	50
2.2	Linear Response Theory for Ground State Properties	53
2.3	Mössbauer Spectroscopy	60
2.4	Optical Spectroscopy	62
2.5	Environment Effects	63
3	Recent Case Studies	64
3.1	High Precision Studies of Organic Radicals	64
3.1.1	Molecular <i>g</i> -Tensors of Phenoxyl Radicals	65
3.1.2	Quinone Radical Anion–Solvent Interactions	66
3.1.3	The Cysteine Radical and Its Unusual <i>g</i> -Tensor Anisotropy	67
3.2	The Spectroscopy of Blue Copper Proteins	70
3.3	Mononuclear Iron Complexes – from Outstanding Model Systems to Fascinating Enzymes	71
3.4	Characterization of an Elusive Fe(V) Species	76
3.5	Binuclear Mixed Valence Copper Complexes – ENDOR and HYSCORE Spectra Simulations from DFT	78
4	Conclusions and Outlook	79
	References	80

Abbreviations

ABS	absorption spectroscopy
CASSCF	complete active space SCF
CD	circular dichroism
DFT	density functional theory
ENDOR	electron-nuclear double resonance
EPR	electron paramagnetic resonance
ESEEM	electron spin echo envelope modulation
ET	electron transfer
EXAFS	extended X-ray absorption fine structure
HDvV	Heisenberg–Dirac–vanVleck Hamiltonian
HF	Hartree–Fock
HFCs	hyperfine coupling constants
HYSCORE	hyperfine sublevel correlation
MB	Mössbauer

MCD	magnetic circular dichroism
MCSCF	multiconfigurational self-consistent field
MD	molecular dynamics
MO	molecular orbital
MRCI	multi reference configuration interaction
MRPT	multi reference perturbation theory
NMR	nuclear magnetic resonance
NQCs	nuclear quadrupole coupling constants
OZ	orbital Zeeman
PCET	proton-coupled electron transfer
QM/MM	quantum mechanical/molecular mechanical
rR	resonance-Raman
SCF	self-consistent field
SH	spin-Hamiltonian
SOC	spin-orbit coupling
SOMF	spin-orbit mean-field
SOMO	semi occupied molecular orbital
SORCI	spectroscopy oriented configuration interaction
SS	spin-spin
TauD	taurine/ α -ketoglutarate dioxygenase
TDA	Tamm-Dancoff approximation
TD-DFT	time-dependent DFT
XAS	X-ray absorption
XC	exchange-correlation
ZFS	zero-field splitting

1

Introduction

The past decade has witnessed an explosive activity in the application of quantum chemical methods to problems of bioinorganic chemistry. It is now almost commonplace to publish an experimental paper together with calculations to back up the conclusions that have been reached or to distinguish between those possibilities that the experimental results left open. In this context the term “quantum chemical methods” has been applied almost synonymously with density functional theory (DFT). Despite of or perhaps even due to the enormous popularity of DFT in bioinorganic chemistry, one frequently voiced complaint from the experimental community is that the calculations often take too little or no notice of the available experimental results. The majority of theoretical papers deal with the calculation of reaction mechanisms which are either proposed or rejected based on theoretically calculated reaction- or transition state energies. Since there are seldom accurate and detailed experimental data available for these thermodynamic and kinetic parameters, the accuracy of the calculations may sometimes be difficult to assess. In this respect it is important to connect the computations to the vast body of experimental data that has been assembled by the ex-

perimental bioinorganic community over the past four decades. Traditionally, a large part of this experimental data is of spectroscopic origin. The reason for this central role that spectroscopy plays in bioinorganic chemistry is that these methods allow one to *selectively* probe the active sites of metalloproteins and metalloenzymes. Secondly, spectroscopic methods can be applied to short-lived intermediates, mixtures or only partially purified proteins, which makes them a versatile tool to follow biochemical events or elementary reaction steps. Thirdly, spectroscopic methods do not only yield geometric structure information but also information at subatomic resolution, e.g. electronic structure information. Since understanding electronic structures is at the heart of understanding the unique reactivities of metalloenzymes, spectroscopy is an extremely useful complement to the X-ray crystallography. The latter technique has the highest information content with respect to geometric structure. However, there are limitations too which concern short-lived species, mixtures and the necessity to perform the experiments on single crystals which may or may not be stable in the X-ray beam.

An important aspect of many active sites of metalloproteins and radical enzymes is that they exist, at least transiently, in open-shell configurations. The presence of unpaired electrons then gives rise to paramagnetism and allows the spectroscopist to probe the active site with various magnetic techniques including (in order of increasing resolution) magnetic susceptibility, magnetic circular dichroism (MCD), electron paramagnetic resonance (EPR), electron-nuclear double resonance (ENDOR), electron spin echo envelope modulation (ESEEM) or hyperfine sublevel correlation spectroscopy (HYSCORE). The EPR, ENDOR, ESEEM, and HYSCORE techniques are covered concisely in a recent book by Schweiger and Jeschke [1], susceptibility is covered in many textbooks on molecular magnetism [2–4], and MCD spectroscopy is covered in a number of references [5–8]. In most cases, the metallic active sites or protein cofactors are also amenable to studies with various optical spectroscopies including standard absorption spectroscopy (ABS), circular dichroism (CD), MCD and also resonance-Raman (rR) spectroscopy [9–11]. Finally, the spectroscopy of inner shells of transition metal atoms and ions is of increasing importance in bioinorganic chemistry. X-ray based techniques such as X-ray absorption (XAS) and extended X-ray absorption fine structure (EXAFS) are heavily employed since they allow element specific studies to be carried out [12–14]. Finally, due to the abundance and importance of iron in bioinorganic chemistry, the technique of Mössbauer (MB) spectroscopy is of considerable utility and yields valuable local information about iron centers including their magnetic properties if the experiments are carried out with an applied external magnetic field [15]. It is impossible to cover all of these techniques at any level of detail in the framework of a single piece of work and therefore most of our examples are selected from the fields of EPR and MB spectroscopy where the theoretical techniques have turned out to be particularly useful.

The central theme of this work is to show how one can combine spectroscopic techniques with theoretical investigations in order to obtain insight into the geometric and electronic structures of metalloproteins and model complexes as well as biologically important radical species. The majority of examples is taken from recent investigations in our own laboratory and should be reasonably representative of the considerable activities that are now underway in many research groups. An attempt will be made to provide the reader with at least a rough idea of some of the methodological aspects that underlies the theoretical approaches used and which are either based on DFT or on simplified correlated *ab initio* methods. More comprehensive information on the various theoretical methods with a full coverage of the recent literature may be found in recent reviews [16–19].

2

Theoretical Methods for Bioinorganic Spectroscopy

2.1

Magnetic Spectroscopy of the Ground State

The active sites of metalloproteins are always of low-symmetry and therefore one can safely assume that the systems under investigation are orbitally nondegenerate. As mentioned above, many metalloprotein active sites exist in open-shell ground state configurations. The presence of unpaired electrons is intimately related to the occurrence of a net molecular magnetic dipole moment and consequently to a rich variety of molecular magnetic phenomena.

In general, a ground state with total spin S gives rise to $2S + 1$ “magnetic sublevels” with $M_S = S, S - 1, \dots, -S$. At the level of the Born–Oppenheimer Hamiltonian, which covers the leading electrostatic interactions between electrons and nuclei, these $2S + 1$ sublevels are all degenerate and this does not change upon the inclusion of the so-called “scalar” relativistic effects. However, if the spin–orbit coupling (SOC) and the direct dipolar spin–spin (SS) coupling as the leading spin-dependent relativistic effects are included in the treatment, the degeneracy of the $2S + 1$ sublevels is lifted. This lifting of degeneracy is referred to as “zero-field splitting” (ZFS) and is parameterized by a tensor with symbol D .

If an external magnetic field is applied, the energies of the magnetic sublevels are further modified. This is the molecular electronic Zeeman effect and it is parameterized by the so-called molecular “g-tensor” (symbol $\mathbf{g}$). The g-tensor arises (mainly) from the SOC induced coupling of the electronic spin- and orbital-moments to the external magnetic field. It may be thought of as the EPR analogue of the NMR chemical shift tensor. For a $S = 1/2$ system the molecular Zeeman effect leads to a linear divergence of the energies of

the $M_S = -1/2$ and $M_S = +1/2$ states at a rate that is proportional to the effective g -value which depends in turn on the relative orientation of the molecule to the external magnetic field. However, while the chemical shift is specific for a given nucleus, the g -tensor is a “system-property” since it measures the distribution of “orbital currents” over the entire molecule which is closely related to the response of the spin-density distribution to the presence of the external magnetic field.

Inclusion of the nuclear spins leads to further magnetic effects which can be studied by magnetic spectroscopies. First of all, the nuclear spin interacts with the external magnetic field, thus giving rise to the nuclear Zeeman effect which is parameterized by the chemical-shift tensor (σ). It occupies a central role in the interpretation of nuclear magnetic resonance (NMR) experiments. Secondly, the nuclear spins interact with the electron spins via three distinct mechanisms: (a) the well-known (isotropic) Fermi contact interaction which is proportional to the probability of finding unpaired electrons at the position of the magnetic nuclei, (b) an anisotropic (traceless) electron spin-nuclear spin dipolar interaction and (c) the anisotropic (nontraceless) SOC induced coupling of the nuclear spins with the orbital motions of the unpaired electrons. The sum of these interactions is parameterized by the hyperfine coupling tensor (symbol A). Finally, the indirect coupling of nuclear spins leads to very small effects which show up as spin-spin couplings in NMR spectroscopy and are parameterized by a spin-spin coupling tensor (symbol J). Larger effects occur for nuclei with a nuclear spin $I > 1/2$. In this case, the nuclear quadrupole moment may interact with the electric field gradient produced by the surrounding charges. This is referred to as quadrupole coupling and it is measured by the quadrupole tensor (symbol Q).

The very powerful concept of the “Spin-Hamiltonian” (SH) allows one to treat all of these complicated interactions in a semi-phenomenological way where a minimum of empirical parameters allows the deduction of the entire spin-physics of the system under investigation. The SH is an effective Hamiltonian and contains only spin-variables of a “fictitious” electron spin S and the nuclear spins $I_A, I_B, \dots$. All reference to the spatial part of the wavefunction and therefore to detailed molecular electronic and geometric structure is implicitly contained in the SH parameters D, g, A, Q, σ and J which are considered as adjustable parameters in the analysis of experiments. Thus, the SH is “a convenient place to rest” in the analysis of experimental data by theoretical means [20]. The SH which includes the interactions covered above is:

$$\hat{H}_{\text{SPIN}} = \hat{S}D\hat{S} + \beta B g \hat{S} + \sum_A \left[\hat{S}A^{(A)} \hat{I}^{(A)} + \beta_N B g_N^{(A)} \hat{I}^{(A)} + \hat{I}^{(A)} Q^{(A)} \hat{I}^{(A)} \right] + \sum_{A < B} \left[\hat{I}^{(A)} J^{(AB)} \hat{I}^{(B)} \right], \quad (1)$$

where the sum over “ A ” refers to the magnetic nuclei, B is the magnetic flux density and β and β_N is the electronic and nuclear Bohr magneton, respectively. The SH acts on a basis of product functions $|SM_S\rangle \otimes |I^{(A)}M_I^{(A)}\rangle \otimes \dots \otimes |I^{(N)}M_I^{(N)}\rangle$.

For not too many spins this basis is often small enough to allow exact diagonalization of the SH and therefore exact quantum mechanical treatments of the spin-physics in the SH framework. For high-dimensional SH problems, both, brute-force (Piligkos et al. 2006) and a variety of perturbation theoretical methods can be employed in order to arrive at exact or good approximate solutions.

The role of theory is then to derive the connection of first principle electronic structure approaches to the SH parameters and to devise practical algorithms for the prediction of the values of these parameters. In this way, calculations can help to develop the full information content of spectra and thus allow conclusions about the geometric and electronic structure of the system under investigation to be drawn (Fig. 1). A great deal of progress has been made in recent years in the calculation of EPR parameters (summarized in [21–23]). The theory of NMR parameters is a well-studied field in theoretical chemistry and mature procedures with well-defined accuracy for the desired quantities have been established [21].

In many cases, it is sensible to divide the system under investigation into distinct fragments with given electron spins $S_A, S_B, \dots$. The interaction of the electron spins of the different spin-systems is usually referred to as exchange coupling and is given by the symbol J [note that unfortunately the same symbol is usually used for the nuclear spin-spin interaction (see e.g. Eq. 1). It should, however, from the context be evident to which interaction J refers]. The interaction consists of the genuine isotropic “exchange interaction” J which is purely electrostatic in origin and only arises from the antisymmetry requirement of the N -electron wavefunction with respect to electron permutations (consequently, the notion of an “exchange interaction” is misleading. In fact, such an interaction does not exist in nature. The “exchange coupling” is a consequence of the interplay between electron–electron repulsion and the Fermion nature of electrons). Unfortunately, there are several popular ways to write the isotropic exchange Hamiltonian (Heisenberg–Dirac–vanVleck Hamiltonian, HDvV). Perhaps the most common way is to write:

$$\hat{H}_{\text{HDvV}} = -2J \hat{S}_A \hat{S}_B. \quad (2)$$

In this formalism two interacting electrons give rise to a singlet-state ($S_t = 0$) and a triplet state ($S_t = 1$) which are split in energy by $|2J|$. Positive values of J lead to $S_t = 1$ below $S_t = 0$ (ferromagnetic coupling), while negative values of J lead to the reverse situation (antiferromagnetic coupling). In addition to the HDvV Hamiltonian, there exists the direct dipole–dipole interaction between the electron spins which also contributes in an anisotropic and trace-

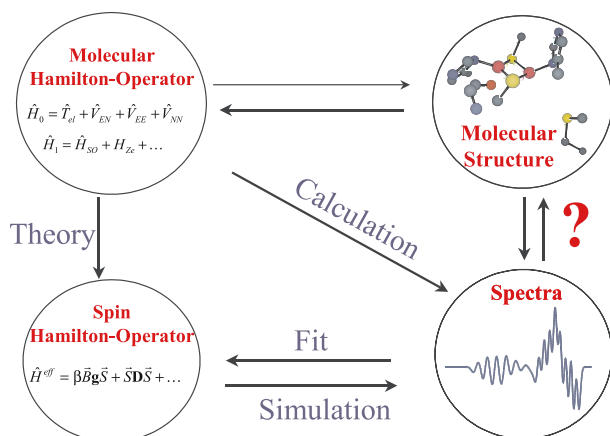


Fig. 1 The relationship of theory and experiment. Electronic structure calculations solve the Schrödinger equation for a system that is characterized by its molecular Hamiltonian and provide the spin Hamiltonian (SH) parameters, e.g. g -tensors, hyperfine coupling constants or Mössbauer parameters. These data are also the result of the spectral simulations and therefore the intersection of theory and experiment. From a careful comparison of simulated and calculated SH parameters, conclusions on electronic and even molecular structures can be drawn

less way to J . Very interesting phenomena arise from the combination of the exchange interaction with the “on-site” spin-Hamiltonians described in Eq. 1. Entire books are devoted to the study of such interactions [2, 24].

2.2

Linear Response Theory for Ground State Properties

All SH parameters can be approached using the powerful language of linear response or analytic derivative techniques. For the present purposes it is sufficient to focus on the time independent case. In order to appreciate the general concepts that are involved, the linear response equations for a Self-Consistent Field (SCF) ground state will be sketched below. This description is appropriate if the state of interest is well described by a HF (Hartree–Fock) or DFT single determinant:

$$\Psi(x_1, x_2, \dots, x_N) = |\psi_1 \dots \psi_N|, \quad (3)$$

where the $\{\psi\}$ constitute the set of occupied spin orbitals and x_i denotes the space and spin coordinates of the i th electron. The ground state energy is written as:

$$E = E_{NN} + \sum_i \langle \psi_i | \hat{h} | \psi_i \rangle + \frac{1}{2} \sum_{ij} (ii|jj) - c_{HF}(ij|ij) + c_{DF}E_{XC}[\rho]. \quad (4)$$

Here, E_{NN} is the nuclear–nuclear repulsion energy, $\hat{h}$ is the one-electron operator, $(ij|kl)$ is the electron–electron repulsion integral

$$(ij|kl) = \iint \psi_i^*(\mathbf{x}_1) \psi_j(\mathbf{x}_1) r_{12}^{-1} \psi_k^*(\mathbf{x}_2) \psi_l(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad (5)$$

and $E_{\text{XC}}[\rho]$ is the exchange–correlation energy functional with the ground state density being given by:

$$\rho(\mathbf{x}) = \sum_i \psi_i^*(\mathbf{x}) \psi_i(\mathbf{x}) . \quad (6)$$

The parameters c_{HF} and c_{DF} are scaling parameters for the HF exchange energy and the XC-energy, respectively. Thus, HF theory corresponds to $c_{\text{HF}} = 1$; $c_{\text{DF}} = 0$, “pure” DFT corresponds to $c_{\text{HF}} = 0$; $c_{\text{DF}} = 1$ while hybrid DFT methods choose $0 < c_{\text{HF}} < 1$. The orbitals are determined self-consistently from the SCF equations:

$$\left\{ \hat{h} + \int \frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d\mathbf{x}' - c_{\text{HF}} \sum_i \hat{K}^{ii} + c_{\text{DF}} \frac{\delta E_{\text{XC}}[\rho]}{\delta \rho(\mathbf{x})} \right\} \psi_i(\mathbf{x}) = \varepsilon_i \psi_i(\mathbf{x}) , \quad (7)$$

where the second term represents the Coulomb-potential, the third term the HF exchange interaction with the nonlocal operator:

$$\langle \psi_k | \hat{K}^{ij} | \psi_l \rangle = (il|kj) . \quad (8)$$

We also define the Coulomb operator:

$$\langle \psi_k | \hat{J}^{ij} | \psi_l \rangle = (ij|kl) . \quad (9)$$

The term in brackets is collectively referred to as $\hat{F}$, the pseudo-one-electron SCF operator. The orbitals are usually approximated by a finite expansion in a basis set $\{\varphi\}$:

$$\psi_i(\mathbf{x}) = \sum_{\mu} c_{\mu i} \varphi_{\mu}(\mathbf{x}) \quad (10)$$

which results in the following expression for the density:

$$\rho(\mathbf{x}) = \sum_{\mu\nu} P_{\mu\nu} \varphi_{\mu}^*(\mathbf{x}) \varphi_{\nu}(\mathbf{x}) \quad (11)$$

with the density matrix

$$P_{\mu\nu} = \sum_i c_{\mu i}^* c_{\nu i} . \quad (12)$$

In order to approach molecular properties one needs to determine the derivatives of the total SCF energy with respect to external perturbations $\lambda, \kappa, \dots$. Since all SH parameters are bilinear in external perturbations, the desired quantity is the second partial derivative. For illustrating the concepts, it is

sufficient to consider the case where the basis functions are chosen to be independent on the external perturbations. To include such a dependence (as is necessary for example for geometric or magnetic field perturbations) is straightforward but would lead to more lengthy equations which are not of interest for the purpose of this work. Since the c 's are determined in a variational procedure, one has:

$$\frac{\partial E}{\partial c_{\mu i}} \frac{\partial c_{\mu i}}{\partial \lambda} = 0. \quad (13)$$

And therefore the first derivative of the energy with respect to a perturbation λ is:

$$\left. \frac{\partial E}{\partial \lambda} \right|_{\lambda=0} = \sum_{\mu\nu} P_{\mu\nu} \langle \varphi_{\mu} | \hat{h} | \varphi_{\nu} \rangle_{\lambda}, \quad (14)$$

where:

$$\langle \varphi_{\mu} | \hat{h} | \varphi_{\nu} \rangle_{\lambda} = \left\langle \varphi_{\mu} \left| \frac{\partial \hat{h}}{\partial \lambda} \right| \varphi_{\nu} \right\rangle \quad (15)$$

if the basis functions are independent of the perturbation. Through straightforward differentiation the second partial derivative becomes:

$$\left. \frac{\partial^2 E}{\partial \lambda \partial \kappa} \right|_{\lambda=0; \kappa=0} = \sum_{\mu\nu} P_{\mu\nu} \langle \varphi_{\mu} | \hat{h} | \varphi_{\nu} \rangle_{\lambda\kappa} + \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial \kappa} \langle \varphi_{\mu} | \hat{h} | \varphi_{\nu} \rangle_{\lambda}. \quad (16)$$

This important equation contains two contributions: the first term is referred to as a first-order contribution since it only depends on the ground state density. The second contribution is a second-order contribution since it requires the knowledge of the first derivative of the density matrix with respect to an external perturbation.

In order to calculate the derivative of the density matrix a number of techniques could be employed. The most straightforward approach is to start from the condition of a SCF solution:

$$FPS - SPF = 0, \quad (17)$$

where F is the matrix of $\hat{F}$ in the AO basis, S is the overlap matrix and P the density matrix. Differentiation with respect to λ yields the equations that are necessary to solve for the perturbed density matrix $P^{(\lambda)}$. We will pursue a more pedestrian approach here and determine the perturbed orbitals $\psi_i^{(\lambda)} \equiv |i^{(\lambda)}\rangle$ themselves in terms of the zero'th order orbitals $\psi_i^{(0)} \equiv |i^{(0)}\rangle$. Differentiation of the SCF equations yields:

$$\left\{ \hat{F}^{(0)} - \varepsilon_i^{(0)} \right\} |i^{(\lambda)}\rangle + \left\{ \hat{F}^{(\lambda)} - \varepsilon_i^{(\lambda)} \right\} |i^{(0)}\rangle = 0. \quad (18)$$

The perturbed orbitals are expanded as:

$$|i^{(\lambda)}\rangle = \sum_a U_{ai}^{(\lambda)} |a^{(0)}\rangle \quad (19)$$

(here and below labels i, j, k, l refer to occupied orbitals and a, b, c, d to unoccupied ones). The unitary matrix U has only occ/virt blocks in the case that the basis functions do not depend on the perturbation. In order to determine the unique elements of U one uses the perturbed SCF equations:

$$\sum_b U_{bi}^{(\lambda)} \langle a^{(0)} | \hat{F}^{(0)} - \varepsilon_i^{(0)} | b^{(0)} \rangle + \langle a^{(0)} | \hat{F}^{(\lambda)} - \varepsilon_i^{(\lambda)} | i^{(0)} \rangle = 0 \quad (20)$$

$$= U_{ai}^{(\lambda)} (\varepsilon_a^{(0)} - \varepsilon_i^{(0)}) + \langle a^{(0)} | \hat{F}^{(\lambda)} | i^{(0)} \rangle = 0. \quad (21)$$

However, $\hat{F}^{(\lambda)}$ depends on the perturbed orbitals. Therefore, one needs to take the derivative of the SCF operator carefully.

$$\begin{aligned} \hat{F}^{(\lambda)} &= \hat{h}^{(\lambda)} + \int \frac{\rho^{(\lambda)}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} - c_{\text{HF}} \sum_{jb} U_{bj}^{(\lambda)*} \hat{K}^{bj} + U_{bj}^{(\lambda)} \hat{K}^{jb} \\ &\quad + c_{\text{DF}} \iint \frac{\delta^2 E_{\text{XC}}[\rho]}{\delta \rho(\mathbf{x}) \delta \rho(\mathbf{x}')} \rho^{(\lambda)}(\mathbf{x}') d\mathbf{x} d\mathbf{x}' \\ &= \hat{h}^{(\lambda)} + \sum_{jb} U_{bj}^{(\lambda)*} \hat{J}^{bj} + U_{bj}^{(\lambda)} \hat{J}^{jb} - c_{\text{HF}} U_{bj}^{(\lambda)*} \hat{K}^{bj} - c_{\text{HF}} U_{bj}^{(\lambda)} \hat{K}^{jb} \\ &\quad + c_{\text{DF}} \int f_{\text{xc}}[\rho] \rho^{(\lambda)}(\mathbf{x}) d\mathbf{x} \\ &= \hat{h}^{(\lambda)} + \sum_{jb} U_{bj}^{(\lambda)*} \left[\hat{J}^{bj} - c_{\text{HF}} \hat{K}^{bj} + c_{\text{DF}} \int f_{\text{xc}}[\rho] \psi_b^{(0)*}(\mathbf{x}) \psi_j^{(0)}(\mathbf{x}) d\mathbf{x} \right] \\ &\quad + U_{bj}^{(\lambda)} \left[\hat{J}^{jb} - c_{\text{HF}} \hat{K}^{jb} + c_{\text{DF}} \int f_{\text{xc}}[\rho] \psi_j^{(0)*}(\mathbf{x}) \psi_b^{(0)}(\mathbf{x}) d\mathbf{x} \right], \end{aligned} \quad (22)$$

where the “XC-kernel” $f_{\text{xc}}[\rho]$ has been defined as the second functional derivative with respect to ρ and it has been tacitly assumed that for all functionals in use this yields a factor $\delta(\mathbf{x} - \mathbf{x}')$ which reduces the double integral to a single integral. This yields the first-order equations:

$$\begin{aligned} &U_{ai}^{(\lambda)} (\varepsilon_a^{(0)} - \varepsilon_i^{(0)}) + \langle a^{(0)} | \hat{h}^{(\lambda)} | i^{(0)} \rangle \\ &\quad + \sum_{jb} U_{bj}^{(\lambda)*} [(bj|ai) - c_{\text{HF}}(ba|ji) + (ai|f_{\text{xc}}|jb)] \\ &\quad + U_{bj}^{(\lambda)} [(jb|ai) - c_{\text{HF}}(ja|bi) + (ai|f_{\text{xc}}|jb)] = 0. \end{aligned} \quad (23)$$

At this point it is useful to distinguish two different types of perturbation: First, “electric field-like perturbations” yield purely real $\langle a^{(0)} | \hat{h}^{(\lambda)} | i^{(0)} \rangle$ and

consequently, also purely real and symmetric U matrices. In this case one has:

$$A^{(E)} U^{(\lambda)} = - V^{(\lambda)} . \quad (24)$$

With:

$$A_{ia,jb}^{(E)} = \delta_{ij} \delta_{ab} \left(\varepsilon_a^{(0)} - \varepsilon_i^{(0)} \right) + 2(jb|ia) + 2c_{DF}(ai|f_{xc}|jb) - c_{HF} \{ (ba|ji) + (ja|bi) \} \quad (25)$$

$$V_{ai}^{(\lambda)} = \left\langle a^{(0)} | \hat{h}^{(\lambda)} | i^{(0)} \right\rangle . \quad (26)$$

Note that the A -matrix (the “electric hessian”) is independent of the nature of the perturbation and that the U and V matrices have been written as vectors with a compound index (ai).

Secondly, “magnetic field like perturbation” yield purely hermitian imaginary $\langle a^{(0)} | \hat{h}^{(\lambda)} | i^{(0)} \rangle$ and consequently also purely imaginary and hermitian U matrices. This leads to:

$$A^{(M)} U = - V^{(\lambda)} \quad (27)$$

with

$$A_{ia,jb}^{(M)} = \delta_{ij} \delta_{ab} \left(\varepsilon_a^{(0)} - \varepsilon_i^{(0)} \right) + c_{HF} \{ (ib|ja) - (ba|ij) \} . \quad (28)$$

Thus, magnetic field-like perturbations yield much easier response (or “coupled perturbed”) equations in which the contributions from any local potential vanish. In fact, in the absence of HF exchange the A -matrix becomes diagonal and the linear equation system is trivially solved. This then leads to a “sum-over-orbital” like equation for the second derivative which resembles in some way a “sum-over-states” equation. One should, however, carefully distinguish the sum-over-states picture from linear response or analytic derivative techniques since they have a very different origin. For electric field-like perturbations or magnetic field-like perturbations in the presence of HF exchange one thus has to solve a linear equation system of the size N (occupied) $\times$ N (virtual) which may amount to a dimension of several hundred thousand coefficients in large-scale applications. However, there are efficient iterative techniques to solve such large equations systems without ever explicitly constructing the full A -matrix. The key step of all existing procedures is the formation of a so-called “sigma” vector which is the product of the A -matrix with a trial U -vector.

In any case, once the perturbed orbitals have been determined the perturbed density is easily found as:

$$P_{\mu\nu}^{(\lambda)} = \sum_i U_{ai}^{(\lambda)*} c_{\mu a}^{(0)} c_{\nu i}^{(0)} + U_{ai}^{(\lambda)} c_{\mu i}^{(0)} c_{\nu a}^{(0)} . \quad (29)$$

In order to illustrate the use of the equations above in a nontrivial example, let us study the dominant contribution to the g -tensor. The g -tensor reads in

the SH formalism:

$$\hat{H}_{\text{spin}} = \beta \mathbf{B} \mathbf{g} \hat{\mathbf{S}}, \quad (30)$$

where β is Bohr's magneton and $\mathbf{B}$ the magnetic flux density. First, it is important to note that the order in which the perturbations are applied is immaterial owing to the "interchange theorem"; thus the more easily handled perturbation is conveniently treated first. In the case of the g -tensor, where the two perturbations in question are the spin-independent orbital Zeeman (OZ) operator and the spin-dependent SOC operator, it is more convenient to treat the OZ operator as the λ and the SOC operator as the κ perturbation. The SOC operator is taken in an accurate mean-field approximation (SOMF) developed by Hess et al. [25] and efficiently implemented into the ORCA program recently [26]. The SOMF operator reads:

$$\hat{h}^{\text{SOMF}} = \sum_i \hat{\mathbf{z}}^{\text{SOMF}}(i) \hat{\mathbf{s}}(i). \quad (31)$$

With the matrix elements of the k 'th component of the SOMF operator given by:

$$\begin{aligned} \langle \varphi_\mu | \hat{z}_k^{\text{SOMF}} | \varphi_\nu \rangle = & \langle \varphi_\mu | \hat{h}_k^{\text{el-SO}} | \varphi_\nu \rangle + \sum_{\kappa\tau} P_{\kappa\tau} \left[(\varphi_\mu \varphi_\nu | \hat{g}_k^{\text{SO}} | \varphi_\kappa \varphi_\tau) \right. \\ & \left. - \frac{3}{2} (\varphi_\mu \varphi_\kappa | \hat{g}_k^{\text{SO}} | \varphi_\tau \varphi_\nu) - \frac{3}{2} (\varphi_\tau \varphi_\nu | \hat{g}_k^{\text{SO}} | \varphi_\mu \varphi_\kappa) \right] \end{aligned} \quad (32)$$

and:

$$\hat{h}_k^{\text{el-SO}}(\mathbf{r}_i) = \frac{\alpha^2}{2} \sum_i \sum_A Z_A r_{iA}^{-3} \hat{l}_{iA;k} \quad (33)$$

$$\hat{g}_k^{\text{SO}}(\mathbf{r}_i, \mathbf{r}_j) = -\frac{\alpha^2}{2} \hat{l}_{ij;k} r_{ij}^{-3}. \quad (34)$$

Here, $\mathbf{P}$ is the total charge density matrix, Z_A the nuclear charge of atom A , r_{iA} is the position of electron i relative to nucleus A and $\hat{l}_{iA;k}$ is the k 'th component of the angular momentum of the i 'th electron relative to atom A . Likewise, r_{ij} is the distance between electrons i and j and $\hat{l}_{ij;k}$ is the k 'th component of the angular momentum of electron i relative to electron j . The OZ operator referred to a global gauge origin (O) is:

$$\hat{h}^{\text{OZ}} = \beta \mathbf{B} \sum_i \mathbf{l}^{(O)}(i). \quad (35)$$

Thus, one solves three sets of coupled-perturbed equations in order to obtain three sets of U -coefficients for the three canonical directions of the magnetic field. In the spin-unrestricted case, each U -matrix falls into distinct spin-up and spin-down sets as long as the perturbation is spin-independent which is

the case for the OZ operator. Finally, one inserts the solution into the second-order part of Eq. 16 and makes use of the fact that within a given multiplicity all matrix elements of the SOC operator can be written in terms of the $\hat{s}_z$ component only. One arrives at:

$$\begin{aligned} \left. \frac{\partial^2 E}{\partial B_\lambda \partial S_\kappa} \right|_{\lambda=0; \kappa=0} &= \frac{1}{2} \beta \sum_{\mu\nu} \frac{\partial P_{\mu\nu}^{(\alpha)}}{\partial B_\lambda} \langle \varphi_\mu | \hat{z}_\kappa^{\text{SOMF}} | \varphi_\nu \rangle \\ &\quad - \frac{1}{2} \beta \sum_{\mu\nu} \frac{\partial P_{\mu\nu}^{(\beta)}}{\partial B_\lambda} \langle \varphi_\mu | \hat{z}_\kappa^{\text{SOMF}} | \varphi_\nu \rangle \\ &= \frac{1}{2} \beta \sum_{\mu\nu} \frac{\partial P_{\mu\nu}^{(\alpha-\beta)}}{\partial B_\lambda} \langle \varphi_\mu | \hat{z}_\kappa^{\text{SOMF}} | \varphi_\nu \rangle. \end{aligned} \quad (36)$$

Thus, spin-dependent perturbations lead to the response spin-density matrix instead of the response electron density matrix which is present in the electric-field case. Comparing these matrix elements with the matrix of the SH for spin S one finds:

$$g_{\lambda\kappa}^{(\text{OZ/SOC})} = \frac{1}{2S} \sum_{\mu\nu} \frac{\partial P_{\mu\nu}^{(\alpha-\beta)}}{\partial B_\lambda} \langle \varphi_\mu | \hat{z}_\kappa^{\text{SOMF}} | \varphi_\nu \rangle \quad (37)$$

(note that both, the response spin density and the SOC matrix elements are purely imaginary. In an actual implementation one operates with skew symmetric real-valued matrices in which case a factor of $i^2 = -1$ arises in front of the g -tensor equation).

In a similar way all terms in the SH can be calculated by analytic derivative techniques. Details of the actually involved procedures may be found in the literature. Importantly, Eq. 16 is not only valid in the case of a SCF ground state description but is of much wider applicability. In the case of correlated *ab initio* methods, the equations to be solved in order to determine the effective density and its response naturally become more complicated than the relatively simple coupled-perturbed SCF equations sketched in this work. The general line of thought is, however, identical and merely the “mechanics” of the calculation become more involved. This statement holds for any nonrelativistic variational or nonvariational methods (e.g. for configuration interaction, Möller–Plesset or coupled-cluster methods) but does not apply to quasi-relativistic (two-component) methods which treat the SOC together with the electron-electron repulsion. For such methods, first-order perturbation theory is usually sufficient to calculate SH parameters. However, as long as these methods are based on single determinants, severe problems with internal consistency and symmetry breakings occur the discussion of which is outside the scope of this work. Note also, that these methods are also about an order of magnitude more expensive than the nonrelativistic one-component methods and it may therefore be speculated that they will be reserved for

rather special applications in which it turns out to be mandatory to treat the SOC to infinite order.

2.3

Mössbauer Spectroscopy

An important spectroscopy in a bioinorganic chemistry context is Mössbauer spectroscopy. It is a method which selectively probes the charge and spin distribution around iron centers [27, 28]. In zero-applied magnetic field, the Mössbauer spectrum of an isolated iron center consists of a single pair of lines which are called quadrupole doublet. The center of the pair is referred to as “isomer shift” and the splitting is known as “quadrupole splitting”. Both properties are readily calculated if the electron density of the ground state is known. The quadrupole splitting of the iron nucleus is given by

$$\Delta E_Q = \frac{1}{2} e_0^2 Q q \sqrt{1 + \frac{1}{3} \eta^2}, \quad (38)$$

where $q = V_{\max}$ is the numerically largest eigenvalue of the symmetric field-gradient tensor V , $\eta = [V_{\text{mid}} - V_{\text{min}}]/V_{\max}$ is the asymmetry parameter and e_0 is the elementary charge. The proportionality constant Q is the quadrupole moment of the ^{57}Fe nucleus in the $I = 3/2$ excited state which is not known precisely. In this study a value of 0.16 barn has been assumed in keeping with a number of previous calculations. This value was also verified as the best choice in a recent calibration study on a series of iron complexes [29]. The electric field gradient tensor is $(\mu, \nu = x, y, z)$:

$$V_{\mu\nu}^{(A)} = \sum_{\kappa, \tau} P_{\kappa\tau} \langle \varphi_{\kappa} | r_A^{-5} (r_A^2 \delta_{\mu\nu} - 3 r_{A;\mu} r_{A;\nu}) | \varphi_{\tau} \rangle, \quad (39)$$

where $\mathbf{r}_A = \mathbf{r} - \mathbf{R}_A$ is the position of the electron relative to nucleus A at position $\mathbf{R}_A$, $\mathbf{P}$ is the density matrix from a ground-state DFT calculation with the basis set $\{\varphi\}$.

The Mössbauer isomer shift δ can be determined to good accuracy from ground-state DFT calculations using the calibration procedure outlined in [30]. Since the isomer shift is known from basic principles to be directly proportional to the electron density at the iron nucleus [27], it can be determined from a linear regression analysis, e.g. a plot of the calculated electron density at the nucleus versus the measured isomer shift. For our method of choice, the B3LYP functional together with the flexible core properties basis set CP(PPP) with an extended set of polarization functions, it can be computed from the regression [30]

$$\delta = 6.5481 - 0.3666[\rho(0) - 11800] \quad \text{in mm s}^{-1}, \quad (40)$$

where $\rho(0)$ is the nonrelativistically calculated electron density at the iron nucleus. Indeed, the calculations in Fig. 2 show a very good correlation between the two quantities with a regression coefficient of 0.986.

It is important to note, however, that the calculated electron density at the iron nucleus is far from the exact density. First of all, using Gaussian basis functions, the basis set limit is difficult to approach since Gaussians have the wrong behavior close to the nucleus. The nuclear cusp can be reasonably well modelled with very steep basis functions. To get very close to the basis set limit is nevertheless almost impossible given the finite accuracy with which numbers can be represented by a digital computer. It has been our strategy to use a basis set with sufficient flexibility in the core region to allow for the distortion of the inner *s*-, and *p*-shells and have termed this basis set CP(PPP). Secondly, different density functionals lead to grossly different electron densities at the nucleus and therefore a given calibration is only valid for a given functional. This large functional dependence may be expected since present day functionals do not only have the wrong asymptotic behavior in the long range but also in the short range close to the nucleus. Third, it is important to ensure that the numerical integration accuracy is sufficient to lead to converged results in the presence of steep, uncontracted basis functions in the core region. Fourth, the neglect of scalar relativistic effects is serious and leads to an underestimation of the electron density of more than 3000 au^{-3} . This seems unacceptable since the variation of the electron density over the range of molecules studied is only $\sim 7 \text{ au}^{-3}$. Recent scalar relativistic calculations using the ZORA method have greatly improved on the value of $\rho(0)$ but the quality of the regression remained unchanged [29, 31]. The reason why

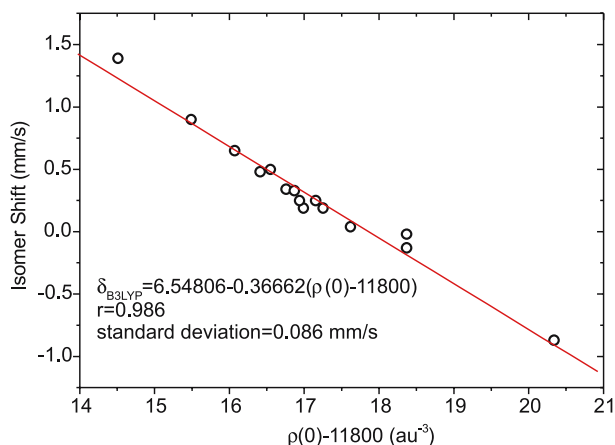


Fig. 2 The linear relationship between calculated electron densities at the iron nuclei $\rho(0)$ and measured isomer shifts of 15 different complexes. The obtained parameters can be used in further application studies employing the same basis set CP(PPP) and functional B3LYP

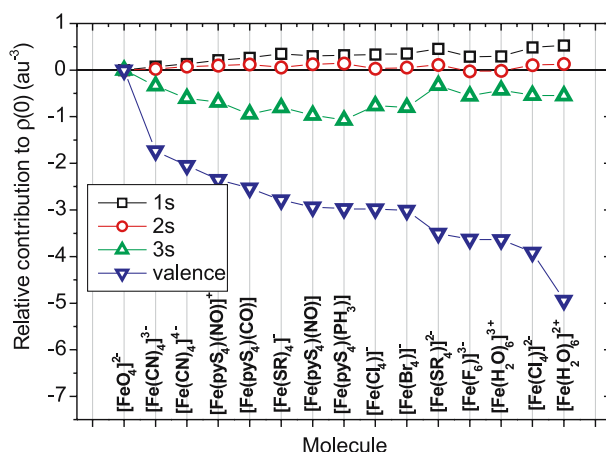


Fig. 3 Relative contributions to the electron densities at the iron nuclei $\rho(0)$. The figure clearly shows that changes in $\rho(0)$ can be attributed to changes in the 3s and valence shells, while the contributions from the 1s and 2s shells are almost constant, independent of the charge and spin states of the systems

the correlation with experiment is still as good as has been observed in the nonrelativistic calculations is shown in Fig. 3. It is demonstrated that the contributions of the 1s- and 2s-orbitals to the total electron density at the iron nucleus is very constant over the range of bonding environments typically met in iron containing complexes. Thus, very large but very systematic errors cancel out in the linear regression analysis and ensure that DFT is an accurate and reliable tool in the study of Mössbauer isomer shifts provided that the results are standardized with respect to functional and basis set. Given these precautions, the regression does not depend on the charge or spin state of the iron and has proven in many applications to be trustworthy [29–33].

2.4

Optical Spectroscopy

Optical spectroscopy will only be touched upon in a cursory manner since it will only be briefly addressed in the application section. The application of DFT methods to excited states is difficult at first glance since the Hohenberg–Kohn theorems only apply to the ground state. However, response theory once more offers a solution to the problem. To this end, one needs to extend the formalism developed above to the case of time-dependent oscillatory perturbations such as the oscillatory electric field of a monochromatic, linearly polarized light beam. The second-order response of the system to such a perturbation is the frequency dependent polarizability. This quantity has poles at the positions where the frequency of the external field equals the excitation frequency of the system. In practice, one finds these poles by the

diagonalization of the response matrix. The random-phase approximation leads to a generalized eigenvalue problem in which both, the electric and magnetic Hessian matrices appear. Solution of this generalized eigenvalue problem then directly yields the excitation frequencies. From the eigenvectors one can calculate the electric and magnetic transition densities and therefore various transition properties such as oscillator strengths and rotation strengths of CD spectra. An interesting approximation is provided by the so-called Tamm–Dancoff approximation (TDA) which amounts to the diagonalization of the electric Hessian $A^{(E)}$ [34]. The TD-DFT/TDA equations very much resemble the configuration interaction problem with single excitations only (CIS) and can be efficiently solved at the similar computational expense. Under most circumstances TD-DFT/TDA, in particular in combination with hybrid density functionals, is substantially more accurate than CIS. However, one should be aware of a number of limitations of this methodology. In particular, TD-DFT fails for charge transfer and covalent-to-ionic transitions, does not describe multiplet effects correctly, does not produce any double excitation and does not produce pure spin-eigenfunctions in the spin-polarized case. If one is aware of these limitations good use can still be made of this method since it is applicable to large molecules and analytic gradients for geometry optimizations on excited states can be implemented with reasonable effort in a computationally efficient way [35–37]. For more details, the reader is referred to a recent review by Head-Gordon and Dreuw [38].

A radically different approach to excited states is offered by multireference *ab initio* methods. These usually start with a multiconfigurational self-consistent field (MCSCF) description of the desired states followed by the treatment of dynamic correlation effects with configuration interaction (MRCI) or perturbation theoretical (MRPT) methods. A particularly successful MRPT variant is the CASPT2 method developed and widely used by Roos and co-workers [39]. A computationally affordable variant of MRCI is the spectroscopy oriented configuration interaction (SORCI) procedure developed in this laboratory and successfully used in recent studies [40]. These methods are reviewed in some detail elsewhere [16].

2.5

Environment Effects

Bioinorganic active sites and model complexes are in the vast majority of cases not studied in the gas phase but in condensed media. Consequently, environment effects are always present in the experimental data and should be modelled at some level of sophistication in order to obtain reliable results.

Continuum dielectric approaches are a very convenient way to model long-range interactions between the system under study and a continuum, which

is characterized by its dielectric constant [41, 42]. In these models, the solute is placed in a cavity and interacts with the unstructured, but polarizable surrounding. These methods do not increase the degrees of freedom of the calculations, which makes them routinely applicable also for larger systems. Furthermore, it is possible to perform geometry optimizations of the solute in the solvent. Within the last years interesting studies appeared that have been devoted to the calculation of solvent shifts of spectroscopic parameters employing continuum models, including ABS [43–45], NMR [46–48], and EPR data [49–55]. It is clear that these approaches are especially well suited for the study of solute–solvent interactions, but they also have been used to study protein active sites [56]. Nevertheless, it must be mentioned that continuum models are not able to describe short-range interactions like solvent to metal ion coordination or hydrogen bonding effects.

A more rigorous way to model environment effects of protein active sites is the use of mixed quantum mechanical/molecular mechanical (QM/MM) approaches [57, 58]. Here, the cofactor of interest is modeled by the QM method, while the protein backbone is described by the much faster MM approach. In consequence, these methods allow the treatment of very large systems—entire proteins can be considered in this way. QM/MM studies that include spectroscopic applications are available in the literature to an increasing degree [18, 46, 59, 60].

More approximate than QM/MM methods is the supermolecule approach, in which the active site or solute and a rather small part of the surrounding are treated on equal footing employing QM methods. While it was shown that this method can give accurate spectroscopic parameters, for example for radicals in protic solution [50, 61], the number of solvent molecules that can be treated is limited and conformational problems can occur.

Finally, *abinitio* molecular dynamics (MD) simulations are feasible, but can be computationally fairly expensive. However, the results for the calculation of spectroscopic parameters are very promising [62].

3

Recent Case Studies

3.1

High Precision Studies of Organic Radicals

An accurate description of the electronic structures of organic radicals with a single unpaired electron is much easier to achieve than for transition metal clusters. Molecular *g*-tensors, hyperfine coupling constants (HFCs) and nuclear quadrupole coupling constants (NQC) can be straightforwardly calculated. The high accuracy that can be achieved for these systems allows for the study of very subtle details like small conformational changes or solvent

effects. This is demonstrated here by discussing three different case studies, including investigations on phenoxy radicals, benzosemiquinone radical anion–solvent interactions, and the cysteine thiyl radical conformational changes. In all cases it was possible to draw important conclusions on the molecular and electronic structures of the systems by comparing calculated and measured spectroscopic parameters. It is however, extremely important to be aware about the achievable accuracy of the employed methods. It must be ensured that the investigated effects are larger than the relative errors in the calculations. Typically, this is clarified by benchmark calculations on similar, but experimentally well-understood systems.

3.1.1

Molecular *g*-Tensors of Phenoxy Radicals

A recent study has underlined the high value of combining experimental and theoretical work on molecular *g*-tensors [63]. Three different phenoxy radicals, hydrogen bonded to an imidazolium, were investigated. These systems serve as models of the Tyr_D tyrosyl radical in photosystem II. The one electron oxidation of the parent neutral compounds to yield the radical cations may either take place via a proton-coupled electron transfer (PCET), or via a pure electron transfer (ET) as displayed in Fig. 4. In order to discriminate between the possible reaction products, a high-field (W-band) EPR study was undertaken. It was possible in these measurements to resolve the *g*-tensor anisotropies. The interpretation of the obtained data was guided by DFT calculations. Complete geometry optimizations were performed employing the BP functional [64–66] in combination with the TZVP basis set [67]. The *g*-tensors were calculated from the coupled perturbed Kohn–Sham equations [68] employing the B3LYP hybrid functional [69, 70] and the EPR-II basis set [71]. Furthermore, the COSMO dielectric continuum model was employed in the geometry optimizations and single-point calculations [72]. Calculations were performed on both conceivable products, the “OH” and “NH” species shown in Fig. 4. The analysis of the data al-

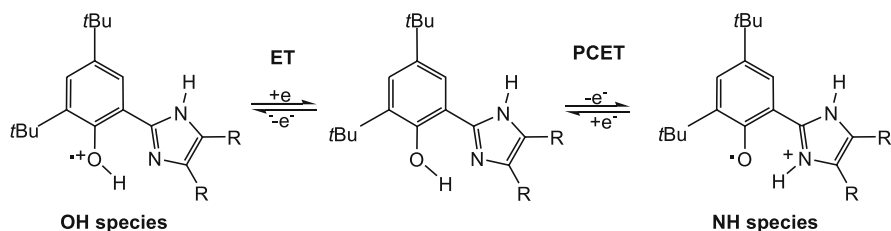


Fig. 4 One electron oxidation of the charge neutral phenol species can take place via proton-coupled electron transfer (PCET) or electron transfer (ET). Here, R = Ph is considered as an example

lowed only one conclusion: the radical that was found in the experiments must be the NH form. Especially the g_x ($\equiv g_{\max}$) value calculated for the OH form (2.0034) deviates by more than 3000 ppm from the experimental value of $g_x = 2.0065$. This deviation is about one order of magnitude larger than the typical accuracy of g -tensor predictions for such radicals. Furthermore, the experimental g_x -value is almost perfectly reproduced in the calculations on the NH form ($g_x = 2.0066$). In consequence, the calculations provided strong evidence that the NH-form has been formed under the experimental conditions and that the one electron oxidation must occur via a PCET mechanism.

3.1.2

Quinone Radical Anion–Solvent Interactions

Magnetic properties can also be used as very valuable parameters to obtain information on the surrounding of a given open shell species. The effect of H-bonding, certainly one of the most important intermolecular interactions in biochemistry, was recently studied by means of the parabenzoquinone radical anion in water [61]. Different model systems were tested, employing the radical anion and one, two, four or 20 solvent molecules within the supermolecule approach. Complete geometry optimizations were performed with the B3LYP hybrid functional in combination with a double- ζ basis set, augmented by polarization and diffuse functions. Again, the spectroscopic parameters were obtained from B3LYP calculations with the well-proven EPR-II basis set, including g -values, as well as HFCs and NQCs of deuterium solvent nuclei that form hydrogen bonds to the carbonyl oxygen atoms of the quinone radical anion. This type of interaction was also found in biological systems [73].

In general, better results were obtained with increasing size of the model systems. This shows that the second solvation shell has a non-negligible influence on the calculated properties. Through the complete calculation of all relevant SH parameters (g -values, HFCs, NQCs and Euler angles) it was possible to simulate the ENDOR spectra from DFT (Fig. 5). Only the line width parameters, which are not available from the theoretical work, were adjusted. From the comparison with the experimental spectra two important conclusions could be drawn: (i) Present-day DFT calculations can be used to generate high-quality initial guess parameters for otherwise underdetermined spectral simulations, and (ii) the calculated geometry of the system must represent a realistic scenario. The second point is especially important for the deduction of H-bond lengths. While H-bond angles can be obtained from the experimental Euler angles, the hydrogen bond lengths have to be estimated from the dipolar HFCs or NQCs of a bridging deuterium employing crude approximations or fit equations. Hence, it is a tempting idea to take H-bond lengths directly from DFT calculations. This requires, however,

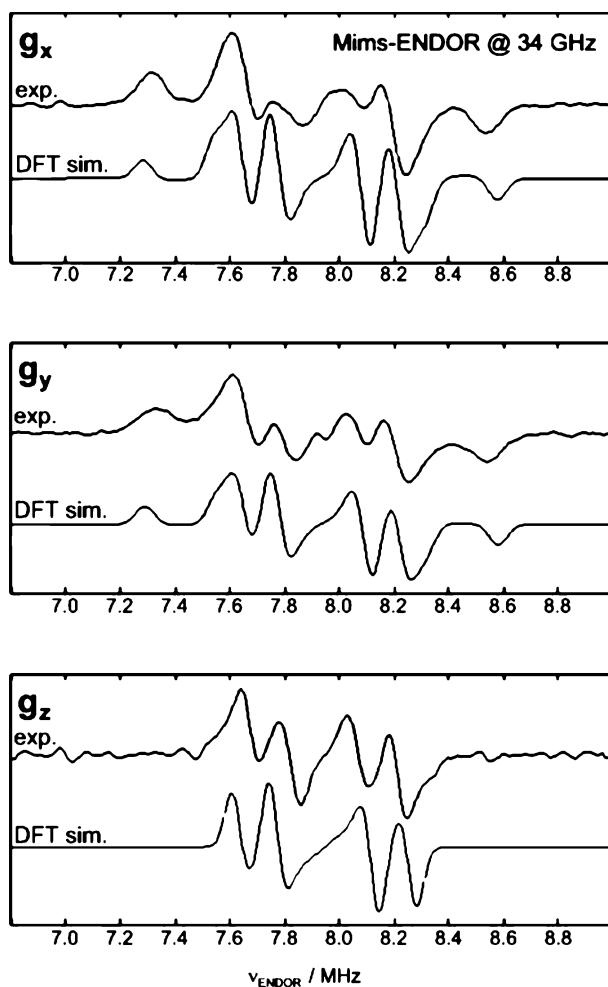


Fig. 5 Simulation of ^2H ENDOR spectra from calculated HFCs, NQCs, g -values and Euler angles for the benzosemiquinone radical anion in water (model system with 20 solvent molecules). Comparison with Q-band pulse ^2H ENDOR (Mims) absorption spectra (first derivative)

that a consistent set of calculated spectroscopic parameters is available and in good agreement with all experimental data.

3.1.3

The Cysteine Radical and Its Unusual g -Tensor Anisotropy

Our next example is a theoretical study of the cysteine thiyl radical [74]. Such sulfur centred radicals play an important role in biochemistry. They are met,

for instance, as catalytic intermediates in all three classes of ribonucleotide reductases. The large spin-orbit coupling constant of sulfur in comparison to oxygen, nitrogen or carbon leads to unique spectroscopic features of this species. One example is the noticeable anisotropy of the molecular g -tensor with principle values that resemble transition-metal complexes, rather than organic radicals.

The aim of the study of van Gastel et al. was to obtain a correlation of g -values and HFCs of the thiyl radical with its structure and the polarity of its surrounding. The carboxyl and amino groups were modeled in their charge neutral form since calculations on zwitterionic model systems produce artifacts in gas-phase computations. The effect of a protic environment was included in this work by additional calculations, where one or two water molecules that form hydrogen bonds to the cysteine sulfur atom were added to the model system. Again, geometry optimizations were performed employing the B3LYP functional in combination with triple- ζ basis sets, augmented by polarization and diffuse functions. The g -values were obtained from coupled perturbed SCF theory as discussed above [68]. Furthermore, ^1H and ^{33}S HFCs were calculated. All these properties were obtained from B3LYP calculations employing the IGLO-III basis set [75]. In addition, excitation energies were computed employing the SORCI approach [40].

Three different conformers with respect to rotations around the $\text{C}^\alpha - \text{C}^\beta$ bond of the radical were found in the geometry optimizations and characterized as local minima with the help of frequency calculations (Fig. 6). The obtained structures are close in energy with differences not larger than 630 cm^{-1} . Transition energies to the first excited state are given for all isomers in Table 1. The results indicate that the involved orbitals are almost degenerated in energy for the conformer 2, while distinctly larger excitation energies were obtained for the 1st and 3rd minimum. Calculated g -values are also given in Table 1. A large g_x shift was found for all conformers. It is mainly determined by electronic transitions from the SOMO-1 to the SOMO orbitals and therefore nicely correlates with the calculated excitation energies.

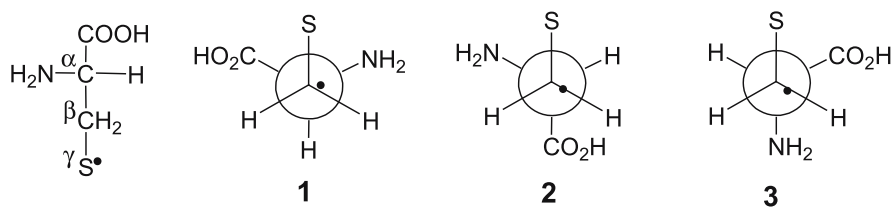


Fig. 6 Fischer projection of the cysteine thiyl radical (*left*). DFT geometry optimizations yielded three different conformers 1–3 with respect to rotations around the $\text{C}^\alpha - \text{C}^\beta$ bond. While their energies were found to be similar within 630 cm^{-1} , noticeably different g -tensors were calculated for the conformers. A molecular coordinate system was chosen with g_x parallel to the $\text{C}^\beta - \text{S}^\gamma$ bond and with g_z located in the $\text{C}^\alpha - \text{C}^\beta - \text{S}^\gamma$ plane

In the next step, a polar environment was modeled by adding water molecules to the system. Hydrogen bonds were found between the water protons and the sulfur atom with $S \cdots H$ distances between 2.39 and 2.60 Å. For all three conformers, every additional water molecule decreased the calculated g_x value. This effect can be traced back to a lifting of the degeneracy of the energies of the sulfur $3p_z$ (SOMO-1) and $3p_y$ orbitals (SOMO), which increases the energy denominator. The results obtained with two water molecules are also given in Table 1.

Hyperfine coupling constants were calculated in this study for the two β -protons and the ^{33}S nucleus. The ^1H β -HFCs were found to be mainly determined by the conformation of the cysteine radical, while the influence of the number of solvent molecules on the calculated HFCs was found to be extremely small. Furthermore, a large isotropic ^{33}S HFC of about 40 MHz was obtained with variations of about 10 MHz depending on the model system.

The experimental g -values of various thiyl radicals were found to strongly depend on the chosen species. While values between 2.10 and 2.49 were in general reported for g_x , numbers in the range of $g_x = 2.15$ – 2.25 were obtained for the majority of investigated thiyl radicals. These data were nicely reproduced in the study of van Gastel et al. [74]. It was concluded that radicals with g_x values of about 2.15 are present in the first minimum with a H-bonded sulfur atom, while the experimental g_x values of about 2.25 were traced back to the same conformer without H-bonds. Such findings can be very valuable with respect to the wide variety of spectroscopic parameters that have been found so far for cysteine radicals.

Table 1 Calculated properties of the three conformers 1–3 of the cysteine thiyl radical from Fig. 6. Excitation energies $E_{0 \rightarrow 1}$ (cm^{-1}) from the ground states to the first excited states were calculated from SORCI. Molecular g -values were obtained from DFT (second order) employing the pure radicals (gas phase) and the radicals ligated by two water molecules that form H-bonds to the sulfur atom ($2\text{H}_2\text{O}$). Third order g -values are given for the model systems without water molecules in parenthesis. They were found to be necessary to reproduce the negative g_z shifts from the experiments

		1	2	3
gas phase	$E_{0 \rightarrow 1}$	1732	1083	3429
	g_x	2.2649 (2.2649)	2.3692 (2.3692)	2.1680 (2.1680)
	g_y	2.0215 (2.0043)	2.0214 (1.9877)	2.0161 (2.0092)
	g_z	2.0024 (1.9852)	2.0023 (1.9686)	2.0023 (1.9954)
$2\text{H}_2\text{O}$	g_x	2.1407	2.1952	2.0928
	g_y	2.0212	2.0229	2.0156
	g_z	2.0024	2.0025	2.0023

3.2

The Spectroscopy of Blue Copper Proteins

Transition metal active sites are the chief ingredient of bioinorganic chemistry. On one hand it is a highly interesting task to explore the properties of these systems, but on the other hand it is still an outstanding challenge to accurately calculate the spectroscopic properties of transition-metal complexes. In order to arrive at a correct microscopic description of their physical properties one needs to be able to deal with near-degeneracy situations, complicated multiplet problems, spin-state energetics, excited states of vastly different characters, spin-polarization, large spin-orbit coupling effects and interfragment spin-couplings simultaneously. In DFT one encounters problems in the balanced description of the partially ionic, partially covalent metal-ligand bonds, the correct balance between spin polarization and spin contamination, the construction of spin-eigenfunction and multiplet states and the accurate calculation of excitation energies. In addition, an appropriate approximation of the spin-orbit coupling operator and adequate, often large, basis sets are required. By contrast, configuration interaction-based methods suffer from the high computational effort and from the fact that the number of spectroscopic properties that can be calculated with these approaches is very limited in most of the available quantum chemical programs. However, in principle, these methods offer a solution to all of the problems referred to above which is a significant motivation for their further development.

The influence of the protein surrounding on magnetic, electric and optical spectroscopic parameters of metalloproteins was recently investigated in a QM/MM study on copper plastocyanin [18]. In these proteins, the metal is coordinated to two histidines, a cysteine and a methionine amino acid residue with an outstanding long Cu – S^{Met} bond of ~ 2.9 Å. This system has been intensively studied over several decades by experimental and theoretical means and its electronic structure and spectroscopy is well understood. Hence, it is an ideal benchmark system for a theoretical study.

Starting with the pdb file of the protein, a solvent box of water molecules was added to the protein in the theoretical work and a MD simulation was performed. Subsequently, QM/MM geometry optimizations were carried out. The obtained structures were used for single point calculations on the active site (a) in the gas phase, (b) in a continuum, and (c) in the protein environment employing the MM point charges. Furthermore, nonrelativistic and scalar relativistic results were compared. Best agreement between calculated and measured *g*-tensors, HFCs and NQCs was obtained for the scalar relativistic calculation in the protein environment, which was also the highest level of theory employed in this study. The influence of the surrounding on the calculated spectroscopic properties was found to be quite substantial and can alter the properties by up to 30%. Interestingly, this effect can be at-

tributed in similar extents to the protein backbone atoms and to the water cage molecules.

Much effort of this work was devoted to the calculation of the Cu HFCs. The Fermi-Contact term was underestimated, which is a typical result for transition-metal ions. A comparison with the results obtained for other metals indicates a correlation with the number of unpaired electrons. The spin-dipolar contribution is expected to be rather accurately predicted. The huge spin-orbit contribution was also distinctly underestimated, which was attributed to the fact that the linear response of the density functionals is too stiff with respect to the external perturbations. Such problems of present-day density functionals express themselves in the errors of many properties like g -tensors, second-order contributions to the HFCs, or transition energies. Hence, a very careful handling of the computed results is necessary.

3.3

Mononuclear Iron Complexes – from Outstanding Model Systems to Fascinating Enzymes

This paragraph is dedicated to high-valent Fe(IV) complexes. Such systems have been known for a very long time to be highly important in biochemistry in the form of the oxo-complexes $(\text{FeO})^{2+}$. They were characterized or postulated as transient intermediates in a wide variety of heme- and nonheme iron enzymes. Because of their transient nature, spectroscopy is the method of choice for studying such high-valent systems. An additional complication in interpreting the spectroscopic data obtained for Fe(IV) sites is the fact that until recently there have been only a limited number of model systems available that could be used for calibration and comparison [76]. Furthermore, the spin states of $(\text{FeO})^{2+}$ systems must be determined: While the $(\text{FeO})^{2+}$ species found in native systems like taurine/ α -ketoglutarate dioxygenase (TauD) were proposed to be in the $S = 2$ state, only model systems with $S = 1$ ground states have been reported so far in the literature. The only exception we are aware of is the $[\text{FeO}(\text{H}_2\text{O})_5]^{2+}$ model compound studied by Pestovsky et al. [77] as well as a non-oxo compound by Collins and co-workers [78].

Hence, it is helpful for the interpretation of the experimental data to know, how the spin states influence the spectroscopic parameters and which geometric features determine the ground-state spin. This is a typical problem that can be solved by quantum chemical calculations, where the system can be forced into the desired state. Such a study was performed on the $[\text{Fe(IV)O}(\text{NH}_3)_5]^{2+}$ system, a prototypical but hypothetical $(\text{FeO})^{2+}$ model system [79]. Ammonia ligands were chosen to obtain a model complex that is small enough to allow the use of CI-based methods but that, at least crudely, resembles the cyclam ligands of some of the real model systems. The iron, oxygen and nitrogen atoms were described by basis sets of triple- ζ quality in the CI calculations and DFT geometry optimizations. Spectroscopic par-

ameters were obtained employing the core property basis set CP(PPP) for iron [30], and the SOMF operator. Results from DFT (BP86 and B3LYP) were compared with the results from CI (SORCI based on CASSCF reference wave functions).

Starting with the results from DFT geometry optimizations for the complex in the $S = 1$ and $S = 2$ states, the $S = 1$ state was found to be slightly more stable by 2 (B3LYP) and 6 kcal/mol (BP). SORCI single-point calculations at the DFT geometries gave the opposite result: the ground state was assigned to the $S = 2$ state, being 10.8 kcal/mol lower in energy than the $S = 1$ state. This shows that it is dangerous to draw conclusions only on the basis of energy calculations. Electron and spin populations were found to be similar in the DFT and *ab initio* calculations. By varying the iron–ligand distances it was demonstrated that the equatorial ligand field determines the spin state of the system: weaker ligands favor the $S = 2$ spin state.

In the next step, the spectroscopic parameters of the two spin states were compared. Starting with the zero-field splitting parameters, large and positive D values were found for both spin states employing DFT (Table 2). From SORCI, a value of $D = +10.1 \text{ cm}^{-1}$ was calculated for the more stable $S = 2$ state, which is a factor of ~ 2 larger than the BP86 result. It is also expected to be more accurate based on model calculations on small molecules and the more realistic description of the d^4 multiplet structure in the *ab initio* approach.

The calculated g -values were found to be rather close to the free electron g -values and are therefore not well suited for discriminating between the two spin states. The iron hyperfine data revealed an interesting difference between the $S = 1$ and $S = 2$ states: While a small, negative $A_{||}$ value of -4.9 MHz was

Table 2 Comparison of calculated spectroscopic parameters (B3LYP) of the $[\text{Fe(IV)O}(\text{NH}_3)_5]^{2+}$ model system in the $S = 1$ and $S = 2$ states: Zero-field splitting parameters D and E/D , g -values, hyperfine coupling constants A , and Mössbauer parameters δ_{MB} , ΔE_{q} , and η . Noticeably large differences between the spin states were found for $A_{||}$ (^{57}Fe), and the Mössbauer parameters

	$S = 1$	$S = 2$
$D \text{ (cm}^{-1}\text{)}$	5.20	4.25
E/D	0.00	0.00
$g_{\perp}$	2.0181	2.0139
$g_{ }$	2.0075	2.0061
$A_{\perp} \text{ (}^{57}\text{Fe) (MHz)}$	-30.1	-29.6
$A_{ } \text{ (}^{57}\text{Fe) (MHz)}$	-4.9	-50.9
$\delta_{\text{MB}} \text{ (mm/s)}$	0.133	0.292
$\Delta E_{\text{q}} \text{ (mm/s)}$	0.54	-1.21
η	0.07	0.03

predicted for the $S = 1$ state, a larger HFC of $A_{||} = -50.9$ MHz was found for the $S = 2$ state employing the B3LYP hybrid functional. It is not expected that the calculated numbers are very accurate, but the large qualitative difference should be solid. Hence, the iron HFC might be an important probe for experimental studies on the spin states of $(\text{FeO})^{2+}$ systems.

Considering the calculated MB parameters, it was found that the isomer shift of the $S = 2$ species is similar to that of many low- or intermediate spin Fe(III) isomers. A distinctly smaller value was obtained for the $S = 1$ species and is typical for such systems which generally show isomer shifts around 0 mm/s. The larger value for the $S = 2$ state was mainly attributed to the strongly increased equatorial bond distances and is similar to the value found for the TauD reaction intermediate for which this spin state was proposed. The computed quadrupole splittings strongly depend on the symmetry of the systems. The computed change in the sign from positive quadrupole splittings for $S = 1$ to negative ΔE_Q values for $S = 2$ was also found in experimental work: While a positive value was reported from Rohde et al. for a $S = 1$ model system (Que's model compound), a negative quadrupole splitting was found by Price et al. for the $S = 2$ intermediate from TauD (Fig. 7) [80, 81]. Finally, the absorption spectra were calculated employing TD-DFT and SORCI with a small CAS(4,5) reference space. It was found by both methods that the $S = 1$ systems features $d-d$ transitions of lower energy than the system in the $S = 2$ state, but the ordering of the excited states and the predicted transition energies are very different. Again, more accurate results are expected from the SORCI approach.

We now switch to biological systems and start with the high-valent oxo-ferryl intermediate of TauD. These enzymes couple the decarboxylation of α -ketoglutarate with the hydroxylation of an unactivated C–H bond of taurine and play therefore an important role in diverse biological processes [82]. The transient $(\text{FeO})^{2+}$ species was recently well characterized by Mössbauer spectroscopy, but its exact molecular structure is still not known [81, 83]. The proposed structure of this intermediate, which is called J, is given in Fig. 7. The metal is bound to a His₂–Asp facial triad and a succinate molecule. Fur-

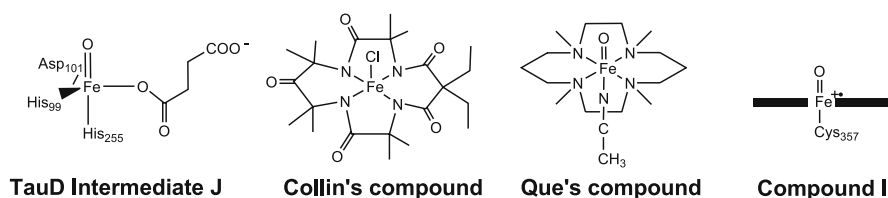


Fig. 7 The proposed molecular structure of the oxo-ferryl intermediate J of TauD, the structures of the Fe(IV) model compounds of Collins and Que, and the structure of the Fe(IV)-oxo porphyrin radical cation intermediate from P450 (Compound I). The *thick bar* represents the heme plane

thermore, it is strongly assumed that this species is in the quintet state, which is supported by the calculations performed by Krebs et al. [83]: The measured Mössbauer parameters δ and ΔE_Q were only well reproduced in calculations on a trigonal-bipyramidal model system with an axial oxo ligand in the quintet state (Table 3). The accuracy of the employed methods was judged by test calculations on the well known $(\text{FeO})^{2+}$ model compounds from Fig. 7. It is interesting to compare the MB data of intermediate J with those calculated for the model system $[\text{Fe(IV)O}(\text{NH}_3)_5]^{2+}$ (Table 2): The data obtained for the $S = 2$ state of $[\text{Fe(IV)O}(\text{NH}_3)_5]^{2+}$ are in much better agreement with the MB parameters of the TauD oxo-ferryl intermediate than the data of the $S = 1$ state.

While we have discussed so far only nonheme iron systems, iron porphyrins are certainly of outstanding relevance in bioinorganic chemistry and we will now focus on a study of the high valent ferryl-oxo porphyrin radical cation of cytochrome P450 as a representative case for this class (Fig. 7) [84]. The work that we have picked out includes also an extensive modeling of the protein environment with QM/MM methods. It is divided in three parts that deal with (a) small model systems containing the $(\text{FeO})^{2+}$ motif, (b) the interaction between a porphyrin radical and the $(\text{FeO})^{2+}$ moiety, and (c) the treatment of the protein active site with and without surrounding. The use of small reference systems made the application of *ab initio* methods possible, while the workhorse was DFT, again.

DFT was employed for geometry optimizations, the calculation of magnetic hyperfine tensors, g -tensor and MB properties, while the *ab initio* methods were used for the calculation of excitation energies and exchange coupling constants. The small, symmetric $[\text{Fe(O)}(\text{NH}_3)_4(\text{H}_2\text{O})]^{2+}$ model system (C_{4v}) was used to analyze the character of the Fe – O bond and the MO diagram in detail, but also optical spectra, g -tensors and zero-field splittings of this species were provided. Que's model compound (Fig. 7) was also

Table 3 Comparison of measured and calculated Mössbauer parameters of the high-valent iron-oxo intermediate from TauD. Calculated data from Krebs et al. [83]. Experimental results from Price et al. (TauD) [81], Kostka et al. (Collins model compound) [95], and Rohde et al. (Que's model compound) [80]

		δ (mm/s)	ΔE_Q (mm/s)
TauD intermediate J	Exptl.	+0.31	–0.88
	Calcd.	+0.22	–1.0
Collins model compound	Exptl.	–0.04	–0.89
	Calcd.	–0.08	–0.76
Que's model compound	Exptl.	+0.17	+1.24
	Calcd.	+0.11	+1.2

employed in this study to calibrate the computational methods. A comparison of calculated geometrical parameters with those from X-ray evidenced again the suitability of the BP86 density functional for the optimization of molecular structures, which is very efficient in combination with the resolution of identity (RI) approximation [85, 86]. However, the typical overestimation of the weaker nitrogen–metal bond distances was also found in this work. Strong effort was dedicated to the question of the electronic ground state of both model compounds, which can be either of triplet or quintet symmetry.

The calculated MB properties of Que's model compound were found to be very sensitive with respect to the employed geometries. This indicates that the developed fit parameters for the calculation of isomer shifts depend not only strongly on the chosen functionals and basis sets, but also on the underlying geometries. Typically, the fit parameters were developed employing molecular structures from BP86 optimizations, or from X-ray crystallography.

The calculated ^{57}Fe and ^{17}O hyperfine data were found to be very stable with respect to different geometries. Axially symmetric tensors were obtained in both cases. Provided that the strong underestimation of the Fermi-Contact term is corrected (e.g. by scaling [29]), good agreement between theory and experiment was found for the ^{57}Fe HFCs, while no experimental data are available for the oxo-group. For both model systems, small g -shifts were found, but very large D -values. From a comparison with the experimental data it became clear that the employed DFT methods tend to noticeably underestimate the *zero-field* splitting D .

The next challenge of this study was an appropriate description of the $S = 1$ FeO unit ferromagnetically coupled to an $S = 1/2$ porphyrin radical. An analysis of the spin populations has shown that the FeO moiety is only marginally affected by the porphyrin radical with spin populations mainly located at the methane bridges, and to a lesser extent at the pyrrole nitrogen atoms. It was possible to calculate accurate MB parameters and ^{57}Fe HFCs, and exchange coupling constants for this system.

The successful studies on these model compounds gave a solid base for the investigation of the Compound I. Employing the QM/MM methodology, a huge influence of the protein environment on the calculated data was found by comparing results that were obtained with and without QM/MM point charges. Especially, spin density is taken away from the sulfur atom of the axial cysteine ligand and a decrease of charge at the iron center was found if the protein environment was included. Since antiferromagnetic coupling was observed in this species, the doublet ground state was studied in addition to the quartet state employing broken symmetry DFT [87, 88]. The MB parameters were found to be very similar in both spin states. This indicates that the electron density at the metal center is very similar. Hyperfine data were analyzed in detail for ^{57}Fe , ^{17}O , ^{14}N and ^1H , which are sensitive probes for electronic and geometrical details of the system. Especially the ^1H superhy-

perfine coupling constants of the cysteine ligand noticeably depend on the conformation of the side chain.

The results of this work had significance beyond providing a set of calibrated theoretical parameters which could be used for the identification of compound I. A key finding was that the exchange coupling in compound I is expected to be small and negative while the zero-field splitting was large and positive. In this situation the notation of a pure spin state partially loses its meaning since at room temperature the system has three Kramers doublets that are almost equally populated and that are heavy mixtures of doublet and quartet components lowest in energy. Thus, all three doublets could be the starting point for further reactions. This has an important bearing for the concept of two-state reactivity in which the potential energy surfaces of the doublet and quartet states play a key role [89, 90]. A second key finding of this work was that there is a second manifold of sextet and quartet states low in energy which arise from spin coupling of the $S = 2$ state of the $(\text{FeO})^{2+}$ unit with the porphyrin/thiyl radical. That these states are low lying and could also be implicated in the observed reactivity is evident following the results of the model studies. In fact, shortly after the appearance of [84] this possibility was explored theoretically [91], and it was indeed found that the sextet/quartet channel is competitive along the reaction coordinate.

3.4

Characterization of an Elusive Fe(V) Species

While the highest conclusively identified oxidation state of iron in biochemistry is Fe(IV), the possibility that Fe(V) might exist and play a role has been proposed many times. However, chemical precedence was largely missing. In this respect, the synthesis and characterization of a genuine Fe(V) species presents an important step in this field. Wieghardt and co-workers initially succeeded to generate a Fe(V)-nitrido species in immobilized samples by photolysis of an Fe(III)-azide precursor (Fig. 8). The main evidence for the generation of a high-valent iron species came from the Mössbauer isomer shift of the photoproduct which was observed to be slightly negative [92]. The spin state of the photogenerated species was plausibly assigned as $S = 3/2$ based on the observations made for analogous isoelec-

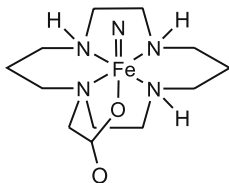


Fig. 8 Molecular structure of the $[\text{Fe}^{\text{V}}(\text{cyclam-ac})(\text{N})]^+$ complex

tronic Mn(IV) and Cr(III) species. The photochemistry was recently reinvestigated experimentally and theoretically with the goal to obtain more evidence that the proposed species is correct [93]. To this end, the Mössbauer measurements were complemented by X-ray absorption spectroscopy at the iron K-edge. These experiments revealed a pre-edge peak which was one eV higher in energy than observed in $(\text{FeO})^{2+}$ complexes which is consistent with the oxidation state Fe(V). Furthermore, analysis of the EXAFS region revealed that a new short 1.61 Å Fe–X (X = N,O) bond is formed. The experiments were supplemented by DFT calculations. Initially, the calculations were performed for the $[\text{Fe}(\text{cyclam-ac})(\text{N})]^+$ complex (cyclam = 1,4,8,11-tetrazacyclotetradecane) and a $S = 3/2$ ground state. Neither the calculated isomer shift, nor the quadrupole splitting nor the optimized bond distance of 1.75 Å were found to be in any reasonable agreement with the experimental data. This could be taken as evidence either against the formation of a high-valent Fe(V)-nitrido species or as a sign of a wrong ground state. All attempts to reproduce the experimental data with protonated species or other conceivable decay products remained without success. Therefore, the $S = 1/2$ spin state was investigated theoretically. It quickly became evident that not only was the energy of the $S = 1/2$ state much lower than that of the $S = 3/2$ state but also the calculated Fe–N distance, quadrupole splitting and isomer shift agreed very well with the experiments. In addition, calculations on the Fe(III) azide precursor also gave good agreement with the experimentally known structure and spectroscopic parameters. Thus, the calculations provided strong evidence that the photolysis indeed led to a high-valent Fe(V)-nitrido species and that this species had an unanticipated ground state spin of $S = 1/2$. In order to confirm this result, magnetic susceptibility experiments were undertaken and revealed unambiguously that the ground state spin is indeed $S = 1/2$. The reason for this immediately became evident from an inspection of the calculated electronic structure. For an approximately octahedral site, the five d -orbitals split into a sigma-antibonding e_g and a π -antibonding t_{2g} set. The interaction with the extremely strongly π -donating N^{3-} ligand, however, strongly splits the t_{2g} set in a low-lying nonbonding d_{xy} orbital and a nearly degenerate π -antibonding d_{xz}, d_{yz} set. Energetically, spin-pairing in the d_{xy} orbital to produce the $(d_{xy}^2)(d_{xz,yz}^1)$ ($S = 1/2$) configuration is simply much more favorable than the alternative $(d_{xy}^1)(d_{xz}^1, d_{yz}^1)$ ($S = 3/2$) configuration. It is noteworthy that the $(d_{xy}^2)(d_{xz,yz}^1)$ ($S = 1/2$) is (almost) orbitally degenerate and Jahn–Teller active. Thus, there are two low-lying Kramers doubles which explains the extremely unusual EPR properties of the system. In fact, clean and reproducible EPR data could not be obtained.

It should be evident from the preceding description how a close interaction between theory and experiments can lead to essential insights on previously elusive species.

3.5

Binuclear Mixed Valence Copper Complexes – ENDOR and HYSCORE Spectra Simulations from DFT

Our last example deals with a combined theoretical and experimental study on a set of two delocalized class III mixed valence $\text{Cu}^{1.5}/\text{Cu}^{1.5}$, $S = 1/2$, complexes [94]. The structures of the two investigated systems are given in Fig. 9. The investigated species serve as model systems for the active sites of nitrous oxide reductases and cytochrome *c* oxidases.

From the experimental point-of-view, W- and X-band ENDOR, HYSCORE, and 2D-TRIPLE were used to obtain ^1H , ^2H , and ^{14}N HFCs and signs. The main goal of the accompanying DFT calculations was to provide help for the spectral simulations and the assignment of the experimental parameters as there was a substantial amount of spectral crowding and the analysis of the experimental data turned out to be extremely challenging. In addition, this study can be regarded as a further benchmark for the assessment of the accuracy that can be obtained in DFT calculations on transition-metal systems. In order to probe the influence of molecular structures on the calculated EPR parameters, results from X-ray crystallography and geometry optimizations employing the BP86 functional were compared.

Considering the electronic structures, similar SOMO's were calculated for the two complexes. They are formed by the antibonding combination of the two d_{z^2} orbitals, but with covalent admixtures of the axial nitrogen ligands. The spin densities of the two model systems are mainly determined by the shape of the SOMO's. Noticeable spin populations were found on all eight nitrogen atoms of both complexes. Thus, a large number of superhyperfine ^{14}N and ^1H contributions to the spectra can be expected in addition to the molecular *g*- and Cu hyperfine tensors. The calculated and measured *g*-values and Cu HFCs of model complex Cu_2RGT are given in Table 4 and are in unexpected good agreement with the experiments. This was explained by the fact that DFT works better for a delocalized system since the exchange hole is also

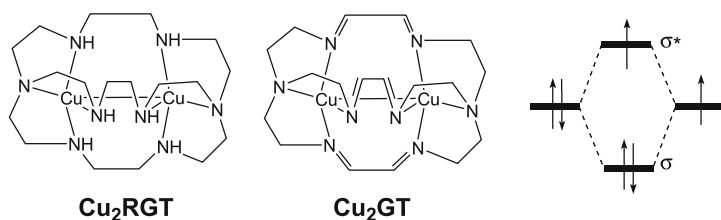


Fig. 9 The binuclear mixed valence $\text{Cu}^{\text{I}}/\text{Cu}^{\text{II}}$ complexes with azacryptand ligands (*left* and *middle*). The overlap of the two d_{z^2} orbitals with three electrons results in a rather strong doubly occupied Cu – Cu σ bond (*right*), which defines also the *z*-axis of the system ($g_{\parallel}$). The single unpaired electron resides in the antibonding combination of the two d_{z^2} orbitals and is completely delocalized over the two metal centers

Table 4 Comparison of calculated and experimental g -values, HFCs and NQCs (MHz) obtained for the model complex Cu_2RGT from Fig. 9

	Calcd.	Exptl.
$g_{\perp}$	2.1378	2.148
$g_{\parallel}$	2.0033	2.004
$\text{Cu } A_{\perp}$	-305.5	-308
$\text{Cu } A_{\parallel}$	0.3	10
$N_{\text{ax}}A_{\perp}$	14.3	14.5
$N_{\text{ax}}A_{\parallel}$	31.5	26.5
$N_{\text{ax}}e^2Qq/h$	-4.0	-3.7
$N_{\text{eq}}A_{\perp}$	5.1	5.5
$N_{\text{eq}}A_{\parallel}$	9.8	7
$N_{\text{ax}}e^2Qq/h$	-3.4	3.0

too delocalized in comparison to the exact one. Data of almost similar high quality were obtained for the Cu_2GT species. Interestingly, the agreement between theory and experiment is typically much worse for mononuclear Cu^{II} complexes. Nitrogen HFCs and NQCs were also well reproduced in the calculations (Table 4). These parameters were used for the calculation of W-band and HYSCORE spectra of Cu_2RGT , which impressively underlines the possibilities of present-day electronic structure calculations. Finally, ^1H HFCs were reported. Focusing on the N – H protons, an overestimation of the anisotropic HFCs was found, while the predicted Fermi-contact terms were too small.

Most importantly, however, is the fact that the DFT calculations gave a qualitatively correct description of the relative sizes of the various HFCs in the system. Perhaps even more importantly, the calculations provide numbers and insight for the orientation of the various HFC and NQC tensors which are required as input for the simulation routines. Even if the calculated numbers are slightly in error, they provide an essential help for analyzing puzzlingly complex experimental spectra which are typical of modern high-resolution EPR experiments. In this respect, the study in [94] provided an important benchmark study for showing how complicated experiments and advanced theoretical methods can come together to arrive at a very detailed experimentally calibrated spin density distribution picture in a large and complex molecule.

4

Conclusions and Outlook

Quantum chemical calculations have been proven to be an inherent part of modern spectroscopy. Among them, DFT calculations can be performed rou-

tinely for systems with more than 100 atoms and provide a large number of spectroscopic parameters at low computational cost. The strengths of the DFT approach are the prediction of trends in properties, tensor orientations and signs. Especially computed Euler angles can be highly relevant for experimentalists who want to simulate spectra with a large number of parameters. Considering the accuracy of the computed numbers, careful calibration studies are necessary. It was shown that very accurate results can be obtained for organic doublet radicals, while the investigation of transition-metal complexes is more complicated and the quality of the results depends strongly on the chosen systems.

Considering the calculation of hyperfine coupling tensors, basis sets are required that are uncontracted and flexible in the core region. Nevertheless, DFT tends to underestimate the isotropic Fermi-contact contributions, while the spin-dipole contributions are calculated quite well. Furthermore, second-order contributions must be taken into account for transition metals and heavier elements like sulfur. The predictions of these spin-orbit contributions can be substantially in error, which is evident in calculated HFCs and *g*-shifts. Considering the calculation of zero-field splitting parameters, more calibration studies are required.

In general, one can expect a more extensive use of CI-based methods for the calculation of spectroscopic properties in future applications, especially for transition-metal systems. The multideterminant character of the wave functions can be directly considered and it is possible to use spin eigenfunctions. However, the implementation of property calculations based on these *ab initio* methods is a challenging task but the attractive formal properties of these methods make them an attractive target.

Above all technical details, we hope that the present piece of work has provided a feeling of what can be achieved presently with theoretical spectroscopy approaches and how fruitful the interaction between theory and experiments can be in the elucidation of the structure and properties of complex and unstable molecules. While this line of research has already produced a substantial number of important contributions there is every reason to expect that the hand-in-hand use of experiments and calculations will become more and more commonplace in the immediate future.

References

1. Schweiger A, Jeschke G (2001) Principles of Pulse Electron Paramagnetic Resonance. Oxford University Press, Oxford
2. Kahn O (1993) Molecular Magnetism. VCH Publishers, Weinheim
3. Blundell S (2001) Magnetism in Condensed Matter. Oxford University Press, New York
4. Jiles D (1998) Introduction to Magnetism and Magnetic Materials. CRC Press, Boca Raton, Florida

5. Praneeth VKK, Neese F, Lehnert N (2005) *Inorg Chem* 44:2570
6. Stich TA, Buan NR, Brunold TC (2004) *J Am Chem Soc* 126:9735
7. Quintanar L, Gebhard M, Wang TP, Kosman DJ, Solomon EI (2004) *J Am Chem Soc* 126:6579
8. Kirk ML, Peariso K (2003) *Curr Opin Chem Biol* 7:220
9. Neese F, Solomon EI (1998) *J Am Chem Soc* 120:12829
10. Neese F, Zaleski JM, Zaleski KL, Solomon EI (2000) *J Am Chem Soc* 122:11703
11. Loeb KE, Zaleski JM, Westre TE, Guajardo RJ, Mascharak PK, Hedman B, Hodgson KO, Solomon EI (1995) *J Am Chem Soc* 117:4545
12. Robblee JH, Messinger J, Cinco RM, McFarlane KL, Fernandez C, Pizarro SA, Sauer K, Yachandra VK (2002) *J Am Chem Soc* 124:7459
13. Garner CD (1991) *Adv Inorg Chem* 36:303
14. Solomon EI, Szilagyi RK, DeBeer George S, Basumallick L (2004) *Chem Rev* 104:419
15. Johnson CE (1975) In: Gonser U (ed) *Mössbauer Spectroscopy (Topics in Applied Physics)*. Springer, Berlin, p 139
16. Neese F, Petrenko T, Ganyushin D, Olbrich G (2006) *Coord Chem Rev* (in press)
17. Neese F (2006) *J Biol Inorg Chem* 11:702
18. Sinnecker S, Neese F (2006) *J Comput Chem* 27:1463
19. Neese F (2006) *Special Per Rep EPR Spec* (in press)
20. Griffith JS (1964) *The Theory of Transition Metal Ions*. Cambridge University Press, Cambridge
21. Kaupp M, Bühl M, Malkin VG (2004) *Calculation of NMR and EPR Parameters. Theory and Applications*. Wiley, Weinheim
22. Neese F (2003) *Curr Opin Chem Biol* 7:125
23. Neese F, Solomon EI, Miller JS, Drillon M (2003) *Magnetism: Molecules to Materials*. Wiley, Weinheim, p 345
24. Bencini A, Gatteschi D (1990) *EPR of Exchange Coupled Systems*. Springer, Berlin
25. Hess BA, Marian CM, Wahlgren U, Gropen O (1996) *Chem Phys Lett* 251:365
26. Neese F (2005) *J Chem Phys* 122:204107
27. Gütlich P, Link R, Trautwein A (1978) *Mössbauer Spectroscopy and Transition Metal Chemistry*. Springer, Heidelberg
28. Gonser U (1975) In: Gonser U (ed) *Mössbauer Spectroscopy (Topics in Applied Physics)*. Springer, Berlin, p 1
29. Sinnecker S, Slep LD, Bill E, Neese F (2005) *Inorg Chem* 44:2245
30. Neese F (2002) *Inorg Chim Acta* 337:181
31. Zhang Y, Oldfield E (2003) *J Phys Chem B* 107:7180
32. Zhang Y, Mao JH, Godbout N, Oldfield E (2002) *J Am Chem Soc* 124:13921
33. Liu TQ, Lovell T, Han WG, Noodleman L (2003) *Inorg Chem* 42:5244
34. Hirata S, Head-Gordon M (1999) *Chem Phys Lett* 302:375
35. Furche F, Ahlrichs R (2002) *J Chem Phys* 117:7433
36. Röhrig UF, Frank I, Hutter J, Laio A, VandeVondele J, Rothlisberger U (2003) *Chemphyschem* 4:1177
37. Burcl R, Amos RD, Handy NC (2002) *Chem Phys Lett* 355:8
38. Dreuw A, Head-Gordon M (2005) *Chem Rev* 105:4009
39. Roos BO, Andersson K, Fülscher MP, Malmqvist PA, SerranoAndres L, Pierloot K, Merchán M (1996) *Adv Chem Phys* 93:219
40. Neese F (2003) *J Chem Phys* 119:9428
41. Cramer CJ, Truhlar DG (1999) *Chem Rev* 99:2161
42. Tomasi J (2004) *Theor Chem Acc* 112:184
43. Caricato M, Mennucci B, Tomasi J (2004) *J Phys Chem A* 108:6248

44. Yan WZ, Xue Y, Zhu H, Zeng J, Xie DQ (2004) *J Comput Chem* 25:1833
45. Stoyanov SR, Villegas JM, Rillema DP (2004) *J Phys Chem B* 108:12175
46. Benzi C, Crescenzi O, Pavone M, Barone V (2004) *Magn Reson Chem* 42:S57
47. Pecul M, Ruud K (2004) *Magn Reson Chem* 42:S128
48. Ciofini I (2004) *Magn Reson Chem* 42:S48
49. Ciofini I, Reviakine R, Arbuznikov A, Kaupp M (2004) *Theor Chem Acc* 111:132
50. Rinkevicius Z, Telyatnyk L, Vahtras O, Ruud K (2004) *J Chem Phys* 121:5051
51. Ciofini I, Adamo C, Barone V (2004) *J Chem Phys* 121:6710
52. Saracino GAA, Tedeschi A, D'Errico G, Improta R, Franco L, Ruzzi M, Corvaia C, Barone V (2002) *J Phys Chem A* 106:10700
53. Langgard M, Spanget-Larsen J (1998) *J Mol Struct THEOCHEM* 431:173
54. Barone V (1996) *Chem Phys Lett* 262:201
55. Sinnecker S, Rajendran A, Klamt A, Diedenhofen M, Neese F (2006) *J Phys Chem A* 110:2235
56. Siegbahn PEM, Blomberg MRA, Pavlov M (1998) *Chem Phys Lett* 292:421
57. Gao J (1995) In: Lipkowitz KD, Boyd DB (eds) *Reviews in Computational Chemistry*, Vol 7. VCH Publishers Inc., New York, p 119
58. Mordasini TZ, Thiel W (1998) *Chimia* 52:288
59. Altun A, Thiel W (2005) *J Phys Chem B* 109:1268
60. Lin H, Schöneboom JC, Cohen S, Shaik S, Thiel W (2004) *J Phys Chem B* 108:10083
61. Sinnecker S, Reijerse E, Neese F, Lubitz W (2004) *J Am Chem Soc* 126:3280
62. Asher JR, Doltsinis NL, Kaupp M (2004) *J Am Chem Soc* 126:9854
63. Benisvy L, Bittl R, Bothe E, Garner CD, McMaster J, Ross S, Teutloff C, Neese F (2005) *Angew Chem Int Ed Engl* 44:5314
64. Becke AD (1988) *Phys Rev A* 38:3098
65. Perdew JP (1986) *Phys Rev B* 33:8822
66. Perdew JP (1986) *Phys Rev B* 34:7406
67. Schäfer A, Huber C, Ahlrichs R (1994) *J Chem Phys* 100:5829
68. Neese F (2001) *J Chem Phys* 115:11080
69. Becke AD (1993) *J Chem Phys* 98:5648
70. Lee CT, Yang WT, Parr RG (1988) *Phys Rev B* 37:785
71. Barone V, Chong DP (1996) *Recent Advances in Density Functional Methods*. World Scientific Publ Co, Singapore, p 287
72. Klamt A, Schüürmann G (1993) *Perkin Trans*, p 799
73. Lubitz W, Feher G (1999) *Appl Magn Reson* 17:1
74. van Gastel M, Lubitz W, Lassmann G, Neese F (2004) *J Am Chem Soc* 126:2237
75. Kutzelnigg W, Fleischer U, Schindler M, Diehl P, Fluck E, Günther H, Kosfeld R, Seelig J (1990) *NMR Basic Principles and Progress*. Springer, Heidelberg, p 165
76. Costas M, Mehn MP, Jensen MP, Que L (2004) *Chem Rev* 104:939
77. Pestovsky O, Stoian S, Bominaar EL, Shan XP, Münck E, Que L, Bakac A (2005) *Angew Chem Int Ed Engl* 44:6871
78. Ghosh A, de Oliveira FT, Yano T, Nishioka T, Beach ES, Kinoshita I, Munck E, Ryabov AD, Horwitz CB, Collins TJ (2005) *J Am Chem Soc* 127:2505
79. Neese F (2006) *J Inorg Biochem* 100:716
80. Rohde JU, In JH, Lim MH, Brennessel WW, Bukowski MR, Stubna A, Münck E, Nam W, Que L (2003) *Science* 299:1037
81. Price JC, Barr EW, Tirupati B, Bollinger JM, Krebs C (2003) *Biochemistry* 42:7497
82. Que L, Ho RYN (1996) *Chem Rev* 96:2607
83. Krebs C, Price JC, Baldwin J, Saleh L, Green MT, Bollinger JM (2005) *Inorg Chem* 44:742

84. Schöneboom JC, Neese F, Thiel W (2005) *J Am Chem Soc* 127:5840
85. Baerends EJ, Ellis DE, Ros P (1973) *Chem Phys* 2:41
86. Dunlap BI, Connolly JWD, Sabin JR (1979) *J Chem Phys* 71:3396
87. Noodleman L (1981) *J Chem Phys* 74:5737
88. Noodleman L, Davidson ER (1986) *Chem Phys* 109:131
89. Shaik S, de Visser SP, Kumar D (2004) *J Biol Inorg Chem* 9:661
90. Shaik S, de Visser SP, Ogliaro F, Schwarz H, Schröder D (2002) *Curr Opin Chem Biol* 6:556
91. Hirao H, Kumar D, Thiel W, Shaik S (2005) *J Am Chem Soc* 127:13007
92. Meyer K, Bill E, Mienert B, Weyhermüller T, Wieghardt K (1999) *J Am Chem Soc* 121:4859
93. Aliaga-Alcalde M, DeBeer George S, Mienert B, Bill E, Wieghardt K, Neese F (2005) *Angew Chem Int Ed Engl* 44:2908
94. Kababya S, Nelson J, Calle C, Neese F, Goldfarb D (2006) *J Am Chem Soc* 128:2017
95. Kostka KL, Fox BG, Hendrich MP, Collins TJ, Rickard CEF, Wright LJ, Münck E (1993) *J Am Chem Soc* 115:6746

First-Principles Approach to Vibrational Spectroscopy of Biomolecules

Carmen Herrmann · Markus Reiher (✉)

Laboratorium für Physikalische Chemie, ETH Zürich, Hönggerberg Campus HCI,
 Wolfgang-Pauli-Str. 10, CH-8093 Zürich, Switzerland
 markus.reiher@phys.chem.ethz.ch

1	Introduction	86
2	Calculation of the Vibrational Normal Modes and Frequencies	91
2.1	Born–Oppenheimer Approximation	91
2.2	Harmonic Approximation	92
2.3	Efficient Calculation of Vibrational Normal Modes and Frequencies	93
3	Calculation of the Electronic Potential Energy Hypersurface	95
3.1	Ground State Hypersurfaces	96
3.2	Excited State Hypersurfaces	96
3.3	Environment Effects and Molecules in Solution	97
4	Calculation of Intensities in Vibrational Spectroscopy:	
	General Considerations	98
4.1	Interaction Between Radiation and One Molecule	98
4.2	Orientalional Averaging	102
4.3	Consideration of Experimental Conditions	103
4.4	Efficient Calculation of Intensities for Large Molecules	103
5	Selective Vibrational Spectroscopic Techniques	105
5.1	Difference-IR and -Raman	105
5.2	Two-Dimensional Techniques	107
5.3	Techniques Sensitive to Chirality: VCD and ROA	110
5.3.1	VCD Theory	112
5.3.2	ROA Theory	113
5.3.3	Molecular Conformation	114
5.3.4	Efficient Computational Procedures for Large (Bio-)Molecules	115
5.3.5	Interpretation of Calculated Spectra for Large (Bio-)Molecules	116
5.3.6	Solvent Effects	116
5.3.7	Local Contributions	118
5.4	Raman Techniques in Electronic Resonance	118
5.4.1	Spontaneous Resonance Raman Scattering	118
5.4.2	Resonance CARS	122
5.4.3	Resonance ROA	124
5.5	NRVS	124
6	Conclusion	125
	References	126

Abstract Vibrational spectroscopy of biomolecules like enzymes, nucleic acids, carbohydrates, lipids, and their components, is in most cases the vibrational spectroscopy of large molecules in aqueous solution or in vivo. Since large molecules in solution are likely to yield conventional infrared (IR) and Raman spectra with many close-lying peaks, spectroscopic techniques which filter out information selectively are of special interest in this field. Because of the large size of the investigated molecules and the lack of reliable *rules of thumb* for many special techniques, accurate first-principles calculations are an important means of interpreting the resulting spectra. First-principle calculations on biomolecules in solution have to cope with the challenges arising from the size of the systems under study, which make selective computational techniques an essential tool in order to be able to investigate biomolecular systems of reasonable size with sufficient computational accuracy. This is why the focus of this work is on the selective first-principles calculation of vibrational spectra of biomolecules obtained by special techniques such as difference IR and Raman, Vibrational Circular Dichroism, Raman Optical Activity, resonance Raman, and also Coherent Anti-Stokes Raman Spectroscopy, two-dimensional IR and Nuclear Resonance Vibrational Spectroscopy. For each of these techniques, a short introduction of their relevance for studies on biomolecules is given. Theoretical as well as practical aspects of calculating the corresponding intensities are discussed and complemented by references to original work on these topics.

Keywords Biomolecules · First-principles methods · Vibrational spectroscopy

Abbreviations

CARS	coherent anti-Stokes Raman scattering
CD	circular dichroism
CIS	configuration interaction using single excitations
DNA	deoxyribonucleic acid
DFT	density functional theory
HF	Hartree–Fock theory
IR	infrared
KS	Kohn–Sham
MD	molecular dynamics
PCM	polarizable continuum model
PES	potential energy surface
ROA	(vibrational) Raman optical activity
RPA	random phase approximation
TDA	Tamm–Dancoff approximation
TDDFT	time-dependent density functional theory
TDHF	time-dependent Hartree–Fock
VCD	vibrational circular dichroism
2D-IR	two-dimensional infrared (spectroscopy)

1

Introduction

In order to understand important processes of life, such as biological energy conversion and storage, carbon and nitrogen fixation, and the transcription

of genes, it is necessary to acquire a detailed knowledge of the underlying biochemical processes. Therefore, information is needed on the molecular structures of biologically relevant systems such as enzymes, deoxyribonucleic acid (DNA), viruses and their components, in particular amino acids and peptides, nucleic acids, sugars, and metalloenzyme active sites. Structural information on biomolecules is in principle available by a number of spectroscopic techniques such as X-ray crystallography [1, 2], NMR spectroscopy [3, 4], and vibrational spectroscopy. Using X-ray crystallography, a very detailed picture of the nuclear positions within a biomolecule may be obtained directly from experimental data, but the applicability of this technique is limited to molecules which can be prepared in crystalline form (however, see also the recent reports on X-ray diffraction by single molecules [5, 6] and noncrystalline samples [7]). Furthermore, the resolution may not be sufficient for an unambiguous assignment of the positions of comparatively light nuclei such as hydrogen or, in some cases, carbon, nitrogen or oxygen [8], which may be key atoms in biochemical reaction mechanisms. For example, the identity of the central ligand in the FeMo cofactor of nitrogenase (which is supposed to be N, but might also be O or C) is still not resolved [9, 10]. In addition, the need for crystalline samples may restrict X-ray structure investigations to inactive forms of the system under study. Vibrational spectroscopy, on the other hand, is complementary to X-ray crystallography in the sense that it can monitor biomolecules and their reactions (including active species and intermediates as well as excited states) in their natural environment, which is in most cases in (aqueous) solution, or even *in vivo*. However, the information vibrational spectroscopy provides on molecular structures is not obtained directly from experiment as in X-ray crystallography. It rather has to be extracted from a set of vibrational frequencies and corresponding intensities either by well-established empirical rules based on experimental experience and symmetry considerations, or by comparison to computed vibrational spectra.

Most experimental experience on vibrational spectroscopy has so far been gained with traditional infrared (IR) and Raman vibrational spectroscopy. As most biomolecules are comparatively large and have low (mostly no) symmetry, however, the resulting IR and Raman vibrational spectra contain a large number of vibrational bands which may overlap depending on the experimental resolution. As a consequence, new vibrational spectroscopic techniques such as resonance Raman, Vibrational Circular Dichroism (VCD), and Vibrational Raman Optical Activity (ROA) have been developed and made routinely applicable, which filter out certain vibrational bands based on special selection criteria. Because of the complicated molecular structure of most biomolecules and since experimental experience with these techniques is often still limited compared to traditional IR and Raman, theoretical calculations are needed to assign the spectra. In order to be of use for more than a few special cases, such simulations should be sufficiently accurate and applicable to any class of biomolecules and type of spectroscopy.

This is provided by a first-principles approach to the simulation of vibrational spectra of biomolecules, which is the focus of this work. A thematically related and very comprehensive recent review on theoretical and experimental techniques for the vibrational spectroscopy of biomolecules is provided in [11]. It is complementary to this work in the sense that it comprises numerous examples to demonstrate the current technical state-of-the-art, whereas this work concentrates rather on conceptual issues and uses examples to illustrate an underlying idea, which is the targeted filtering of information from vibrational spectra in experiment and in particular in first-principles calculations.

It is possible to circumvent the harmonic approximation in the calculation of molecular vibrations by first-principles molecular dynamics (MD) simulations [12] (see also Kirchner and collaborators, 2006, in this volume). In first-principles MD simulations, the motion of the nuclei is described by classical mechanics—only the interaction potential is not. This review, however, concentrates on static quantum chemistry, where the motion of the nuclei is described in a statistical manner, that is no explicit reference is made to the time variable in the equations of motion. A first-principles approach to vibrational spectroscopy implies that, in contrast to molecular dynamics based on (classical) molecular mechanics, the elementary particles in the theoretical treatment are the atomic *nuclei* and the electrons, not the atoms linked by some (modeled) interaction potentials. These particles are described based on the laws of quantum mechanics, and consequently first-principles interaction potentials are free from any empirical parameters other than fundamental natural constants. They are thus physically correct, provided the electronic structure of the molecules can be described with sufficient accuracy. The quantum mechanical treatment of molecules (see also Sects. 2 and 3 for more details) is usually based on the Born–Oppenheimer approximation [13], which decouples the description of nuclei and electrons. The nuclei are moving on the potential energy hypersurface (PES) obtained by solving the electronic Schrödinger equation. Furthermore, it is common to employ the harmonic approximation, in which the molecular vibrations are treated as a set of decoupled harmonic oscillators.

This procedure entails three problems:

1. Relevant biomolecules may easily be larger than 100–1000 atoms, which makes an accurate theoretical description of their electronic structure and the whole set of vibrational modes a challenging or even impossible task. This problem is aggravated if solvent effects are considered. Thus, it is possible that no advantage can be taken from the detailed correlation between spectral information and molecular structure provided by quantum chemical calculations. In experiment, the size of the systems under study leads to the reverse problem: The finite resolution may render individual peaks inseparable from one another, so that, although all information

is present in principle, it cannot be extracted from the overwhelming amount of data.

2. The Born–Oppenheimer approximation may break down at points on the electronic PES where several electronic states are lying close in energy, such as points near avoided crossings or conical intersections and on excited state hypersurfaces.
3. The system under study or the way in which it is studied may require corrections to the harmonic approximation. Without these anharmonic corrections, neither overtone nor combination bands are included in the theoretical description. For one-dimensional vibrational spectroscopic techniques, the neglect of anharmonicity results first of all in frequency shifts, whereas in two-dimensional vibrational spectroscopy, without anharmonicity the whole spectrum would vanish since positive and negative peaks would cancel.

The anharmonicity problem is thus in particular relevant for biomolecules in the theoretical treatment of two-dimensional vibrational spectroscopic techniques. For one-dimensional techniques, a consideration of anharmonic corrections may be necessary to answer special questions, but in general, a correct description of the intensities associated with certain selected frequencies is more important for an unambiguous correlation between vibrational peaks and molecular vibrations.

The possible breakdown of the Born–Oppenheimer approximation may be of importance for techniques in electronic resonance, where the properties of excited states are probed. Apart from the special case of VCD and far from resonance, employing the Born–Oppenheimer approximation is usually justified.

As far as the dimensionality problem is concerned, the “overcrowding” of vibrational spectra can be avoided in experiment either by reducing the information or by extending the space on which it is presented. The first approach results in special techniques such as resonance Raman, ROA or VCD spectroscopy, which amplify vibrational bands with certain properties and thus filter out information. The second approach is realized in two-dimensional IR and Raman techniques, where the intensities are plotted versus two frequency axes instead of one. From a theoretical point-of-view, the dimensionality problem arising from the size of the system under study can be divided into two subproblems which are separable in the theoretical treatment as well as in the sense that they can be carried out one after the other in actual calculations: (1) the calculation of the vibrational normal modes and frequencies and (2) the calculation of the corresponding intensities. In order not to lose accuracy in solving these problems by mode-selective techniques, a necessary condition is that no (or only well-controlled) approximations in addition to those already made in the calculation of full vibrational spectra are introduced (“well-controlled” means that the approximate size of the introduced

error is known). For the normal modes, this can be achieved by using the mode-tracking protocol [14], which makes the selected calculation of exact (harmonic) normal modes possible based on guessed vibrations that are easily constructed for a particular scientific problem [15–21]. Such iterative refinement schemes have also been applied to large fractions of a vibrational spectrum of a large biomolecule (usually the upper or lower part of the spectrum is obtained [22, 23]) instead of the selected, narrow sections of the spectrum as in mode-tracking. The corresponding intensities can be calculated subsequently by numerical direct differentiation of the appropriate molecular property tensors along those selected normal modes. This combination of techniques makes the calculation of the relevant parts of—in principle—all kinds of vibrational spectra much faster, and extends the range of (bio-) molecules accessible by accurate first-principles methods.

To summarize, whereas the anharmonicity and the Born–Oppenheimer problems are decisive only for certain techniques (often one may accept the numerical error introduced), the dimensionality problem is of importance for the first-principles treatment of all types of vibrational spectroscopies for biomolecules. Thus, this work concentrates on the *selective* calculation of vibrational frequencies and corresponding absorption or scattering intensities for biomolecules, and covers only experimental techniques which filter out special information compared to conventional IR and Raman spectroscopy. Furthermore, the techniques should be able to investigate biomolecules in their natural environment, that is either *in vivo*, in aqueous solution, or, for certain lipophilic molecules, in organic solution. Regarding these requirements, the following techniques are of particular interest: Difference IR and Raman techniques, which can monitor reactions, two-dimensional IR techniques, which may yield information on vibrational couplings and relaxation dynamics, chirality-sensitive techniques such as Vibrational Circular Dichroism (VCD) and Raman Optical Activity (ROA), which probe chirality as well as conformations, and Raman techniques in electronic resonance, which are sensitive for electronically excited molecular states (including Coherent Anti-Stokes Raman Spectroscopy (CARS), which may replace resonance Raman in the case of strongly fluorescent samples). Furthermore, due to its special, nucleus-specific selection criteria, quantum chemical calculations for Nuclear Resonance Vibrational Spectroscopy (NRVS) are shortly summarized, which is suited for the investigation of iron-containing systems in crystalline form.

All these techniques have in common that they probe the same set of molecular vibrations, but differ in which (and how) subsets of these vibrations are filtered out via the corresponding intensities. As normal modes and intensities can be calculated separately, they are treated here in separate sections: Sects. 2 and 3 summarize the calculation of vibrational normal modes and frequencies, the latter focusing on the calculation of the electronic PES, while Sect. 4 outlines the theoretical background of absorption or scattering intensity calculations in general and Sect. 5 summarizes the first-principles

description of several selective vibrational spectroscopic techniques. Since the focus in our group is currently on resonance Raman and Vibrational Raman Optical Activity, these two techniques are treated in greater detail than the remaining ones.

As far as the notation is concerned, vectors are printed in bold-face and small letters, matrices in bold-face and capitals. As an exception, nuclear coordinates are indicated by bold-face capitals, $\mathbf{R}$ referring to cartesian and $\mathbf{Q}$ to normal coordinates.

2

Calculation of the Vibrational Normal Modes and Frequencies

2.1

Born–Oppenheimer Approximation

The basis of the quantum chemical description of molecular vibrations of (bio-)molecules with more than a few atoms is the Born–Oppenheimer approximation [13], which allows for a separate description of the movements of the atomic nuclei and the electrons. These are coupled in the time-independent nonrelativistic Schrödinger equation

$$\hat{H}_{\text{mol}}\Psi_I = [\hat{T}_K + \hat{T}_e + \hat{V}_{\text{nuc,nuc}} + \hat{V}_{\text{nuc,e}} + \hat{V}_{\text{e,e}}]\Psi_I = E_I\Psi_I, \quad (1)$$

(where I labels the state of the system) by the electron-nucleus-interaction Hamiltonian $\hat{V}_{\text{nuc,e}}$, whereas the contributions to the total molecular Hamiltonian $\hat{H}_{\text{mol}}$ which describe the kinetic energy of the nuclei ($\hat{T}_K$), the kinetic energy of the electrons ($\hat{T}_e$), the interaction between the nuclei ($\hat{V}_{\text{nuc,nuc}}$), and the interaction between the electrons ($\hat{V}_{\text{e,e}}$) refer to either nuclear or electronic coordinates only and thus do not prevent a separation of the Schrödinger equation into a nuclear and an electronic part. The Born–Oppenheimer approximation is based on the adiabatic approximation, in which the wave function $\Psi_I(\mathbf{R}, \mathbf{r})$ is separated into an electronic and a nuclear part. The nuclear part $\chi_a(\mathbf{R})$ depends on the $3M$ coordinates $\mathbf{R}_A$ of all M nuclei collected in the vector $\mathbf{R}$, and the electronic part $\psi_{e,i}(\mathbf{r}, \tilde{\mathbf{R}})$ depends explicitly on the $3N$ coordinates $\mathbf{r}_i$ of all N electrons (summarized as the vector $\mathbf{r}$), and parametrically on the nuclear coordinates¹ (which is expressed by the tilde),

$$\Psi_I(\mathbf{R}, \mathbf{r}) = \chi_a(\mathbf{R}) \cdot \psi_{e,i}(\mathbf{r}, \tilde{\mathbf{R}}). \quad (2)$$

¹ Parametrical dependence of the electronic wave function on the nuclear coordinates means that these coordinates enter the electronic Schrödinger equation as constant values. Different electronic wave functions are thus obtained for different sets of nuclear coordinates, i.e., for different molecular structures.

The indices a and i label nuclear and electronic energy levels, respectively, and are defined by the composite total index I . By neglecting the action of the nuclear kinetic energy operator onto the electronic wave function, the adiabatic approximation turns into the Born–Oppenheimer equation. As a result, after inserting Eq. 2 into Eq. 1, multiplying from the left with $\Psi_{e,i}^*(\mathbf{r}, \tilde{\mathbf{R}})$ and integrating over all electronic coordinates, the nuclear Schrödinger equation

$$[\hat{T}_K + E_{e,i}(\mathbf{R})] \chi_a(\mathbf{R}) = E_{\text{tot}} \chi_a(\mathbf{R}) \quad (3)$$

is obtained, which describes the nuclei as moving on the PES defined by $E_{e,i}(\mathbf{R})$ of the electronic state i . The PES may be obtained pointwise by solving the electronic Schrödinger equation,

$$[\hat{T}_e + \hat{V}_{\text{nuc,nuc}} + \hat{V}_{\text{nuc,e}} + \hat{V}_{\text{e,e}}] \Psi_{e,i}(\mathbf{r}, \tilde{\mathbf{R}}) = E_{e,i}(\mathbf{R}) \Psi_{e,i}(\mathbf{r}, \tilde{\mathbf{R}}), \quad (4)$$

for different positions of the nuclei. This provides the potential in which the nuclei are moving. The solution of the electronic Schrödinger equation in practice is shortly summarized in Sect. 3. As far as the nuclear Schrödinger equation is concerned, it is sufficient to know that it is possible to calculate the electronic PES and its derivatives pointwise.

2.2

Harmonic Approximation

The most simple expression for the nuclear wave function $\chi_a(\mathbf{R})$ is a product of functions which each only depend on one nuclear coordinate. This product will be an exact solution of Eq. 3 if the individual coordinates are not coupled through the terms in the Hamiltonian operator in Eq. 3. The first step to achieve such a decoupling is to express the dependence of the PES on mass-weighted cartesian nuclear coordinates $R_A^{(m)} = M_A^{1/2} R_A$ (where the index A refers to the x , y , or z coordinate of a specific nucleus and its nuclear mass) as a Taylor series around the nuclear equilibrium structure $\mathbf{R}^{(m)} = \mathbf{0}$,

$$\begin{aligned} E_{e,i}(\mathbf{R}^{(m)}) &= E_{e,i}(\mathbf{0}) + \sum_{A=1}^{3M} \left(\frac{\partial E_{e,i}}{\partial R_A^{(m)}} \right)_0 R_A^{(m)} \\ &+ \frac{1}{2} \sum_{A=1}^{3M} \sum_{B=1}^{3M} \left(\frac{\partial^2 E_{e,i}}{\partial R_A^{(m)} \partial R_B^{(m)}} \right)_0 R_A^{(m)} R_B^{(m)} + \dots, \end{aligned} \quad (5)$$

and to truncate it after the second-order term (harmonic approximation, see Fig. 1). Then, considering that the energy gradients are zero at the equilibrium (= minimum) structure, choosing the equilibrium energy as zero and transforming it to a set of $3M$ mass-weighted normal coordinates $Q_A^{(m)}$ which makes all mixed second derivatives of the electronic energy

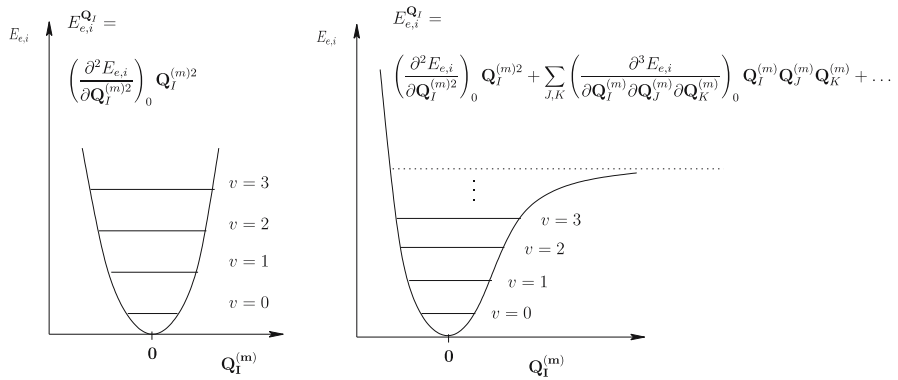


Fig. 1 Qualitative potential energy curves and the first four vibrational energy levels for the anharmonic case (*right-hand side*) and its harmonic approximation (*left-hand side*) along a mass-weighted vibrational normal mode $Q_A^{(m)}$ in the i th electronic state (which may be the ground state or an excited state). The total energy $E_{e,i}(\{Q_I\})$ is related to the energy for elongations along one normal mode $E_{e,i}^{Q_i}$ by $E_{e,i}(\{Q_I\}) = \sum_I E_{e,i}^{Q_i}$. In the anharmonic case, no vibrational energy levels are found above the dissociation level indicated by the *dotted line*

$\left(\partial^2 E_{e,i} / \partial Q_A^{(m)} \partial Q_B^{(m)} \right)_{A \neq B}$ vanish, a decoupled description is possible. The nuclear kinetic energy operator $\hat{T}_K = \sum_A \partial^2 / \partial Q_A^{(m)2}$, is also diagonal with respect to these coordinates, so that the product

$$\chi_a \left(\{Q_I^{(m)}\} \right) = \psi_1 \left(Q_1^{(m)} \right) \cdot \psi_2 \left(Q_2^{(m)} \right) \cdot \dots \cdot \psi_M \left(Q_M^{(m)} \right) \quad (6)$$

is an exact solution of the nuclear Schrödinger equation in harmonic approximation, which is now separable into $3M$ independent equations of harmonic oscillator form with a mass equal to unity (since mass-weighted coordinates are employed). The vibrations of the molecule are thus described as a system of decoupled harmonic oscillators. The eigenfunctions of harmonic oscillator equations are known, and their eigenvalues (the angular frequencies ω_I) are equal to the square root of the second energy derivatives. Of course, since $\omega = 2\pi\nu$ and $\tilde{\nu} = \nu/c$, vibrational frequencies ν and wavenumbers $\tilde{\nu}$ can be easily calculated from the angular frequencies.

2.3

Efficient Calculation of Vibrational Normal Modes and Frequencies

The conventional approach to obtain the transformation from cartesian to normal coordinates is to calculate all entries of the mass-weighted Hessian matrix (the matrix containing the second derivatives $\left(\partial^2 E_{e,i} / \partial R_A^{(m)} \partial R_B^{(m)} \right)_0$ of

the electronic energy with respect to nuclear coordinates), and to diagonalize it,

$$\mathbf{Q}^{(m)T} \mathbf{H}^{(m)} \mathbf{Q}^{(m)} = \mathbf{H}^{(Q,m)}. \quad (7)$$

The mass-weighted normal modes are the columns of the transformation matrix $\mathbf{Q}^{(m)}$. For nonlinear molecules, the diagonalization of the Hessian yields $3M - 6$ vibrational normal modes, as well as three rotational and three translational modes with eigenvalues of zero.

By diagonalizing the Hessian matrix, all vibrational normal modes and frequencies of the system under study are obtained. In the special vibrational spectroscopic techniques which are discussed in this work, however, the focus is on filtering out selected vibrations from an overcrowded vibrational spectrum. A calculation of all modes and frequencies would therefore produce a large amount of rather useless data at high computational cost. This can be circumvented by the mode-tracking protocol, which allows for the selective calculation of vibrational frequencies and normal modes from eigenpairs of the Hessian matrix through subspace iteration [14, 17] (for representative applications, see [15–20, 24, 232]). The mode-tracking protocol can be summarized as follows:

1. Guess a vibration $\mathbf{Q}_\mu^{(m,0)}$, which may be chosen based on chemical intuition or on a precedent vibrational analysis with a lower level of accuracy.
2. Use this guessed vibration as a first basis vector $\mathbf{b}_1$ in a Davidson-like algorithm [25]:

$$\mathbf{b}_1 = \mathbf{Q}_\mu^{(m,0)}. \quad (8)$$

3. Construct the small Davidson matrix $\tilde{\mathbf{H}}^{(i)}$ from the Hessian matrix $\mathbf{H}^{(m)}$ and the matrix $\mathbf{B}^{(i)}$ which contains all basis vectors of iteration i as columns,

$$\tilde{\mathbf{H}}^{(i)} = \mathbf{B}^{(i)T} \mathbf{H}^{(m)} \mathbf{B}^{(i)} \quad (9)$$

and diagonalize it:

$$\tilde{\mathbf{H}}^{(i)} \mathbf{c}_\mu^{(i)} = \lambda_\mu^{(i)} \mathbf{c}_\mu^{(i)} \quad (10)$$

4. Calculate the approximate eigenvectors $\mathbf{Q}_\mu^{(m,i)}$ of the Hessian from $\mathbf{c}_\mu^{(i)}$ and $\mathbf{B}^{(i)}$:

$$\mathbf{Q}_\mu^{(m,i)} = \sum_{j=1}^i c_{\mu,j}^{(i)} \mathbf{b}_j \quad (11)$$

5. Check for convergence.

6. If the convergence criteria are not satisfied, create one new basis vector per tracked normal mode $\mathbf{Q}_\mu^{(m)}$ from the corresponding residuum vector $\mathbf{r}_\mu^{(i)}$,

$$\mathbf{r}_\mu^{(i)} = \sum_{j=1}^i c_{\mu,j}^{(i)} \left[\mathbf{H}^{(m)} \mathbf{b}_j - \lambda_\mu^{(i)} \mathbf{b}_j \right] \quad (12)$$

and go back to step 3 (for details on convergence criteria and creating new basis vectors using a preconditioner see, e.g., [14]).

In order to calculate the matrix product $\mathbf{B}^{(i)T} \mathbf{H}^{(m)} \mathbf{B}^{(i)}$ without computing the Hessian matrix $\mathbf{H}^{(m)}$, a matrix $\boldsymbol{\Sigma} = \mathbf{H}^{(m)} \mathbf{B}^{(i)}$ with columns $\boldsymbol{\sigma}^j$ is defined. These $\boldsymbol{\sigma}^j$ are evaluated by applying the chain rule for differentiation,

$$\begin{aligned} \sigma_{j,I} &= \sum_B \left(\frac{\partial^2 E_{e,i}}{\partial R_A^{(m)} \partial R_B^{(m)}} \right)_0 b_{j,B} = \left(\frac{\partial^2 E_{e,i}}{\partial R_A^{(m)} \partial \mathbf{b}_j} \right)_0 \\ &= \left[\frac{\partial}{\partial \mathbf{b}_j} \left(\frac{\partial E_{e,i}}{\partial R_A^{(m)}} \right) \right]_0. \end{aligned} \quad (13)$$

The components of σ_j can thus be calculated as numerical derivatives of the gradient components along the basis vectors $\mathbf{b}_j$. Several eigenvectors may be optimized simultaneously using the Davidson–Liu (also known as block-Davidson) method [26, 27].

In contrast to other techniques such as partial Hessian vibrational analysis [28], where a block corresponding to an interesting subsystem is cut out of the Hessian matrix and diagonalized separately, mode-tracking will provide exact normal modes and frequencies if the convergence criteria are chosen sufficiently tight. For systems with a repeated structural motif, such as the amide bond in peptides, vibrational modes which are dominated by motions of the atoms belonging to this structural motif may also be calculated selectively in an approximate fashion by transfer of molecular properties from small model systems to larger molecules of interest. A general algorithm relying on the transfer in Cartesian coordinates instead of internal ones has been developed by Bouř and coworkers [29–31], and refined for arbitrary polypeptides in solution by Choi et al. [32–34]. This scheme may also be applied to the property tensors determining the spectral intensities, and has accordingly been applied to IR, Raman, VCD and ROA spectroscopy [29] (see also Sect. 5.3).

3

Calculation of the Electronic Potential Energy Hypersurface

In the context of vibrational spectroscopy, the aim of calculating the electronic PES by solving the electronic Schrödinger equation is to provide

a physically correct potential $E_{e,i}(\mathbf{R})$ in which the nuclei are moving (Eq. 3). Furthermore, the shapes of excited state energy surfaces are needed for spectroscopic techniques in or near electronic resonance. The PES can only be calculated pointwise, but an analytical form describing the dependence of the total energy on the nuclear coordinates, such as the truncated Taylor series around the molecular equilibrium structure in the harmonic approximation, may be approximated based on these energy values at certain points and derivatives of the energy with respect to nuclear displacements.

3.1

Ground State Hypersurfaces

In order to provide physically reasonable electronic energies and its derivatives, a fast but sufficiently accurate electronic structure method is needed. Currently, Kohn–Sham density functional theory (KS-DFT) [35–37], which in principle fully describes electron correlation, is the most general first-principles method which meets these requirements, and in practice the only one which is applicable to large biomolecules. Its efficiency is comparable to that of Hartree–Fock calculations and can be further enhanced by using density-fitting techniques [38, 39] provided pure density functionals are employed. Because of the size of the systems under study, relativistic effects required to be described if heavy elements are present are usually neglected, with the exception of the scalar-relativistic effects included in effective core potentials (see, e.g., [40]).

3.2

Excited State Hypersurfaces

Excited state hypersurfaces can be calculated either by explicit methods or by using linear response theory. The first class comprises methods such as Configuration Interaction Singles (CIS, compare also the Tamm–Dancoff approximation (TDA) [41, 42]) or Complete Active Space theory (CASSCF/CASPT2, see [43] for a review on applications to various organic photosystems), while the latter encompasses time-dependent Hartree–Fock (TDHF, also known as random phase approximation (RPA) in physics) as well as the electron-density based time-dependent DFT (TDDFT). Since many important biomolecules are far too large to be treated entirely by multi-reference methods, it is desirable to be able to calculate accurate excited state energies and properties with first-principles methods based on a single Slater-determinant. An extensive review of such methods is provided by [44]. As an alternative approach, multireference ab-initio methods may be combined with a MM force field within a Quantum Mechanics/Molecular Mechanics (QM/MM) framework, which has been successfully applied, e.g., to the excited state of the visual pigment rhodopsin [45, 46]. For a comparative review

on different quantum chemical methods applied to excited state calculations on biomolecules see [47].

3.3

Environment Effects and Molecules in Solution

In order to study the properties of (bio-)molecules with first-principles methods, it is necessary to separate a suitable subsystem from its environment and treat it as an isolated system [48, 49]. In order to achieve this in the case of molecules in solution, it is often necessary to consider not only the solute, but also the surrounding solvent. Solvent effects may change the nuclear arrangement in the molecular equilibrium structures, as well as vibrational normal modes and frequencies, or electronic properties by changing the wave function/electron density compared to the free molecule (see [50] for the quantum chemical description of solvent effects on a biomimetic arginine fork model). They can be modelled either explicitly in the form of solvent molecules which are attached to the molecule under study at appropriate positions, or implicitly in the form of dielectric continuum models, or by a combination of both. A perfect description of solvent effects is achieved within the Car–Parrinello molecular dynamics framework under periodic boundary conditions. To bridge the gap to quantum chemical isolated-molecule calculations is not trivial [51]. A very extensive up-to-date review on different continuum models and their application to molecular property calculations is provided by [52]. Reasonable positions and orientations of explicit solvent molecules may be either obtained by intuition or from MD snapshots. For a review on Car–Parrinello Molecular Dynamics for biological systems, see the article by Kirchner et al. in this volume. In the case of large biomolecules with a localized region of interest, such as metalloproteins, it may be necessary (and sufficient) to cut out this region from the rest of the molecule and to treat it either separately or to model the remainder in a more approximate fashion, e.g., in the framework of a QM/MM approach [53] (see also the contribution by Senn and Thiel in this volume). For a recent review on the technical state-of-the-art of the QM/MM-treatment of complex biological processes, see [54], and for applications of QM/MM to enzyme catalysis [55–57]. Another approach to such problems is frozen-density embedding [58], where the system under study is surrounded by a larger system whose effect is represented by an orbital-free density which enters in terms of an effective potential in the Kohn–Sham equations for the system under study. This approach is also applicable to excited state calculations within TDDFT [59].

4

Calculation of Intensities in Vibrational Spectroscopy: General Considerations

In order to provide a basis for the theoretical treatment of the more special vibrational spectroscopic techniques, the concepts of the calculation of IR and Raman intensities shall be outlined in the following. For further details on how to compute IR and Raman intensities, we refer to [60]. The theory of light absorption or (noncoherent) scattering by molecules in solution can be divided into three steps, starting from the most simple model case and approaching experimental conditions stepwise:

1. Description of the interaction between radiation and *one* molecule.
2. Orientational averaging if the molecules are randomly oriented.
3. Consideration of experimental conditions like environment (usually solvent) effects, temperature, ...

4.1

Interaction Between Radiation and One Molecule

It is in principle possible to treat both the radiation and the molecule either classically or quantum mechanically. In order to obtain quantitatively correct absorption and scattering intensities, however, the molecule under study has to be described by the laws of quantum mechanics, whereas the radiation may be treated by classical electrodynamics. This combination is called a semiclassical treatment. For simple processes, it yields the same intensity expressions as a (more tedious) quantum mechanical treatment of the radiation (but fails in special cases such as spontaneous emission). A molecule in the electromagnetic field of a light wave is described within the semiclassical approximation by a Hamiltonian

$$\hat{H} = \hat{H}_{\text{mol}} + \hat{H}_{\text{mol-rad}} , \quad (14)$$

where $\hat{H}_{\text{mol}}$ is the Hamiltonian of the isolated molecule defined in Eq. 1, and $\hat{H}_{\text{mol-rad}}$ a Hamiltonian describing its interaction with the radiation. In a fully quantum mechanical treatment, a third term $\hat{H}_{\text{rad}}$ describing the radiation would be included in the total Hamiltonian $\hat{H}$. When a molecule is interacting with radiation, essentially three things (or a combination of them) can happen: The molecule can emit light stimulated by the radiation and change to a lower-energy state, it can absorb light resulting in a transition to a higher-energy state, and it can scatter light and either keep its initial state (Rayleigh scattering) or undergo transitions upon scattering (Raman scattering). The goal of the theoretical treatment is to derive expressions for the emission, absorption, or scattering intensities, respectively, related to each of these processes. The intensity expressions will depend on the energy levels of

the molecule as well as on the energy and thus the frequency of the radiation, and should be so general that they may be adopted for any relation between these energies.

The contribution of the Hamiltonian $\hat{H}_{\text{mol-rad}}$ to the total Hamiltonian is usually considered much smaller than the one of $\hat{H}_{\text{mol}}$, so that interactions with an oscillating electromagnetic field $\tilde{\mathbf{E}}$ of a light wave with an angular frequency ω_L and a (complex) amplitude $\tilde{E}_0$,

$$\tilde{\mathbf{E}} = \tilde{E}_0 e^{-i\omega_L t}, \quad (15)$$

are usually described by perturbation theory. The perturbation operator $\hat{H}_{\text{mol-rad}}$ in the case of a dynamic electromagnetic field with electric and magnetic field amplitudes E_0 and B_0 can be written for an uncharged molecule as [61]

$$\hat{H}_{\text{mol-rad}} = - \sum_{i=1}^3 \hat{\mu}_i (E_i)_0 - \frac{1}{3} \sum_{i=1}^3 \sum_{j=1}^3 \hat{\theta}_{ij} (\Delta_i E_j)_0 - \sum_{i=1}^3 \hat{m}_i (B_i)_0 + \dots, \quad (16)$$

where $\hat{\mu}_i$, $\hat{\theta}_{ij}$, and $\hat{m}_i$ are the x , y , and z components of the electric dipole, the electric quadrupole, and the magnetic dipole operator, respectively, and $(E_i)_0$, $(\Delta_i E_j)_0$, $(B_i)_0$ etc. are the x , y , and z components of the electric and magnetic field and its gradients at the gauge origin, which vary with time, of course. The Hamiltonian $\hat{H}_{\text{mol-rad}}$ is therefore time-dependent, so that a time-dependent version of perturbation theory has to be employed. It has been found that it is often sufficient in this context to restrict the perturbational treatment to first order.

The intensity of the radiation scattered by a molecule can then be obtained by calculating the (oscillating) electric and magnetic multipole moments induced in the molecule by the light wave and treating them as classical sources of radiation. For Rayleigh scattering, the induced multipole moments are oscillating with the same phase and frequency as the incident radiation and can be computed from the expectation values of the molecular multipole operators using the perturbed time-dependent molecular wave functions and subtracting the corresponding permanent molecular multipole moment from this expectation value. Raman scattering can be described within a semi-classical formalism by introducing transition multipole moments between an initial state $|i\rangle$ and a final state $|f\rangle$,

$$\tilde{\mu}_{fi}^{(1)} = \langle f | \tilde{\mu}^{(1)} | i \rangle, \quad (17)$$

which are oscillating with an angular frequency equal to $\omega_L - \omega_{fi}$ (Stokes) or $\omega_L + \omega_{fi}$ (Anti-Stokes), and are unrelated in phase to the incident radiation due to the random phase of the molecular vibrations.

For ordinary Raman scattering, only a linear dependence of the induced electric dipole moment on the electric field amplitude is considered. The

induced electric dipole transition moment $\tilde{\mu}_{fi}^{(1)}$ is thus proportional to the amplitude $\tilde{E}^0$ of the oscillating electric field of the radiation,

$$\tilde{\mu}_{fi}^{(1)} = \tilde{\alpha}_{fi} \tilde{E}^0. \quad (18)$$

Higher-order effects can be included by considering nonlinear dependencies on the electric field described in terms of (higher-order) hyperpolarizabilities. For the sake of simplicity of notation, complex quantities are used in Eq. 18 and in the following, denoted by a tilde, although for Raman scattering, only the real part of the induced dipole moment is needed [62, 63]. The (electric dipole–electric dipole) polarizability $\tilde{\alpha}$ can be obtained from time-dependent perturbation theory as [64]

$$(\tilde{\alpha}_{fi})_{\alpha\beta} = \frac{1}{\hbar} \sum_{r \neq f,i} \left(\frac{\langle f | \hat{\mu}_\alpha | r \rangle \langle r | \hat{\mu}_\beta | i \rangle}{\omega_{ri} - \omega_L - \frac{1}{2}i\Gamma_r} + \frac{\langle f | \hat{\mu}_\beta | r \rangle \langle r | \hat{\mu}_\alpha | i \rangle}{\omega_{rf} + \omega_L + \frac{1}{2}i\Gamma_r} \right), \alpha \in \{x, y, z\}. \quad (19)$$

$\hat{\mu}_\alpha$ is a component of the molecular electric dipole operator, and Γ_r is a damping factor equal to the inverse of the life time of the molecular state r . The oscillating induced dipole described by Eq. 18 is treated as a classical source of radiation, which results in the intensity formula [65]

$$I(\theta) = \frac{\omega^4 |\mu_{fi}^{(1)}|^2 \sin^2(\theta)}{32\pi^2 c^3 \epsilon_0} \quad (20)$$

for the radiation scattered at an angle θ relative to the direction of the incident beam. c is the speed of light, ϵ_0 the permittivity of the vacuum, and ω is the angular frequency of the scattered radiation. In the case of spectroscopic techniques which probe molecular chirality, i.e., VCD and ROA, restricting the theoretical treatment to induced dipole moments in the form of Eq. 18 would yield intensity differences of zero for all vibrations. Therefore, contributions from the molecular induced magnetic dipole and electric quadrupole moments have also to be taken into account, and Eqs. 20 and 18 have to be adjusted accordingly (see [64]).

In practice, the transition polarizability components $(\tilde{\alpha}_{fi})_{\alpha\beta}$ cannot be evaluated in the form given by Eq. 19, since this would require knowledge about all excited molecular states $|k\rangle$. It can be simplified, however, by making assumptions which are justified for special cases, and thus restrict the applicability of the resulting formulae to those special cases. In the case of nonresonant Raman scattering, the frequency of the incident radiation is assumed to be far below electronic transition frequencies of the molecule (which also implies that the damping factors can be neglected), and the electronic ground state to be nondegenerate. As mentioned in Sect. 2, by employing the adiabatic approximation, the molecule can be treated as moving on an electronic

energy hypersurface. In analogy to the simplification of the dependence of the electronic energy on nuclear coordinates in the calculation of vibrational motions and frequencies, the polarizability is therefore commonly expanded into a Taylor series, which is truncated after the first-order term (electric harmonic approximation)²,

$$\alpha(Q) = \alpha_0 + \sum_p \left(\frac{\partial \alpha}{\partial Q_p} \right)_0 Q_p + \dots \quad (21)$$

Then, the real part of Eq. 19 can be simplified to

$$(\alpha_{fi})_{\alpha\beta} = (\alpha_{\alpha\beta})_0 \langle v_f | v_i \rangle + \sum_p \left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \langle v_f | Q_p | v_i \rangle, \quad (22)$$

which is known as Placzek's approximation. $|v_i\rangle$ and $|v_f\rangle$ are abbreviations for the nuclear wave functions of the initial and the final state. Since $|v_i\rangle$ and $|v_f\rangle$ are approximated as products of harmonic oscillator eigenfunctions (Sect. 2), the matrix elements $\langle v_f | Q_p | v_i \rangle$ can be deduced from the quantum mechanical treatment of the harmonic oscillator. For vibrational transitions involving only the change of one vibrational normal mode from quantum number n to $n + 1$, they are equal to $\hbar/2\omega_p$, and zero in all other cases. In the case of resonant Raman scattering, Eq. 19 is simplified by the fact that the summands containing the electronic state in resonance are much larger than the others. This is treated in more detail in Sect. 5.4.

For absorption associated with a molecular transition from the initial state $|i\rangle$ to the final state $|f\rangle$, the transition rate can be calculated via time-dependent perturbation theory and is given by

$$\frac{dP_{f \leftarrow i}}{dt} = \frac{2\pi}{\hbar} |\langle f | \hat{H}_{\text{mol-rad}} | i \rangle|^2 \rho(E_{fi}) = \frac{1}{\hbar^2} |\langle f | \hat{H}_{\text{mol-rad}} | i \rangle|^2 \rho(v_{fi}), \quad (23)$$

which is Fermi's Golden Rule. The transition rate is the change of the probability to find the system (which was originally in the initial state $|i\rangle$) in the final state $|f\rangle$ w.r.t. time, i.e., the number of transitions $|f\rangle \leftarrow |i\rangle$ per time. The prefactor in Eq. 23 depends on whether the density of states per energy range $\rho(E_{fi})$ is expressed using the energy of the states or the associated frequencies. By considering only the electric-dipole term in the perturbation operator in Eq. 16 and based on similar considerations as for the transition polarizability, the transition matrix elements in Fermi's Golden Rule given by Eq. 23 can be calculated from the derivatives of the molecular dipole moment along mass-weighted vibrational normal modes $(\partial \mu / \partial Q_p)_0$ (see [60] for more details).

² In contrast to this, the expansion is truncated after the second-order term in the mechanical harmonic approximation. Since the quantum chemical calculation of Raman intensities involves both approximations, this is then called "the double harmonic approximation".

To summarize, the following approximations are made in the established theoretical double harmonic approximation of IR and Raman spectroscopy:

1. Born–Oppenheimer approximation.
2. (Mechanical) harmonic approximation.
3. Semi-classical treatment: The contributions from classical electrodynamics are expressions for the dependence of the induced multipole moments on the incident radiation, as well as for the dependence of the scattered radiation on the induced multipole moments. Quantum mechanics provides vibrational frequencies and transition matrix elements and thus accounts for all quantities depending on the molecule.
4. The time-dependent perturbation theory employed for the derivation of Eqs. 18 and 23 is restricted to first-order.
5. Only the electric dipole moment induced by the incident radiation is considered, and all other induced electric and magnetic multipole moments are neglected.
6. The electromagnetic field of the incident radiation is considered to be constant over the whole molecule at a given time, since the wavelength of the radiation is usually larger than the molecular region which is affected by the interaction with it. A typical wavelength employed for Raman measurements would be, for example, 5000 Å, whereas typical bond distances are of the order 1–2 Å. In IR spectroscopy, this ratio is more favorable due to the larger wavelengths of the incident radiation.
7. Molecular property tensors are assumed to depend linearly on nuclear distortions close to the equilibrium structure (electrical harmonic approximation).

4.2

Orientational Averaging

In order to obtain IR and Raman intensity expressions for molecular gases, solutions, and fluids, i.e., for a large number of randomly oriented molecules, classical orientational averaging of the intensity expressions derived for one molecule may be employed. Since the radiation scattered in a spontaneous Raman process is incoherent due to the random phase of the molecular vibrations [65], the intensity of the radiation scattered by N molecules is N times the intensity scattered by one molecule, averaged over all possible molecular orientations. Orientational averaging can be achieved by replacing space-fixed polarizability tensor elements by the classical average over molecule-fixed polarizability elements for all possible molecular orientations. As a result, intensity formulae depending on the polarization of the incoming and the detected radiation beam as well as on the angle between both beams are obtained. For details on this topic, we refer to [60], which also contains a detailed comparison of the calculation of IR and Raman spectra.

4.3

Consideration of Experimental Conditions

The third point, the inclusion of experimental conditions, comprises first of all the description of solvent and temperature effects. Solvent effects are essentially a modification of the electronic potential (and property) energy hypersurface and have therefore been addressed in Sect. 3.

Quantum chemical calculations of this hypersurface are usually based on equations valid for a temperature of 0 K, i.e., by excluding all nuclear motion effects. In many cases, the results obtained are close to those which are measured at higher temperature (e.g., room temperature). As far as temperature effects are concerned, they can be included into the intensity expressions by assuming a Boltzmann distribution of the molecules over initial vibrational energy levels. Furthermore, if a molecule has several conformations which are so close in energy to the ground state that at experimental temperatures they are non-negligibly populated in the sample, the influence of the temperature onto the vibrational spectra may be incorporated by calculating spectra for every relevant conformation and superimposing them with weighting factors calculated by assuming a Boltzmann distribution of the different conformations. The relative energies of the molecular conformations may of course be changed by solvation.

Another feature which is affected by the environment is the line broadening of spectral peaks. In contrast to MD simulations (see, e.g., [66]), this information is not obtained from a static quantum chemical calculation. It may be modelled, however, by plotting Gaussian or Lorentzian functions centered at the calculated vibrational frequencies and using an appropriate halfwidth (which may be in the range of 8 to 30 cm^{-1} , for example).

4.4

Efficient Calculation of Intensities for Large Molecules

The main focus of this work is on the efficient calculation of the relevant parts of vibrational spectra obtained by special techniques which filter out information compared to conventional IR and Raman spectroscopy. On the basis of the approximations outlined above, vibrational absorption and scattering intensities are usually calculated from derivatives of components of molecular property tensors with respect to mass-weighted normal coordinates. This applies to traditional IR and Raman spectroscopy as well as to many special techniques such as difference IR or ROA. These derivatives along mass-weighted normal modes are commonly obtained by calculating all cartesian derivatives of the property of interest and then transforming them to a normal mode basis. This automatically yields derivatives along *all* normal modes of the molecule. One is often interested, however, in intensities associated with a few characteristic vibrations or within a certain frequency

range only. In such cases, direct differentiation along selected mass-weighted normal modes allows for an efficient calculation of the relevant derivatives alone [67]. It furthermore allows for restricting the calculation of the normal modes to the interesting ones, in particular by employing the mode-tracking protocol described in Sect. 2.

The numerical derivative of any molecular property A (like, e.g., the polarizability or the electric dipole moment) along the mass-weighted normal mode $Q_p^{(m)}$ is given by [67]

$$\frac{\partial A}{\partial Q_p^{(m)}} = \frac{A(R_{eq} + s_R R_p^{\text{norm}}) - A(R_{eq} - s_R R_p^{\text{norm}})}{2s_{Q_p} |\Delta Q_p^{\text{norm}}|}, \quad (24)$$

where s_R is the step-size in (non-mass-weighted) cartesian coordinates, and s_{Q_p} is a dimensionless normal-mode specific step size equal to

$$s_{Q_p} = s_R \left[\sum_{i=1}^{3M} (Q_{pi}^{\text{norm}})^2 / M_i \right]^{-1/2} \left[\frac{[\text{unit of length}]}{[\text{unit of mass}]^{1/2}} \right], \quad (25)$$

and the superscript “norm” indicates normalized coordinates.

Whereas the calculation of the cartesian property derivatives can be done either analytically or numerically (both possibilities have been implemented in various quantum chemical programs), the direct differentiation along mass-weighted normal modes has in practice to be carried out numerically. In principle, analytical derivatives can be calculated faster than numerical ones, and some authors argue that the lacking implementation of analytical derivatives for certain special techniques such as ROA is a serious drawback for first-principles calculations of the corresponding intensities (see e.g., [68]). This may be true for small molecules, but in the case of many comparatively large biomolecules, the *selective* numerical differentiation has the opposite effect, since it enhances the computational efficiency and renders larger molecules accessible to calculations. In addition, these numerical methods are facile to implement and offer easy parallelization and restart options. These advantages are, in principle, all available in pure analytical schemes, but their development is rather complicated and tedious as analytic electronic energy derivatives are to be derived for various electronic structure methods.

Another possibility of making the prediction of vibrational spectra more efficient is to employ computational methods of different quality for the normal modes and the molecular property tensors. Since molecular polarizabilities, especially the generalized ones needed for the calculation of ROA spectra, require a higher computational effort than the molecular energies, it may speed up the calculation of Raman and ROA spectra considerably to use a small basis set for the polarizabilities. It turned out that Raman and ROA intensities can be calculated with reasonable accuracy within Hartree–Fock

and density functional theory by employing surprisingly small, tailored basis sets for the property tensors in combination with more accurate normal modes [69, 70]. This is due to error cancellations when calculating the derivatives of the property tensors—although the absolute values may be wrong, their dependence on nuclear coordinates may still be reproduced correctly when employing numerical differentiation (see [60] for a discussion of this effect in the context of the electric dipole–electric dipole polarizability tensor).

5

Selective Vibrational Spectroscopic Techniques

In this section, selective vibrational spectroscopic techniques are shortly introduced, and principles and applications of the corresponding quantum chemical intensity calculations are summarized.

5.1

Difference-IR and -Raman

Difference-IR and -Raman spectra can be obtained by recording a series of conventional IR or Raman spectra during the course of a reaction and presenting them as intensity differences with respect to the spectrum of the educts. In contrast to techniques such as resonance Raman or VCD and ROA, the selection of certain bands from the full vibrational spectrum is not achieved by filtering out certain scattering or absorption components via the experimental setup, but by plotting the difference of time-resolved IR or Raman spectra. Molecular vibrations involving mainly atoms which either exhibit large displacements along the reaction path, or which are indirectly affected by the reaction through changes in the electron distribution, will show up as positive or negative peaks in the difference spectra, while the bands associated with all other parts of the molecule cancel upon subtraction. Provided the observed bands can be assigned unambiguously to (local) molecular vibrations, difference spectra can thus answer the question which parts of a molecule are involved in a reaction. Furthermore, information on the reaction kinetics and on the change of electron density distribution can be obtained.

Difference vibrational spectroscopy is a comparatively young technique (early time-resolved difference IR spectra have been reported in 1980 [71]), since it requires at the same time a high sensitivity of the experimental setup and a high stability of the measuring conditions. Difference IR spectroscopy has been applied successfully to protein reactions (see [72] for an up-to-date review, [73] for a review of IR spectroscopy of proteins in general including a section on difference IR, and [74] for a very recent application to light absorption by photoreceptors). Difference Raman spectroscopy, besides its

applicability to proteins [75], is particularly suited for the investigation of localized interactions in nucleic acid complexes (for a review, see [76], for an example of its application to DNA recognition by proteins, [77], and [78] for a survey of difference Raman spectroscopy of biomolecules in general).

Vibrational bands can be experimentally assigned to molecular regions by comparison to vibrational spectra of reference systems obtained either by controlled mutation of one selected amino acid residue in peptides, or by site-specific isotope labeling. Because of the large effort associated with assuring that the desired reference compound is synthesized, and of recording the spectra under exactly the same experimental conditions as the original sample, quantum chemical calculations can be valuable for reliable and efficient band assignments.

Although most reactions take place at a limited region of the protein only, interactions with the peptide chain and possible conformational changes in the protein may play a central role during its course. Thus, it may be necessary to include the whole protein structure in a simulation. Difference IR spectroscopy can be applied to proteins of up to 100 kDa, including proteins of the cell membrane. Since systems of that size are far from being accessible by pure first-principles methods, QM/MM-calculations are of crucial importance. QM/MM simulations of vibrational spectra of biomolecules focussed, for instance, on bacteriorhodopsin [79–81], the GTPase Ras [82, 83], and CO binding in myoglobin [84].

For biological systems which shall be modelled as close to their natural state as possible, the choice of a molecular structure as a starting point of the vibrational analysis is not obvious. In principle, the calculation of vibrational frequencies and normal modes within the harmonic approximation requires that the molecule under study is in an energy minimum with respect to all nuclear degrees of freedom. But a geometry optimization of the whole system might lead to a structure which is far away from the one adopted in solution or during the reaction. However, the harmonic approximation can also be justified for partly optimized structures if only those modes are interpreted which are centered on the optimized molecular subsystem [21]. Therefore, it is possible to perform a constrained optimization for only the QM subsystem starting from an X-ray structure (as done in [84]) or from a snapshot taken from a preceding MD simulation (as in [82]), and to use it for a subsequent vibrational analysis. This vibrational analysis is usually performed by simply diagonalizing that part of the QM/MM Hessian matrix which contains the QM atoms [66, 82, 84], since the calculation and diagonalization of the whole QM/MM Hessian matrix would imply an enormous computational effort producing a lot of useless information. In principle, it is also possible to calculate selectively the exact normal modes of the QM subsystem only from the complete QM/MM Hessian matrix by employing the mode-tracking protocol. This approach will be applied to myoglobin in a separate study to be published by the authors. It has the advantage of yielding normal modes

of the whole QM/MM system, whereas the diagonalization of parts of the Hessian matrix would entail an uncontrollable loss of accuracy due to the neglected coupling of the QM system to the MM system. Molecular dipoles and their derivatives have been implemented into QM/MM schemes [85], so that IR intensities can be calculated within a QM/MM approach. The simplest computational approach to vibrational difference spectra is then to calculate the spectra of educts and products separately and to subtract them. Furthermore, since IR difference spectra can be recorded with a time resolution down to a few ns or even further if the step scan technique is applied [72], it is possible to extract pure difference IR spectra of intermediate products on the reaction path using mathematical techniques such as multivariate curve resolution [86] or factor analysis combined with singular value decomposition [87], both of which do not rely on special assumptions on the nature of the reaction. By comparison of these extracted intermediate spectra to simulated ones for assumed intermediate structures, very detailed information on the reaction path should be obtained.

5.2

Two-Dimensional Techniques

Two-dimensional vibrational spectroscopy is based on the idea of recording IR (or Raman) spectra of molecules in selectively excited vibrational states. Because of the anharmonicity of the PES and vibrational couplings, positive and negative intensity differences arise with respect to the ordinary one-dimensional IR (or Raman) spectrum. There are two different experimental setups, which yield similar two-dimensional spectra [88]: double-resonance spectroscopy (or dynamic hole burning), and pulsed Fourier transform spectroscopy (or heterodyne-detected photon echo spectroscopy). Both are compared from a theoretical as well as from an experimental point-of-view in [88]. In the former approach, spectra are obtained by exciting the sample with a laser beam (pump pulse) with a narrow frequency range and recording a Fourier Transform-IR spectrum with a second laser beam (probe pulse) covering the whole frequency range of interest. By recording such spectra for different frequencies of the pump pulse, a two-dimensional IR spectrum with two frequency axes is obtained. Pulsed Fourier transform spectroscopy consists of sending three consecutive ultrashort laser pulses with defined time delays through the sample and Fourier transforming the outgoing signal with respect to the delay times between the first and the second pulse and the time after the third pulse. Because of the tunable time-delays with a time-resolution down to picoseconds, two-dimensional IR (2D-IR) spectroscopy is particularly suited for the investigation of vibrational relaxation dynamics.

The first 2D-IR experiment on biomolecules was performed by Hamm et al. in 1998, employing the double-resonance variant for the investigation

of amide I bands in peptides [89]. Among others, the particular value of two-dimensional IR spectroscopy for peptides is emphasized by a recent experimental study of a model octapeptide, indicating that 2D-IR spectroscopy allows for a discrimination between α -helical and 3_{10} -helical secondary structures [90], which is difficult to achieve by other spectroscopic techniques. (The introduction of [90] also provides a good overview of 2D-IR spectroscopy of peptides with many literature references. For a review of this topic, see e.g., [91].) 2D-IR may furthermore give valuable insight into the tertiary structure of peptides [89, 92]. For an overview of its methodology and applications, see e.g., [93–95], and [96] for a conceptual comparison to multidimensional NMR spectroscopy.

A second dimension can also be introduced into the IR spectrum by combining it with a UV pump pulse (IR/UV double resonance spectroscopy). [97] provides an application to the investigation of the secondary structure of various gas-phase isolated peptides. In contrast to 2D-IR, however, this method is restricted to molecules which contain a chromophore and will not be discussed further for the sake of brevity. Two-dimensional Raman [98] and resonance Raman [99] spectroscopy is also possible, but has so far been applied primarily to the investigation of liquids and small molecules, respectively, rather than to large biomolecules in solution.

Two-dimensional IR spectroscopy solves the problem of interpreting IR spectra with many overlapping bands not by filtering out some of these bands and thus reducing information, but by spreading out the information onto a larger space, which is provided by two wavenumber axes instead of one. In particular, broad amide I bands which appear structureless in the one-dimensional IR spectrum, may be split into several cross-peaks in the two-dimensional spectrum [100]. In addition, tunable filtering mechanisms are available in the form of the polarization of the pump and probe laser beams. By carefully choosing these polarizations, diagonal peaks may be completely suppressed, so that only off-diagonal coupling remains [101].

2D-IR spectra consist of pairs of positive and negative peaks. Calculations of their intensities often focus on the contribution of anharmonicity of the potentials of the vibrational modes, whereas the contributions of electrical anharmonicity and level-dependent dephasing dynamics are neglected. A comparative study of these three contributions is provided by [102]. The diagonal peaks arise because of the anharmonicity of the vibrational modes (in Fig. 2, this results in a difference in length of arrows 1 and 2: if transition 1 is excited by the pump pulse, the IR spectrum recorded with the probe pulse will show a “hole” (i.e., a negative peak) in comparison to the one-dimensional IR spectrum at the wavenumber corresponding to transition 1, and an additional positive peak at the wavenumber corresponding to transition 2, which due to the anharmonicity is located at slightly lower wavenumbers than the negative peak). Off-diagonal peaks are due to couplings between different vibrations (as can be seen from the difference between arrows 3 and 4).

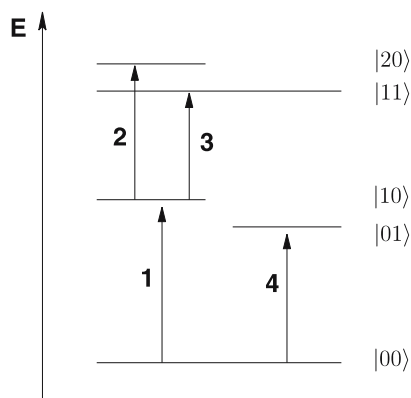


Fig. 2 Qualitative energy levels of an anharmonic oscillator coupled to another one

For a more detailed and sophisticated explanation see e.g., [103]. If all vibrations were strictly harmonic, so that all vibrational levels were even-spaced and no coupling would arise, no peaks would be observed in the 2D-IR spectrum, since positive and negative contributions would appear at the same positions and cancel exactly. In order to describe 2D-IR spectra, it is thus necessary to go beyond the (mechanical) harmonic approximation. When anharmonic contributions are considered, the potential energy surface in terms of mass-weighted normal coordinates takes the form

$$E_{e,i}(\mathbf{Q}^{(m)}) = \frac{1}{2} \sum_{I=1}^{3M} \left(\frac{\partial^2 E_{e,i}}{\partial^2 \mathbf{Q}_I^{(m)2}} \right)_0 \mathbf{Q}_I^{(m)2} \quad (26)$$

$$+ \frac{1}{6} \sum_{I=1}^{3M} \sum_{J=1}^{3M} \sum_{K=1}^{3M} \left(\frac{\partial^3 E_{e,i}}{\partial \mathbf{Q}_I^{(m)} \partial \mathbf{Q}_J^{(m)} \partial \mathbf{Q}_K^{(m)}} \right)_0 \mathbf{Q}_I^{(m)} \mathbf{Q}_J^{(m)} \mathbf{Q}_K^{(m)} + \dots$$

The coupling between different normal coordinates is introduced via the cubic (and higher-order) force constants $\left(\partial^3 E_{e,i} / \partial \mathbf{Q}_I^{(m)} \partial \mathbf{Q}_J^{(m)} \partial \mathbf{Q}_K^{(m)} \right)_0$. Anharmonic force constants up to sixth order with up to three different normal mode indices have been included into predictions of 2D-IR spectra of acetic acid dimers based on DFT calculations [104]. The calculation of fundamental frequencies and anharmonic vibrational constants may be achieved either by vibrational self-consistent field methods or by vibrational perturbation theory (see [105] for a comparison of both). For large molecules, however, the perturbation theory approach may fail due to the large number of anharmonic resonances. The vibrational self-consistent field approach, on the other hand, requires a significantly larger computational effort, so the problem of including mechanical anharmonic effects accurately in calculations of large molecules remains a difficult if even possible task. Electrical anharmonicity,

on the other hand, is commonly believed to play a minor role and thus usually neglected [104].

The goal of a theoretical description of 2D-IR spectra is to predict the positions, intensities, and shapes of the peaks within a selected wavenumber range [106]. Since 2D-IR is a nonlinear technique, the description of the whole spectrum based on first-principles calculations is very involved (for a detailed account of the theory of nonlinear spectroscopy see [107]). Thus, for realistic systems, model Hamiltonians are used instead, whose parameters may be adjusted based on first-principles calculations. Two-dimensional IR spectra of peptides are usually calculated based on a vibrational-exciton model [108, 109], which describes selected vibrational states as a set of coupled oscillators. The associated coupling constants have been estimated so far either based on semi-empirical models or from the anharmonic force constants obtained from DFT calculations [100, 104, 110, 111], in some cases including a continuum model description of the solvent [112]. It is also possible to use molecular dynamics or Monte-Carlo simulations in order to construct fluctuating vibrational-exciton models [113]. Two different variants of this technique have been compared and applied to a 17-residue peptide in [114]. Results from MD simulations have also been compared to structural and dynamical parameters for the subpicosecond conformational dynamics of small peptides obtained from 2D-IR experiments [115].

Because of the restriction of the exciton model to a certain class of vibrations (such as amide I vibrations in peptides), the calculations are automatically selective for a certain subset of vibrational modes.

5.3

Techniques Sensitive to Chirality: VCD and ROA

Vibrational Circular Dichroism (VCD) and Vibrational Raman Optical Activity (ROA) spectra are in principle recorded in the same way as standard IR and Raman spectra, respectively, but instead of total intensities, the intensity differences between right and left circularly polarized light are plotted. These differences are zero for achiral molecules and in general nonzero for chiral molecules. Thus, information related to the chirality of the molecule under study is filtered out from the vibrational spectrum. As far as the basic idea is concerned, the two techniques are similar to electronic circular dichroism (CD), where absorption intensity differences between right and left circularly polarized light in the frequency region of electronic transitions are recorded. Advantages compared to CD are, however, that VCD and ROA allow for a determination of the absolute configuration of molecules which do not contain a chromophore and are thus inaccessible to CD spectroscopy. Also, this determination can be more reliable since it is based on a larger number of peaks. The ability to determine absolute configurations of biomolecules (and pharmaceuticals) is of crucial importance due to the different interactions

of different enantiomers with molecular structures in living organisms. The large number of signals arising from vibrational motions of all parts of the molecule instead of only those involved in electronic transitions furthermore provides more structural information than electronic spectra.

Although the first ROA experiment was carried out in 1973 [116] and thus one year earlier than the first VCD experiment [117], ROA became a routinely applicable technique only recently. This is mainly due to the fact that the scattering intensity differences are of the order of about 10^{-3} times the total scattering intensity at most, and that total scattering intensities are considerably smaller than usual absorption intensities, so that very sensitive and stable instruments are needed. Apart from this, ROA outperforms VCD in two ways: First, it can be recorded in the natural environment of most biomolecules, namely in aqueous solution, and second, ROA spectra can be recorded down to lower wavenumbers ($\approx 100 \text{ cm}^{-1}$ [118]) than VCD spectra ($\approx 600 \text{ cm}^{-1}$ [119]).

VCD has primarily been used for the structure determination of peptides and proteins, but also for nucleic acids and monosaccharides (for reviews, see e.g., [120–122]). The range of compounds investigated by ROA so far [123, 124] is similar to the one of VCD, except that it comprises also disaccharides [125] and even intact viruses [126]. Because of the sensitivity of the intensity differences to conformational changes, both techniques are particularly suited for the investigation of protein secondary structures. ROA, e.g., allows for a discrimination between helical and β -sheet peptide structures based on empirical correlations between assumed protein structures and characteristic signal patterns [127, 128]). Furthermore, experimental data suggest that ROA might provide information on tertiary structures [123, 124]. For a comparison of the configurational and conformational sensitivity of VCD and ROA, see [129] and for a general quantitative comparison of both techniques [119].

In order to improve and verify empirical correlations between structure and spectral intensities, and in particular in order to make unambiguous assignments of absolute configurations possible, quantum chemical calculations are needed for both VCD and ROA. Whereas the development of a quantum chemical ROA theory, whose basic equations were published in 1971 [130], preceded the consistent formulation of the theory of VCD in 1983/1985 [131, 132], the number of VCD calculations published so far is much larger due to the faster experimental development of this technique. An excellent description of the theoretical background of ROA and of the interaction of chiral molecules with radiation in general can be found in [64]. For a short summary of theoretical background, current experimental instrumentation and selected applications of ROA see [133], and on theoretical and experimental vibrational optical activity in general, emphasizing the need of first-principles calculations see [134]. Furthermore, [135] provides a survey on ROA (and Raman) spectroscopy both from an experimental and a theoretical point-of-view.

It is necessary to outline the theoretical background briefly in order to provide a basis for the discussion of special problems related to the calculation of VCD and ROA spectra for biomolecules. For computational details, we refer to the reviews mentioned above and to the literature cited in the following.

5.3.1

VCD Theory

Absorption intensity differences between right and left circularly polarized light in an isotropic sample are proportional to the isotropic rotational strength $R_{f \leftarrow i}$, which can be calculated from the Rosenfeld equation [136],

$$R_{f \leftarrow i} = \text{Im} \left(\langle i | \boldsymbol{\mu} | f \rangle \langle f | \mathbf{m} | i \rangle \right). \quad (27)$$

This equation describes vibrational as well as electronic circular dichroism, depending on whether the initial and the final state $|i\rangle$ and $|f\rangle$ differ only in their vibrational or also in their electronic state. The evaluation of the electric transition dipole matrix element $\langle i | \boldsymbol{\mu} | f \rangle$ is known from IR theory, so the calculation of VCD spectra focusses on the magnetic transition dipole matrix element $\langle f | \mathbf{m} | i \rangle$. For nonmagnetic (closed-shell) molecules, the electronic contribution to $\langle f | \mathbf{m} | i \rangle$ is zero for matrix elements between nondegenerate electronic states within the Born–Oppenheimer approximation, which was the reason for the “delay” in the formulation of the theory of VCD compared to ROA. By replacing the concept of nuclear-coordinate dependent energy and property surfaces by the idea of velocity dependent ones, however, it is possible to formulate computable expressions for $\langle f | \mathbf{m} | i \rangle$ based on Born–Oppenheimer wave functions [131, 132],

$$\langle f | \mathbf{m}(\dot{\mathbf{Q}}) | i \rangle = \sum_p \langle f | \dot{\mathbf{Q}}_p | i \rangle \left(\frac{\partial \mathbf{m}}{\partial \dot{\mathbf{Q}}_p} \right)_0, \quad (28)$$

where the matrix element $\langle f | \dot{\mathbf{Q}}_p | i \rangle$ can be calculated within the (mechanical) harmonic approximation from the harmonic oscillator model. The derivative of the magnetic dipole moment along the time derivative $\dot{\mathbf{Q}}_p$ of the normal mode $\mathbf{Q}_p$ can be expressed by using the magnetic field perturbation method [132, 137, 138] as

$$\left(\frac{\partial \mathbf{m}}{\partial \dot{\mathbf{Q}}_p} \right)_0 = -2\hbar \text{Im} \left\langle \frac{\partial \Psi_0}{\partial \mathbf{B}} \left| \frac{\partial \Psi_0}{\partial \mathbf{Q}_p} \right. \right\rangle. \quad (29)$$

Both wave function derivatives can be calculated either analytically (employing coupled HF (or KS) perturbation theory [139]) or numerically [140]. Equation 29 has been the basis for numerous computational studies on VCD. A summary of the excellent performance of DFT for that purpose is provided by [141], and one on the use of theory to deduce absolute configurations from experiment by [142].

5.3.2 ROA Theory

The derivation of the relevant formulas for the ROA intensity differences is essentially the same as in Raman spectroscopy, except that in addition to the electronic dipole moment, the contribution of the molecular electronic quadrupole and magnetic dipole moments interacting with the radiation to the induced multipole moments have to be taken into account [64, 130, 143, 144]. These three induced multipole moments can be expressed as a function of the electric field amplitude of the incident radiation, and of several complex generalized polarizability tensors. The latter are the (“ordinary”) electric dipole–electric dipole polarizability α , the electric dipole–magnetic dipole polarizability G' (the prime indicating imaginary parts of complex quantities), and the electric dipole–electric quadrupole polarizability A . Depending on whether the right-handed or the left-handed enantiomer is under the influence of radiation with a certain polarization state, the magnitude of the induced multipole moments and thus the intensity of the scattered radiation in a given polarization state will be slightly different.

The next step is to calculate this intensity of the scattered radiation from the induced multipole moments, which can be done within the semi-classical approach based on equations known from classical electrodynamics. Whereas in VCD only one experimental setup is of importance, in ROA different arrangements are possible, which are distinguishable by the polarization of the incident and detected radiation and, analogous to conventional Raman spectroscopy, by the angle to the incident light beam in which the scattered radiation is detected [64, 133]. For each setup, different mathematical expressions for the calculation of the corresponding intensity differences are obtained. Because of the good signal-to-noise ratio and the low susceptibility to measuring artifacts, the backscattering variant of ROA is employed in modern instruments. Therefore, the corresponding expression for the intensity difference between right and left circularly polarized light relative to the total scattered intensity, obtained by applying the principles mentioned above and employing Placzek’s approximation, is given as an example,

$$\Delta(180^\circ) = \frac{I^R - I^L}{I^R + I^L} \quad (30)$$

$$= \frac{8 \left[3 (\alpha_{\alpha\beta})_{mn} (G_{\alpha\beta}^*)_{mn} - (\alpha_{\alpha\alpha})_{mn} (G_{\beta\beta}^*)_{mn} + \frac{1}{3} \omega (\alpha_{\alpha\beta})_{mn} \epsilon_{\alpha\gamma\delta} (A_{\gamma,\delta\beta}^*)_{mn} \right]}{c \left[7 (\alpha_{\lambda,\mu})_{mn} (\alpha_{\lambda,\mu}^*)_{mn} + (\alpha_{\lambda\lambda})_{mn} (\alpha_{\mu\mu}^*)_{mn} \right]},$$

where summation over repeated indices is implied. The quantities indicated by the subscript mn are transition matrix elements between vibrational states $|\nu_n\rangle$ and $|\nu_m\rangle$ of the generalized polarizability tensors, whose dependence on

the normal coordinates is, as in the theory of conventional Raman scattering, expressed as a Taylor series truncated after the linear term (electric harmonic approximation). The combinations of these transition matrix elements in Eq. 30 also appear in the intensity expressions corresponding to the other experimental setups, so they are usually summarized as Raman and ROA invariants (for details, see e.g., [133] or [64]). Within the electric harmonic approximation, the transition matrix elements in Eq. 30 are proportional to derivatives of the corresponding molecular generalized polarizability tensor along mass-weighted normal modes. Since derivatives of the electric dipole–magnetic dipole polarizability G' along nuclear coordinates have not been implemented in an analytical form yet [68], and analytical derivatives of the electric dipole–electric quadrupole polarizability A have become available only very recently [145], all three derivatives are usually calculated numerically at present. The polarizability tensors themselves are usually calculated employing linear response theory [146]. As in the calculation of VCD spectra, it is necessary to ensure gauge invariance of the intensity differences by employing London atomic orbitals [146, 147], since both the electric dipole–magnetic dipole and the electric dipole–electric quadrupole polarizability depend on the choice of the coordinate origin. This origin dependence is not reproduced correctly otherwise. As far as selection rules are concerned, all Raman active modes are in principle also visible in ROA spectra.

ROA calculations on biomolecules have so far essentially been carried out for peptides and proteins (see e.g., [148], which also gives a survey of the current state-of-the-art of computational methods for predicting ROA spectra). Furthermore, very recent reviews on the quantum mechanical prediction of chiroptical vibrational properties [68] and chiroptical properties in general [149] are available.

5.3.3

Molecular Conformation

ROA [118, 133, 150] and VCD [151, 152] calculations have shown that in conformationally flexible molecules, intensity differences arising from different conformations (e.g., of methyl groups or ring structures) can cancel, so that certain vibrational bands vanish in the ROA spectrum. This can be regarded as an additional mechanism filtering out information, which distinguishes ROA from Raman spectroscopy, where contributions from different conformations are always additive. In order to reproduce this behavior accurately in quantum chemical calculations, it is in principle necessary to calculate the relative energies of all relevant conformations (which may depend strongly on the solvent) and to superimpose their spectra multiplied by weighting factors obtained by assuming a Boltzmann distribution of the conformers [118]. For larger biomolecules, where already the simulation of the spectrum of one conformer is a computational challenge, another (though conceptually

less satisfying) approach is to neglect the intensity differences of all vibrational bands which involve mainly the conformationally flexible groups (in case such a partitioning of modes is possible).

5.3.4

Efficient Computational Procedures for Large (Bio-)Molecules

Whereas calculated VCD intensity differences are not considerably more expensive than the corresponding IR intensities, the calculation of ROA intensity differences is much more computer time-consuming than the calculation of the normal modes, vibrational frequencies and also Raman intensities. Two strategies may be combined in order to overcome this problem. First, the molecular property tensors may be computed more efficiently. Currently, efficient density-fitting techniques like those available for the calculation of the molecular polarizability [153] are not yet available for generalized property tensors. Therefore, calculations of the property tensors are accelerated by employing a smaller basis set than the one used for the vibrational modes and frequencies—for ROA, a specific basis set has been developed which yields excellent results despite its limited size. It comprises diffuse *p*-type functions on hydrogen atoms, which are not contained even in much larger basis sets, but which improve the accuracy of the results dramatically [69]. In order to gain additional efficiency, it has been proposed to neglect the contributions of the electric dipole–electric quadrupole polarizability [154], which are often small, and reduce to zero for a certain experimental setup [133]. This reduces the computational efficiency considerably, at the cost of a (hardly controllable) loss of accuracy. It is also possible to replace the linear response theory by a cheaper, but more approximate sum-over-states formalism [155] for the property calculations, which has yielded good results for oligopeptides. In addition, calculated ROA intensity differences (for a given set of normal modes) have been found to be virtually independent of the chosen density functional, so that calculations of the relevant properties with pure density functionals and thus with the speed-up provided by the resolution-of-the-identity technique [38, 39] are in principle possible [70]. As a second way of improving the computational efficiency needed for large biomolecules, the calculation of intensity differences can be restricted to a small set of vibrational modes. For the assignment of absolute configurations as well as for the investigation of molecular conformations, calculations of intensity differences are not necessarily needed for *all* vibrational bands, but may be sufficient for a few characteristic ones. Intensity differences may be computed selectively for these modes from property derivatives (and additionally wave function derivatives, in the case of VCD) obtained by direct differentiation along normal modes. For ROA, the direct numerical differentiation along normal modes has been implemented and applied to deca-alanine [232]. Which vibrational normal modes are considered to be characteristic may not

be obvious from the beginning, but can be determined by defining a set of modes assumed to have large associated intensity differences based on chemical intuition. This set can be enlarged stepwise if it turns out to be not sufficient for an unambiguous assignment of the molecular configuration. In the case of calculated ROA spectra for peptides, the relevant normal modes chosen will usually be the amide modes (especially amide III), since the intensity differences of these modes are especially sensitive to the backbone conformation. An approximate calculation of VCD spectra of large molecules has also been performed based on the transfer of molecular property tensors in cartesian coordinates, which relates spectra of large molecules with repeated similar subunits (such as peptides) to molecular properties computed for the fragments [29, 34].

5.3.5

Interpretation of Calculated Spectra for Large (Bio-)Molecules

When calculating VCD or ROA spectra of large biomolecules, the possibility that individual peaks may cancel, and the fact that the mechanical harmonic approximation has to be employed, creates a problem with respect to the interpretation of the computed spectra. The intensity differences of closely lying vibrational peaks may have different signs, and due to the mechanical harmonic approximation as well as the neglect of solvent effects, the corresponding wavenumbers may easily be shifted by 20 cm^{-1} or more. In contrast to small molecules, where the individual peaks are sufficiently separated from each other, the high density of peaks in the spectra of large molecules may lead to cancellations of peaks in the experimental spectrum which can turn into a sequence of a positive and a negative peak in the calculated spectrum, and vice versa. Since the harmonic approximation cannot be avoided in calculations of vibrational spectra of large molecules, especially ROA spectra calculated by first-principles methods will have difficulties to predict the detailed structure of a given spectral region (such as amide III) with certainty, but only the intensity difference integrated over the whole region is likely to be characteristic [232]. This may be accounted for by plotting ROA spectra in a band-averaged way in addition to a Lorentzian (or Gaussian) peak broadening (Fig. 3). How the relevant information can be extracted from calculated ROA spectra of large molecules is considered in detail by [232].

5.3.6

Solvent Effects

As far as the calculation of VCD and ROA spectra is concerned, solvent effects can become even more important than for techniques which are not sensitive to chirality, since not only the relative magnitudes, but also the signs of vibrational bands may change in different solvents [156, 157], either indirectly

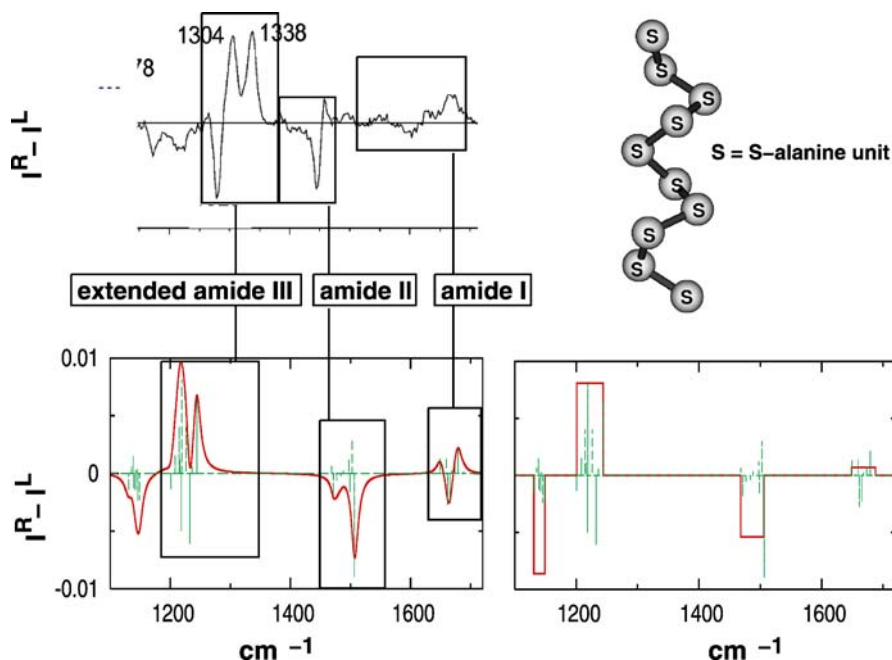


Fig. 3 ROA spectroscopy of helical all-S deca-alanine. *Top left*: Experimental ROA spectrum of poly-L-alanine in 30% dichloroacetic acid/70% CHCl_3 (reproduced from data presented in [127]). *Top right*: Symbolical representation of (all-S)-deca-alanine. *Bottom*: Selectively calculated ROA intensity differences of (all-S)-deca-alanine, at an excitation wavelength of 488.80 nm. *Bottom left*: Lorentzian intensity distribution, half-width of peaks 15.0 cm^{-1} . *Bottom right*: Band-averaged ROA intensity differences. Individual peaks have been included as a line spectrum scaled by 0.04. Calculated spectra were reproduced from data given in [232]. Intensity differences in $\text{\AA}^4 \text{ a.m.u.}^{-1}$; normal modes and frequencies: DFT(BP86/RI), TZVP; generalized polarizability tensors: DFT(BLYP), Zuber and Hug's rDSP:3-21

due to changes in the conformational equilibrium or directly through solvent-induced changes in the electronic structure of the solute and its dependence on the nuclear coordinates. In [156] for example, solvent effects on VCD spectra of proline, modeled by a combination of the integral equation formulation of the polarizable continuum model (PCM) and explicit water molecules, are found to influence the relative energies of different conformations considerably, whereas the structural parameters are quite invariant to solvent effects. A complementary study has been carried out for ROA [157], employing the same PCM version but no explicit solvent molecules, which points out the importance of considering the solvent influence on both the vibrational normal modes and the property tensors. It has been reported in previous studies that the explicit consideration of water molecules in the first solvation shell of solvated alanine improves ROA results considerably in combination with an

Onsager continuum model [158]. Further investigations on the relative performance of continuum and explicit solvent models for the prediction of ROA spectra are in preparation and will be published by the authors in a separate study. Furthermore, DFT investigations combined with the PCM on VCD spectra have shown the need for better quantitative accuracy in the predicted IR and VCD intensities [151, 152], which suggests that solvent effects may have to be considered more accurately in some cases than is possible with the PCM. In particular when hydrogen bonds between solvent and solute are important, explicit solvent molecules will, in general, be necessary. The larger the number of solvent molecules, the more important mode-selective techniques will become for the feasibility of such studies.

5.3.7

Local Contributions

Additional insight into the generation of VCD and ROA signals may be gained from assessing and comparing the contributions from different parts of the molecule under study. Local contributions to VCD intensity differences, which may be visualized as a continuous distribution over the whole molecule in the form of an electron transition current density, have been derived by Freedman, Nafie and coworkers [159–162]. A scheme to decompose ROA and Raman intensities into contributions from individual pairs of atoms (which may be grouped to pseudo-atomic contributions with a certain arbitrariness comparable to electron population analysis) has been presented by Hug [163, 164]. It reveals that long-range couplings may play an important role in ROA spectroscopy [165]. The possibility of local decomposition is a great advantage of first-principles calculations over data gained from experimental work, since it provides information on where signals are generated and thus might finally lead to a set of rules for the prediction of VCD and ROA spectra.

5.4

Raman Techniques in Electronic Resonance

Traditional as well as special Raman scattering techniques have a resonance variant, where the energy of the incident light is close to an electronic transition energy of the molecule under study. This usually results in the amplification of a few peaks and thus to a simplification of the vibrational spectrum.

5.4.1

Spontaneous Resonance Raman Scattering

Whereas Raman intensities far from electronic resonance depend mainly on molecular ground-state properties, resonance Raman spectra contain in-

formation on those electronically excited states whose excitation frequencies are close to the frequency of the incident laser beam (usually only one or a few; for a qualitative scheme of the former case see Fig. 4). From a simplified point-of-view, the intensities of those vibrational modes along which the equilibrium structure is shifted most upon electronic excitation are largely amplified in comparison to ordinary Raman spectroscopy. Thus, resonance Raman spectroscopy is of particular interest in order to filter out those vibrational modes which are located on a chromophore in large biomolecules [166–168]. Its importance for the investigation of biomolecules is highlighted, e.g., in a special issue on the structure and dynamics of biomolecules probed by (mainly resonance) Raman spectroscopy in *J Raman Spectrosc* 2005:36(4). For a review of the applications of resonance Raman and other Raman techniques to problems relevant for the life sciences see [169]. Resonance Raman spectroscopy has been applied to enzymes whose active site is a chromophore such as heme proteins (for a review on structural and functional properties of hemoglobins revealed by resonance Raman spectroscopy see e.g., [170]), where the vibrational frequencies are sensitive to the oxidation and the spin state as well as to the coordination number of the central iron porphyrin group [171–175], and to redox-active tyrosines in various metalloenzymes [176]. Furthermore, it has been used for the resonance enhancement of vibrations of the peptide backbone [177, 178] and of aromatic side chains [179] in order to elucidate their conformation and thus the secondary structure. Resonance Raman spectroscopy can also be em-

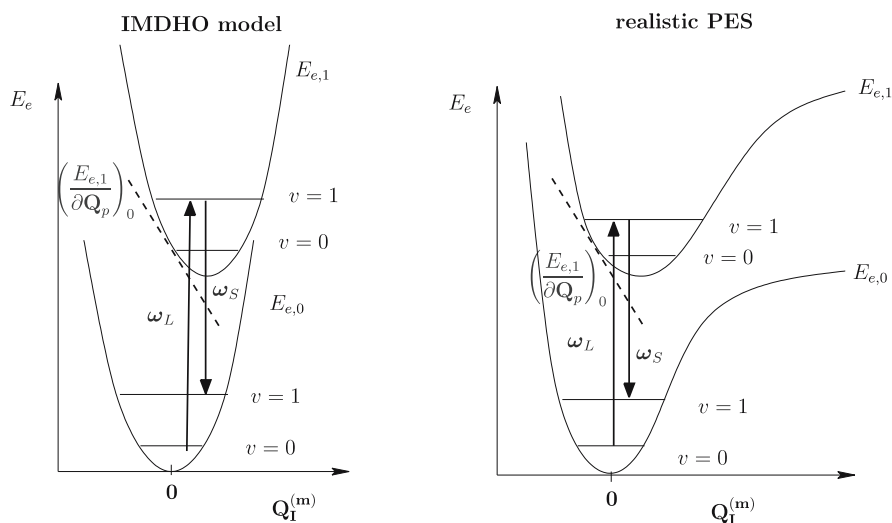


Fig. 4 Schematic representation of resonance Raman scattering for one electronically excited state in resonance (IMDHO = independent mode, displaced harmonic oscillator model)

ployed as a time-resolved technique down to subpicosecond resolution, e.g., for the investigation of the excited-state conformational dynamics of nickel porphyrins [180] or the dynamics of nitric oxide reductases and hemecopper oxidases [181, 182].

The calculation of resonance Raman intensities starts from the same expression for the transition polarizability obtained from time-dependent perturbation theory as the treatment of nonresonant Raman scattering (Eq. 19). However, since the energy difference between transitions to electronically excited molecular states and the incident radiation is not large as assumed in the derivation of the Placzek approximation [62], this drastic simplification has to be replaced by better suited ones. If the frequency difference between the incident radiation and an electronic transition becomes small, one or a few denominator(s) of the terms on the left-hand side of Eq. 19 will approach zero and the corresponding terms will dominate over the others, which consequently can be neglected. Following the same line of reasoning, the terms on the right-hand side of Eq. 19 can also be neglected since they will all be considerably smaller than those dominant terms. The sum therefore reduces to one or a few terms which need to be evaluated. In practice, mainly two approaches are employed: the transform theory as presented by Peticolas and Rush [183] based on Kramers–Kronig relations between the optical absorption and the frequency-dependence of the dispersive part of the polarizability [184], and a time-dependent treatment from which a simple short-time approximation can be deduced [185–187]. The performance of the two approaches as well as the limitations of the underlying approximations have been assessed in detail in [188]. Both have been found to yield comparable results for various molecules [188, 189]. On the basis of the assumptions that (1) the Born–Oppenheimer and (2) the (mechanical) harmonic approximation are valid, that (3) only one excited electronic state s is in resonance, that (4) only Franck–Condon-type scattering occurs, and that (5) the excited state normal modes differ from their ground state counterparts only in a shift of their equilibrium position (independent mode, displaced harmonic oscillator (IMDHO) model), transform theory predicts the resonance Raman intensity of a normal mode Q_p as

$$I_{Q_p} = \omega_L \omega_S^3 |\mu_{0s}^{\text{el}}|^4 \Delta_{q_p}^2 |\Phi(\omega_L) - \Phi(\omega_L - \omega_p)|^2. \quad (31)$$

ω_L and ω_S refer to the angular frequencies of the incident and the scattered (Stokes-)laser beam, respectively, ω_p to the vibrational angular frequency associated with the normal mode Q_p , $|\Phi(\omega_L) - \Phi(\omega_L - \omega_p)|^2$ is a Kramers–Kronig scaling factor which may be obtained from the (either calculated or measured) absorption spectrum, and $\Delta_{q_p}^2$ denotes the displacement of the equilibrium structure of the excited state s relative to the ground state along the dimensionless normal mode q_p , which is related to the mass-weighted

normal mode Q_p as $q_p = Q_p \sqrt{2\pi c \omega_p / \hbar}$. The short-time approximation is based on considering scattering as connected to ground state vibrational wave packets with short life-times propagating on the excited state energy PES [186, 187]. It describes the intensities associated with preresonant Raman scattering as

$$I_{Q_p} = \omega_p^2 \Delta_{q_p}^2. \quad (32)$$

Within the IMDHO model, and assuming that only short-time dynamics of the nuclei are important, the excited-state displacements along dimensionless normal modes may be estimated from the excited state gradients with respect to the dimensionless normal modes,

$$\left(\frac{\partial E_{e,s}}{\partial q_p} \right)_0 = -\omega_p \Delta q_p. \quad (33)$$

The excited state gradients can be efficiently calculated within TDDFT [190–194]. It is obvious from both methods that resonance Raman scattering is largely determined by how much the vibrational normal mode under consideration is shifted in the excited electronic state compared to the ground state. For theoretical details on both approaches, see [188], and for a recent comparison of experimental resonance Raman spectra of large biomolecules to calculated ones see a study on Ruthenium complexes to be published by the authors of this paper and [195, 196]. In addition, a more general version of the Placzek approximation polarizability has been proposed based on the short-time approximation [197–201], which reduces to the conventional Placzek approximation in the nonresonant case and which has the advantage of automatically including all possible excited states. It has recently been applied to uracil as a test molecule and found to compare well with experiment as well as to the short-time approximation based on excited state gradients for this system [202].

The computational effort associated with calculations based on transform theory is in general much higher than for those relying on the short-time approximation. Both require a ground-state structure optimization and a subsequent calculation of the ground-state normal modes and frequencies, but in the case of transform theory, an additional structure optimization in the relevant excited state(s) is needed, whereas in the case of the short-time approximation, a comparatively fast computation of the excited state gradients at the ground-state equilibrium structure is sufficient.

An alternative approach to the calculation of (pre-) resonance Raman spectra is based on performing CPMD simulations in the ground state as well as in the excited state of interest, and to deduce vibrational frequencies and associated preresonance Raman intensities from the obtained data. This technique has been suggested and applied to a model chromophore of the DsRed protein in [203].

5.4.2

Resonance CARS

In contrast to nonresonant Raman scattering, resonance Raman spectra may suffer from the fluorescence of the sample, since the frequency of the scattered Stokes radiation (which is usually detected instead of the Anti-Stokes radiation, since the latter is much less intense) may be close to the frequency of the radiation emitted by the fluorescent species. This problem may be overcome, however, by employing resonance Coherent Anti-Stokes Raman (CARS) scattering (see [204] for a general review). Although CARS can be principally used in as well as out of electronic resonance, only its resonance form is treated here, because it is the more widely used version of both for the investigation of biomolecules. In contrast to spontaneous Raman scattering, where the phase of the molecular vibrations of each molecule is arbitrary, the radiation scattered by different molecules in CARS has a fixed phase relation. This is achieved by using a set of coherent intense laser pulses (which have to be intense enough to make nonlinear effects observable, and short enough in order not to decompose the sample despite their high intensity): The molecules are excited simultaneously by a pump beam (with angular frequency ω_L , corresponding to a molecular electronic transition for resonance CARS) and a Stokes beam (with angular frequency $\omega_S < \omega_L$). The result is a set of molecules in one or a few selected vibrational state(s), vibrating with a phase determined by the two exciting beams. These artificially excited molecules are probed in principle like in conventional Raman spectroscopy, using a probe beam with angular frequency ω_P , usually equivalent to ω_L , and detecting an Anti-Stokes radiation beam with angular frequency $\omega_{AS} = \omega_L + \omega_P - \omega_S$. When the difference between the angular frequencies of the two incident beams, $\omega_L - \omega_S$, approaches the angular frequency of a molecular vibrational transition ω_M , the intensity of the coherent scattered radiation will greatly enhance. Thus, by tuning the frequency of the incident laser beams, one or a few molecular vibrations may be filtered out (provided the duration of the laser pulses is long enough to allow for a sufficient frequency resolution, which usually is the case when the time-scale is around a few picoseconds). This tunable filtering mechanism represents an advantage over the linear selective spectroscopic techniques discussed so far, where the choice of which modes are filtered out is essentially controlled by the properties of the molecule under study. Since the scattered radiation is coherent, its intensity is nonzero essentially only in one direction depending on the directions of the three incident laser beams. This directed signal and the fact that an intense Anti-Stokes signal is recorded allows for a distinction from fluorescence, which is detected at lower wavenumbers [205, 206]. CARS has been applied to various biomolecules, such as porphyrins [207, 208], bacteriorhodopsin [209], and low-frequency modes around 150 cm^{-1} of β -carotene which are assumed to play an important role in the energy transfer in photosyn-

thesis [210], including applications as a microscopic imaging technique [211, 212] and the investigation of living cells [213].

CARS spectra may be recorded with femtosecond time resolution (see [214] for a review) which allows for investigating ultrafast molecular dynamics. Ultrashort laser pulses are also used in the context of CARS microscopy [215]. The broad frequency range of the incident laser beams, which is hardly avoidable when using ultrashort laser pulses, may be overcome by feedback-controlled optimization of pulse shapes [216], or by additionally varying the relative delay between the exciting laser pulses [214], so that mode-selectivity can be achieved. There are at least two ways of measuring time-dependent CARS spectra: First, an initial pump laser may be used, which is followed after a time delay by a CARS experiment where the three CARS pulses are interacting with the sample simultaneously. Second, a time-delay can be introduced in a normal CARS experiment between the simultaneous pump and Stokes pulses on the one hand and the probe pulse on the other hand. For an application of the former technique to the ground state repopulation dynamics via S_1 to S_0 internal conversion of β -carotene see [217], and for a comparison of this technique to the latter one see [218]. Both techniques are compared and reviewed from a theoretical point-of-view in [219]. More references may be found in a special issue on Femtosecond Coherent Raman Spectroscopy (J Raman Spectrosc 2000:31).

A rigorous first-principles calculation of a CARS spectrum has to the best of our knowledge not been reported in the literature so far. Because of its coherent nonlinear nature, the theoretical description of CARS spectroscopy is much more involved than the one of spontaneous Raman scattering (for detailed accounts of the theoretical foundations, see [220–222]). The CARS intensities of a bulk of molecules cannot necessarily be calculated as the orientational-averaged sum of the scattering intensities of all molecules, since the scattered radiation is coherent. The derivation of CARS intensities is thus usually formulated in a more general way than for linear Raman techniques. The incident radiation introduces a polarization in the sample of molecules under study, which depends on the third-order susceptibility of the sample. This third-order susceptibility is not a single-molecule property as in the theoretical treatment of spontaneous Raman scattering, but a bulk property of the sample under study. If first-principles calculations are performed in the context of CARS spectroscopy, they have so far rather been restricted to the molecular equilibrium structure and vibrational normal modes [207], but do not include anisotropies.

It is possible to predict excitation profiles, that is the dependence of CARS intensities on the frequency of the pump laser, either by employing a sum-over-states formalism or transform theory [223]. However, both have not been employed as true first-principles methods yet, but in connection with a set of fitted parameters. In addition, the observed profiles are not necessarily reproduced for all laser wavelengths and vibrational modes [223]. The

challenge of calculating resonance CARS intensities is furthermore illustrated by a study aiming at predicting CARS signals for a sample of dipicolinic acid [224]. This study is based on considering equations of motion of the sample in terms of density matrix elements, coupled to Maxwell's equations, and takes into account the particle density as well as propagation effects of the laser pulses. However, the complicated system under study is reduced to a four-level system and needs estimated radiative decay and dephasing times as an input.

CARS spectra may be converted to the corresponding Raman spectra [206], so in principle, it might seem sufficient to be able to predict resonance Raman spectra. However, accurate first-principles calculations of CARS spectra appear to be valuable nonetheless. The CARS signal is often obscured by interference with the nonresonant background, which may be suppressed by techniques such as polarization-sensitive CARS experiments [207]. It can be helpful to be able to predict this background interference with theoretical methods.

5.4.3

Resonance ROA

Resonance variants have also been introduced for special forms of Raman spectroscopy. Resonance ROA spectroscopy is still at a very early stage of development. The first (and, so far, the only) experiment has been published in 1998 [225], two years after the derivation of the underlying theory [226]. In analogy to the connection between resonance Raman and electronic absorption spectra, the theory of ROA is closely related to the one of electronic circular dichroism. It predicts (and is confirmed by the experiment) that in contrast to ROA spectroscopy far from electronic resonance, resonance ROA signals of chiral molecules all have the same sign, provided that only one excited electronic state is in resonance. In this case, the resonance ROA spectrum is equal to the resonance Raman spectrum scaled by a factor equal to one or one-half the magnitude of the electronic CD anisotropy ratio of the electronic state which is in resonance, depending on the experimental setup. The resonance ROA spectra of two enantiomers are related by a reflection at the wavenumber axis. Altogether, resonance ROA may be regarded as a means to assess whether only one excited electronic state is in resonance (by comparison to resonance Raman spectra recorded with unpolarized light), combined with a detection tool for absolute configurations.

5.5

NRVS

Nuclear Resonance Vibrational Spectroscopy (NRVS) is an exception in the group of spectroscopic techniques included in this review, since it is the only

technique which cannot be applied to biomolecules in solution, but which is restricted to solid samples, measured at very low temperature (typically around 30 K). However, it possesses outstanding selective properties and is not affected by the optical selection rules. NRVS filters out the vibrations which involve a Mössbauer active nucleus, with intensities proportional to the displacement of that nucleus in a certain vibrational normal mode. The most important Mössbauer active isotope from a biochemical point-of-view is ^{57}Fe , so that NRVS is especially useful for heme-containing compounds. The idea of NRVS is to record the vibrational structures of Mössbauer bands. These structures are not observable in conventional Mössbauer experiments, but need a sophisticated experimental setup based on a tunable X-ray synchrotron source, currently allowing a resolution down to $7\text{--}8\text{ cm}^{-1}$. This demanding experimental setup is one of the main reasons why NRVS experiments have become feasible only recently. An up-to-date review on details of the experimental methodology and applications to iron species, comparing experimental results to DFT calculations, is given in [227]. Since NRVS intensities are directly related to the contributions of the iron atom(s) to the vibrational normal modes and the vibrational frequencies [228, 229], only a diagonalization of the Hessian matrix is necessary to calculate them. Recent applications of NRVS to biomolecules have been focussing on myoglobin [229, 230], cytochrome *c* [230], rubredoxin [229], and the low-frequency modes of the FeMo cofactor in intact nitrogenase [231].

6

Conclusion

This review provides a survey of the quantum chemical foundations and applications of first-principles calculations of vibrational spectra of biomolecules. Because of their large size (which can easily be 100–1000 atoms or more), the conventional vibrational spectra of biomolecules tend to contain much more information on a limited space than can be interpreted. Thus, this work focussed on the theory of techniques which filter out information compared to conventional IR and Raman spectroscopy, in particular resonance Raman, difference IR and Raman, two-dimensional techniques, VCD, ROA, CARS, and NRVS. The complexity of the systems under study and the early stage of development of many special spectroscopic techniques makes first-principles calculations a very valuable tool for the interpretation of vibrational spectra.

For the same reason, i.e., their complexity, many biomolecules are not accessible by conventional quantum chemical methods. This requires either to cut out a central fragment and analyze it with an accurate (but computationally expensive) method such as (in the context of large molecules) DFT, to treat the whole molecule with cheaper, but less accurate (e.g., semi-empirical) methods, or to use a combination of both (like QM/MM). A conceptionally

different approach to this problem is to describe the whole biomolecular system with accurate methods, but to restrict the calculation to the relevant information only. Two such methods, mode-tracking to calculate vibrational normal modes and direct numerical differentiation of molecular property tensor along normal modes to obtain the corresponding intensities, were presented and compared to conventional computational schemes.

By selective calculation of the relevant normal modes and intensities only, large biomolecules are made accessible to accurate quantum chemical calculations in the future, and accordingly, solvent effects, whose treatment by continuum models is often not sufficient to model all spectral features (especially in the context of VCD and ROA), may be treated explicitly for much larger systems.

Acknowledgements M.R. is grateful to the Fonds der Chemische Industrie (FCI) for a Dozentenstipendium. C.H. gratefully acknowledges funding by a Chemiefonds-Doktorandenstipendium of the FCI. The authors thank Dr. Johannes Neugebauer for valuable comments on earlier versions of this manuscript, and Dr. Georg Brehm and Priv.-Doz. Dr. Michael Schmitt for helpful discussions.

References

1. Ladd ME, Palmer R (2003) Structure determination by X-ray crystallography. Kluwer Academic/Plenum Publishers, New York
2. Drenth J (1999) Principles of Protein X-Ray Crystallography. Springer, Berlin
3. Wagner G (1997) Nature Struct Biol 4:841–844
4. Clore GM, Gronenborn AM (1997) Nature Struct Biol 4:849–853
5. Hajdu J (2000) Curr Opin Struc Biol 10:569–573
6. Miao J, Hodgson KO, Sayre D (2001) Proc Natl Acad Sci USA 98:6641–6645
7. Miao J, Chapman HN, Kirz J, Sayre D, Hodgson KO (2004) Annu Rev Biophys Biomol Struct 33:157–176
8. Sommerhalter M, Lieberman RL, Rosenzweig AC (2005) Inorg Chem 44:770–778
9. Einsle O, Tezcan FA, Andrade SLA, Schmid B, Yoshida M, Howard JB, Rees DC (2002) Science 297:1696–1700
10. Yang T-C, Maeser NK, Laryukhin M, Lee H-I, Dean DR, Seefeldt LC, Hoffman BM (2005) J Am Chem Soc 127:12804–12805
11. Jalkanen KJ, Jürgensen VW, Claussen A, Rahim A, Jensen GM, Wade RC, Nardi F, Jung C, Degtyarenko IM, Nieminen RM, Herrmann F, Knapp-Mohammady M, Niehaus TA, Frimand K, Suhai S (2005) Int J Quantum Chem 106:1160–1198
12. Marx D, Hutter J (2000) Ab Initio Molecular Dynamics: Theory and Implementation. In: Grotendorst J (ed) Modern Methods and Algorithms of Quantum Chemistry, Proceedings. NIC Series, Vol 3. John von Neumann Institute for Computing, Jülich
13. Born M, Oppenheimer R (1927) Ann Phys 84:457
14. Reiher M, Neugebauer J (2003) J Chem Phys 118:1634–1641
15. Neugebauer J, Reiher M (2004) J Phys Chem A 108:2053–2061
16. Neugebauer J, Reiher M (2004) J Comput Chem 25:587–597
17. Reiher M, Neugebauer J (2004) Phys Chem Chem Phys 6:4621–4629

18. Kaledin AL (2005) *J Chem Phys* 122:184106
19. Reiher M, Neugebauer J (2005) *J Chem Phys* 123:117101
20. Kaledin AL, Kaledin M, Bowman JM (2005) *J Chem Theory Comput* 2:166–174
21. Herrmann C, Reiher M (2006) *Surf Sci* 9:1891–1900
22. Filippone F, Parrinello M (2001) *Chem Phys Lett* 345:179
23. Filippone F, Meloni S, Parrinello M (2001) *J Chem Phys* 115:636
24. Adler TB, Borho N, Reiher M, Suhm MA (2006) *Angew Chem Int Ed* 45:3440–3445
25. Davidson ER (1975) *J Comput Phys* 17:87–94
26. Liu B (1978) In: Moler C, Shavitt I (eds) *Numerical Algorithms in Chemistry: Algebraic Methods*. Lawrence Berkeley Laboratory LBL-8158, Livermore, CA
27. Murray CW, Racine SC, Davidson ER (1992) *J Comput Phys* 103:382
28. Head JD (1997) *Int J Quantum Chem* 65:827
29. Bouř P, Sopková J, Bednářová L, Maloň P, Keiderling TA (1997) *J Comput Chem* 18:646–659
30. Bouř P, Kubelka J, Keiderling TA (2000) *Biopolymers* 53:380
31. Bouř P, Kubelka J, Keiderling TA (2002) *Biopolymers* 65:45
32. Ham S, Cha S, Choi J-H, Cho M (2003) *J Chem Phys* 119:1451
33. Choi J-H, Ham S, Cho M (2003) *J Phys Chem B* 107:9132
34. Choi J-H, Cho M (2004) *J Chem Phys* 120:4383–4392
35. Hohenberg P, Kohn W (1964) *Phys Rev* 136:864–871
36. Kohn W, Sham LJ (1965) *Phys Rev* 140:1133–1138
37. Parr RG, Yang W (1989) *Density-Functional Theory of Atoms and Molecules*. International Series of Monographs on Chemistry, Vol 16. Oxford Science Publications, New York
38. Baerends EJ, Ellis DE, Ros P (1973) *Chem Phys* 2:41
39. Dunlap BI, Connolly JWD, Sabin JR (1979) *J Chem Phys* 71:3396–3402
40. Andrae D, Häussermann U, Dolg M, Stoll H, Preuss H (1990) *Theor Chim Acta* 77:123
41. Tamm I (1945) *J Phys (USSR)* 9:449
42. Dancoff SM (1950) *Phys Rev* 78:382
43. Garavelli M (2006) *Theor Chem Acc* 116:87–105
44. Dreuw A, Head-Gordon M (2005) *Chem Rev* 105:4009–4037
45. Andruniów T, Ferré N, Olivucci M (2004) *Proc Natl Acad Sci USA* 101:17908–17913
46. Sinicropi A, Andruniow T, Ferre N, Basosi R, Olivucci M (2005) *J Am Chem Soc* 127:11534–11535
47. Helms V (2002) *Curr Opin Struc Biol* 12:169–175
48. Reiher M (2003) *Found Chem* 5:23–41
49. Reiher M (2003) *Found Chem* 5:147–163
50. Kirchner B, Reiher M (2005) *J Am Chem Soc* 127:8748–8756
51. Kirchner B (2005) *J Chem Phys* 123:204116
52. Tomasi J, Mennucci B, Cammi R (2005) *Chem Rev* 105:2999–3093
53. Cramer CJ (2004) *Essentials of Computational Chemistry: Theories and Models*. 2nd edn. Wiley, New York
54. Riccardi D, Schaefer P, Yang Y, Yu H, Ghosh N, Prat-Resina X, König P, Li G, Xu D, Guo H, Elstner M, Cui Q (2006) *J Phys Chem B* 110:6458–6469
55. Friesner RA, Guallar V (2005) *Annu Rev Phys Chem* 56:389–427
56. Shurki A, Warshel A (2003) *Adv Prot Chem* 66:249–313
57. Gao J, Truhlar DG (2002) *Annu Rev Phys Chem* 53:467–505
58. Wesolowski TA, Warshel AJ (1993) *J Phys Chem* 93:8050–8053

59. Neugebauer J, Louwerse MJ, Baerends EJ, Wesolowski TA (2005) *J Chem Phys* 122: 094115
60. Neugebauer J, Reiher M, Kind C, Hess BA (2002) *J Comput Chem* 23:895–910
61. Barron LD, Gray CG (1973) *J Phys A* 6:59
62. Placzek G (1934) Rayleigh-Streuung und Raman-Effekt. In: Marx E (ed) *Handbuch der Radiologie*, Vol 6. Akademische Verlagsgesellschaft, Leipzig
63. Born M, Kuang K (1954) *Dynamical Theory of Crystal Lattices*. Clarendon Press, Oxford
64. Barron LD (2004) *Molecular Light Scattering and Raman Optical Activity*. 2nd edn. Cambridge University Press, Cambridge
65. Long DA (2002) *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. Wiley, New York
66. Klähn M, Mathias G, Kötting C, Nonella M, Schlitter J, Gerwert K, Tavan P (2004) *J Phys Chem A* 108:6186–6194
67. Reiher M, Neugebauer J, Hess BA (2003) *Z Phys Chem* 217:91–103
68. Polavarapu PL (2006) *Int J Quantum Chem* 106:1809–1814
69. Zuber G, Hug W (2004) *J Phys Chem A* 108:2108–2118
70. Reiher M, Liegeois V, Ruud K (2005) *J Phys Chem A* 109:7567–7574
71. Siebert F, Mäntele W, Kreutz W (1980) *Biophys Struct Mech* 6:139–146
72. Kötting C, Gerwert K (2005) *Chem Phys Chem* 6:881–888
73. Barth A, Zscherp C (2002) *Quarterly Reviews of Biophysics* 35:369–430
74. Kottke T, Batschauer A, Ahmad M, Heberle J (2006) *Biochemistry* 45:2472–2479
75. Tuma R (2005) *J Raman Spectrosc* 36:307–319
76. Benevides JM, Overman SA, Thomas GJ Jr (2005) *J Raman Spectrosc* 36:279–299
77. Serban D, Benevides JM, Thomas GJ Jr (2003) *Biochemistry* 42:7390–7399
78. Callender R, Deng H (1994) *Annu Rev Biophys Biomol Struct* 23:215–245
79. Hayashi S, Tajkhorshid E, Pebay-Peyroula E, Royant A, Landau EM, Navarro J, Schulten K (2001) *J Phys Chem B* 105:10124–10131
80. Rajamani R, Gao JL (2002) *J Comput Chem* 23:96–105
81. Rousseau R, Kleinschmidt V, Schmitt UW, Marx D (2004) *Angew Chem Int Ed* 43:4804–4807
82. Klähn M, Schlitter J, Gerwert K (2005) *Biophys J* 88:3829–3844
83. Topol IA, Cachau RE, Nemukhin AV, Grigorenko BL, Burt SK (2004) *Biochim Biophys Acta* 1700:125–136
84. Freindorf M, Shao Y, Brown ST, Kong J, Furlani TR (2006) *Chem Phys Lett* 419:563–566
85. Cui Q, Karplus M (2000) *J Chem Phys* 112:1133–1149
86. Ruckebusch C, Duponchel L, Sombret B, Huvenne JP, Saurina J (2003) *J Chem Inf Comput Sci* 43:1966–1973
87. Hessling B, Souvignier G, Gerwert K (1993) *Biophys J* 65:1929
88. Cervetto V, Helbig J, Bredenbeck J, Hamm P (2004) *J Chem Phys* 121:5935–5942
89. Hamm P, Lim M, Hochstrasser RM (1998) *J Phys Chem B* 102:6123–6138
90. Maekawa H, Toniolo C, Moretto A, Broxterman QB, Ge N-H (2006) *J Phys Chem B* 110:5834–5837
91. Ozaki Y, Murayama K, Wu Y, Czarnik-Matusiewicz B (2003) *Spectroscopy* 17:79–100
92. Hamm P, Lim M, De Grado WF, Hochstrasser RM (1999) *Proc Natl Acad Sci USA* 96:2036–2041
93. Hamm P, Hochstrasser RM (2001) In: Fayer MD (ed) *Ultrafast Infrared and Raman Spectroscopy*. Marcel Dekker, New York
94. Jonas DM (2003) *Annu Rev Phys Chem* 54:425–463

95. Hochstrasser RM (2006) Dynamical Models for Two-Dimensional Infrared Spectroscopy of Peptides. In: Berry S, Jortner J (eds) *Advances in Chemical Physics*, Vol 132. John Wiley, Hoboken, New Jersey
96. Scheurer C, Mukamel S (2002) *J Chem Phys* 116:6803–6816
97. Chin W, Piuzzi F, Dimicoli I, Mons M (2006) *Phys Chem Chem Phys* 8:1033–1048
98. Tokmakoff A, Lang MJ, Larsen DS, Fleming GR, Chernyak V, Mukamel S (1997) *Phys Rev Lett* 79:2702–2705
99. Chen PC, Joyner CC (2006) *J Phys Chem A* 110:7989–7993
100. Woutersen S, Hamm P (2001) *J Chem Phys* 115:7737–7743
101. Voronine D, Abramavicius D, Mukamel S (2006) *J Chem Phys* 124:034104
102. Khalil M, Tokmakoff A (2001) *Chem Phys* 266:213–230
103. Bredenbeck J (2005) *Transient 2D-IR Spectroscopy: Towards Ultrafast Structural Dynamics of Peptides and Proteins*. Thesis, Universität Zürich
104. Dreyer J (2005) *Int J Quantum Chem* 104:782–793
105. Neugebauer J, Hess BA (2003) *J Chem Phys* 118:7215–7225
106. Golonzka O, Khalil M, Demirdöven N, Tokmakoff A (2001) *Phys Rev Lett* 86:2154–2157
107. Mukamel S (1995) *Principles of Nonlinear Optical Spectroscopy*. Oxford Series in Optical and Imaging Sciences, Vol 6. Oxford University Press, Oxford
108. Zhang WM, Chernyak V, Mukamel S (1999) *J Chem Phys* 110:5011–5028
109. Wang J, Hochstrasser RM (2004) *Chem Phys* 297:195
110. Abramavicius D, Zhuang W, Mukamel S (2004) *J Phys Chem B* 108:18034–18045
111. Wang J, Chen J, Hochstrasser RM (2006) *J Phys Chem B* 110:7545–7555
112. Gorbunov RD, Kosov DS, Stock G (2005) *J Chem Phys* 122:224904
113. Scheurer C, Piryatinski A, Mukamel S (2001) *J Am Chem Soc* 123:3114–3124
114. Zhuang W, Abramavicius D, Hayashi T, Mukamel S (2006) *J Phys Chem B* 110:3362–3374
115. Woutersen S, Mu Y, Stock G, Hamm P (2001) *Proc Natl Acad Sci USA* 98:11254–11258
116. Barron LD, Boogard MP, Buckingham AD (1973) *J Am Chem Soc* 95:603
117. Holzwarth G, Hsu EC, Mosher HS, Faulkner TR, Moscovitz A (1974) *J Am Chem Soc* 96:251–252
118. Kapitán J, Baumruk V, Kopecký V, Bouř P (2006) *J Phys Chem A* 110:4689–4696
119. Qu X, Lee E, Yu G-S, Freedman TB, Nafie LA (1996) *Appl Spectrosc* 50:649–657
120. Polavarapu PL, Zhao C (2000) *Fresenius J Anal Chem* 366:727–734
121. Taniguchi T, Miura N, Nishimura S-I, Monde K (2004) *Mol Nutr Food Res* 48:246–254
122. Keiderling TA (2002) *Curr Opin Chem Biol* 6:682–688
123. Barron LD, Hecht L, Bell AF, Wilson G (1996) *Appl Spectrosc* 50:619–629
124. Barron LD, Ford SJ, Bell AF, Wilson G, Hecht L, Cooper A (1994) *Faraday Discuss* 99:217–232
125. Alasdair LH, Bell F, Barron LD (1994) *J Am Chem Soc* 116:5155–5161
126. Blanch EW, Hecht L, Syme CD, Volpetti V, Lomonosoff GP, Nielsen K, Barron LD (2002) *J Gen Virol* 83:2593–2600
127. Mc Coll IH, Blanch EW, Hecht L, Barron LD (2004) *J Am Chem Soc* 126:8181–8188
128. Mc Coll IH, Blanch EW, Gill AC, Rhie AGO, Ritchie MA, Hecht L, Nielsen K, Barron LD (2003) *J Am Chem Soc* 125:10019–10026
129. Tam CN, Bouř P, Keiderling TA (1997) *J Am Chem Soc* 119:7061–7064
130. Barron LD, Buckingham AD (1971) *Mol Phys* 20:1111
131. Nafie LA, Freedman TB (1983) *J Chem Phys* 78:7108–7116
132. Stephens PJ (1985) *J Phys Chem* 89:748–752
133. Barron LD, Hecht L, Mc Coll IH, Blanch EW (2004) *Mol Phys* 102:731–744

134. Nafie LA (1997) *Annu Rev Phys Chem* 48:357–386
135. Polavarapu PL (1998) *Vibrational Spectra: Principles and Applications with Emphasis in Optical Activity. Studies in Physical and Theoretical Chemistry*, Vol 85. Elsevier, New York
136. Rosenfeld LZ (1928) *Z Physik* 52:161
137. Galwas PA (1983) Thesis, Cambridge University
138. Buckingham AD, Fowler PW, Galwas PA (1987) *Chem Phys* 112:1–14
139. Amos RD, Handy NC, Jalkanen KJ, Stephens PJ (1987) *Chem Phys Lett* 133:21–26
140. Lowe MA, Segal GA, Stephens PJ (1986) *J Am Chem Soc* 108:248–256
141. Stephens PJ, Develin FJ (2000) *Chirality* 12:172–179
142. Freedman TB, Cao X, Dukor RK, Nafie LA (2003) *Chirality* 15:743–758
143. Atkins PW, Barron LD (1969) *Mol Phys* 16:453
144. Buckingham AD, Dunn MB (1971) *J Chem Soc A*, p 1988
145. Quinet O, Liégeois V, Champagne B (2005) *J Chem Theory Comput* 1:444–452
146. Helgaker T, Ruud K, Bak KL, Jørgensen P, Olsen J (1994) *Faraday Discuss* 99:165
147. Ruud K, Helgaker T, Bouř P (2002) *J Phys Chem A* 106:7448–7455
148. Pecul M, Ruud K (2005) *Int J Quantum Chem* 104:816–829
149. Crawford TD (2006) *Theor Chem Acc* 115:227–245
150. Kapitán J, Baumruk V, Bouř P (2006) *J Am Chem Soc* 128:2438–2443
151. He J, Petrovich A, Polavarapu PL (2004) *J Phys Chem A* 108:1671–1680
152. He J, Petrovic AG, Polavarapu PL (2004) *J Phys Chem B* 108:20451–20457
153. van Gisbergen SJA, Snijders JG, Baerends EJ (1999) *Comput Phys Comm* 118:119–138
154. Zuber G, Goldsmith M-R, Beratan DN, Wipf P (2005) *Chem Phys Chem* 6:595–597
155. Bouř P (2001) *J Comput Chem* 22:426–435
156. Cappelli C, Monti S, Rizzo A (2005) *Int J Quantum Chem* 104:744–757
157. Pecul M, Lamparska E, Cappelli C, Frediani L, Ruud K (2006) *J Phys Chem A* 110:2807–2815
158. Jalkanen KJ, Nieminen RM, Friemand K, Bohr J, Bohr H, Wade RC, Tajkhorshid E, Suhai S (2001) *Chem Phys* 265:125–151
159. Nafie LA (1997) *J Phys Chem A* 101:7826–7833
160. Freedman TB, Shih M-L, Lee E, Nafie LA (1997) *J Am Chem Soc* 119:10620–10626
161. Freedman TB, Gao X, Shih M-L, Nafie LA (1998) *J Phys Chem A* 102:3352–3357
162. Freedman TB, Lee E, Nafie LA (2000) *J Phys Chem A* 104:3944–3951
163. Hug W (2001) *Chem Phys* 264:53–69
164. Hug W, Haesler J (2005) *Int J Quantum Chem* 104:695–715
165. Zuber G, Hug W (2004) *Helv Chim Acta* 87:2208–2234
166. Spiro TG, Czernuszewicz RS (2000) In: Que L Jr (ed) *Physical Methods in Bioinorganic Chemistry*. University Science Books, Sausalito, CA
167. Loehr TM (1998) In: Solomon EI, Hodgson KO (eds) *Spectroscopic Methods in Bioinorganic Chemistry*. American Chemical Society, Washington, DC
168. Czernuszewicz RS, Spiro TG (1999) In: Solomon EI, Lever ABP (eds) *Inorganic Electronic Structure and Spectroscopy*, Vol 1. Wiley, New York
169. Petry R, Schmitt M, Popp J (2003) *Chem Phys Chem* 4:14–30
170. Egawa T, Yeh S-R (2005) *J Inorg Biochem* 99:72–96
171. Kitagawa T, Mizutani Y (1994) *Coord Chem Rev* 135/136:685–735
172. Spiro TG, Wasbotten IH (2005) *J Inorg Biochem* 99:34–44
173. Hudeček J, Anzenbacherová E, Anzenbacher P, Munro AW, Hildebrandt P (2000) *Arch Biochem Biophys* 383:70–78
174. Spiro TG (1983) In: Lever ABP, Gray HB (eds) *Iron Porphyrins*, Part 2. Addison-Wesley, Reading, MA

175. Spiro TG, Li XY (1988) In: Spiro TG (ed) *Resonance Raman Spectra of Heme- and Metalloproteins*. Wiley, New York
176. Berthomieu C, Hienerwadel R (2005) *Biochim Biophys Acta* 1707:51–66
177. Asher SA, Mikhonin AV, Bykov S (2004) *J Am Chem Soc* 126:8433–8440
178. Mikhonin AV, Ahmed Z, Ianoul A, Asher SA (2004) *J Phys Chem B* 108:19020–19028
179. Huang C-Y, Balakrishnan G, Spiro TG (2006) *J Raman Spectrosc* 37:277–282
180. Jeong DH, Kim D, Cho DW, Jeoung SC (2001) *J Raman Spectrosc* 32:487–493
181. Pinakoulaki E, Koutsoupakis C, Stavarakis S, Aggelaki M, Papadopoulos G, Daskalakis V, Varotsis C (2005) *J Raman Spectrosc* 36:337–349
182. Negrerie M, Cianetti S, Vos MH, Martin J-L, Kruglik SG (2006) *J Phys Chem B* 110:12766–12781
183. Peticolas WL, Rush T III (1995) *J Comput Chem* 16:1261–1270
184. Hassing S, Mortensen OS (1980) *J Chem Phys* 73:1078–1083
185. Myers AB (1996) *Chem Rev* 96:911
186. Lee S-Y, Heller EJ (1979) *J Chem Phys* 71:4777
187. Heller EJ, Sundberg RL, Tannor D (1982) *J Phys Chem* 86:1822
188. Neugebauer J, Hess BA (2004) *J Chem Phys* 120:11564–11577
189. Neugebauer J, Baerends EJ, Efremov EV, Ariese F, Gooijer C (2005) *J Phys Chem A* 109:2100
190. Caillie CV, Amos RD (1999) *Chem Phys Lett* 308:249–255
191. Caillie CV, Amos RD (2000) *Chem Phys Lett* 317:159–164
192. Furche F, Ahlrichs R (2002) *J Chem Phys* 117:7433–7447
193. Hutter J (2003) *J Chem Phys* 118:3928–3934
194. Rappoport D, Furche F (2005) *J Chem Phys* 122:064105
195. Jarzęcki AA, Spiro TG (2005) *J Phys Chem A* 109:421–430
196. Petrenko T, Ray K, Wieghardt KE, Neese F (2006) *J Am Chem Soc* 128:4422–4436
197. Schulman JM, Detrano R, Musher JI (1972) *Phys Rev A* 5:1125
198. Schulman JM, Detrano R (1974) *Phys Rev A* 10:1192
199. Schulman JM, Lee S-Y (1981) *J Chem Phys* 74:4930
200. Lee S-Y (1982) *J Chem Phys* 76:3064
201. Lee S-Y (1983) *J Chem Phys* 78:723
202. Jensen L, Zhao LL, Autschbach J, Schatz GC (2005) *J Chem Phys* 123:174110
203. Tozzini V, Giannozzi P (2005) *Chem Phys Chem* 6:1786–1788
204. Zheltikov AM (2000) *J Raman Spectrosc* 31:653–667
205. Begley RF, Harvey AB, Byer RL (1974) *Appl Phys Lett* 25:387–390
206. Rinia HA, Bonn M, Müller MU (2006) *J Phys Chem B* 110:4472–4479
207. Schlücker S, Koster J, Nissum M, Popp J, Kiefer W (2001) *J Phys Chem A* 105:9482–9488
208. Otto C, Voroshilov A, Kruglik SG, Greve J (2001) *J Raman Spectrosc* 32:495–501
209. Popp J, Popp A, Ujj L, Atkinson GH, Sheves M, Ottolenghi M (1996) *J Ram Spectrosc* 27:87–95
210. Wohlleben W, Buckup T, Herek JL, Motzkus M (2005) *Chem Phys Chem* 6:1
211. Knutsen KP, Messer BM, Onorato RM, Saykally RJ (2006) *J Phys Chem B* 110:5854–5864
212. Zumbusch A, Holtom G, Sunney Xie X (1999) *Phys Rev Lett* 82:4142–4145
213. van Manen H-J, Kraan YM, Roos D, Otto C (2005) *Proc Natl Acad Sci USA* 102:10159–10164
214. Kiefer W, Materny A, Schmitt M (2002) *Naturwissenschaften* 89:250–258
215. Volkmer A, Book LD, Xie XS (2002) *Appl Phys Lett* 80:1505–1507
216. Konradi J, Singh AK, Scaria AV, Materny A (2006) *J Raman Spectrosc* 37:697–704

217. Siebert T, Maksimenka R, Materny A, Engel V, Kiefer W, Schmitt M (2002) *J Raman Spectrosc* 33:844–854
218. Siebert T, Schmitt M, Gräfe S, Engel V (2006) *J Raman Spectrosc* 37:397–403
219. Stock G, Domcke W (1997) *Adv Chem Phys* 100:1
220. Eesley GL (1981) *Coherent Raman Spectroscopy*. Pergamon Press, Oxford
221. Yuratich MA, Hanna DC (1977) *Mol Phys* 32:671
222. Teets RE (1984) *Opt Lett* 9:226–228
223. Funk J-M, Schmitt U, Kiefer W, Materny A (2000) *J Raman Spectrosc* 31:743–753
224. Sariyanni Z-E, Rostovtsev Y (2004) *J Mod Optic* 51:2637–2644
225. Varghek M, Freedman TB, Lee E, Nafie LA (1998) *Chem Phys Lett* 287:359–364
226. Nafie LA (1996) *Chem Phys* 205:309–322
227. Scheidt WR, Durbin SM, Sage JT (2005) *J Inorg Biochem* 99:60–71
228. Sage JT, Durbin SM, Sturhahn W, Wharton DC, Champion PM, Hession P, Sutter J, Alp EE (2001) *Phys Rev Lett* 86:4966–4969
229. Xiao Y, Wang H, George SJ, Smith MC, Adams MWW, Jenney FE Jr, Sturhahn W, Alp EE, Zhao J, Yoda Y, Dey A, Solomon EI, Cramer SP (2005) *J Am Chem Soc* 127:14596–14606
230. Adams KL, Tsoi S, Yan J, Durbin SM, Ramdas AK, Cramer WA, Sturhahn W, Alp EE, Schulz C (2006) *J Phys Chem B* 110:530–536
231. Xiao Y, Fisher K, Smith MC, Newton WE, Cass DA, George SJ, Wang H, Sturhahn W, Alp EE, Zhao J, Yoda Y, Cramer SP (2006) *J Am Chem Soc* 128:7608–7612
232. Herrmann C, Ruud K, Reiher M (2006) *Chem Phys Chem* 7:2189–2196

Car–Parrinello Molecular Dynamics Simulations and Biological Systems

Jens Thar · Werner Reckien · Barbara Kirchner (✉)

Institute for Physical and Theoretical Chemistry, Wegelerstr. 12, 53115 Bonn, Germany
 kirchner@thch.uni-bonn.de

1	Introduction	134
2	Molecular Dynamics Simulations	137
2.1	Equations of Motion	137
2.2	Thermostats	139
3	Electronic Structure	141
3.1	Hohenberg–Kohn Theorems	141
3.2	Kohn–Sham Equations	142
3.3	Exchange Correlation Functional and Dispersion Corrections	143
4	Basis Sets and Pseudopotentials	145
4.1	Plane Waves	145
4.2	Pseudopotentials	147
5	Car–Parrinello Simulations	149
5.1	Fictitious Mass μ	150
6	QM/MM	151
7	Literature on Biological Systems	155
7.1	Proteins, Peptides, and Amino Acids	155
7.2	Metal-Containing Biomolecules: Metallo-enzymes	156
7.3	DNA and RNA	157
7.4	Miscellaneous, Biomimetics	158
7.5	Excited States MD	158
8	Case Study: Biomimetic Nitrogen Fixation	159
9	Conclusion	165
	References	166

Abstract The basic concepts of Car–Parrinello molecular dynamics (CPMD) and its application to biological systems are reviewed. In Car–Parrinello simulations the electronic structure is calculated on the fly, i.e., during the course of simulations the potential is adjusted according to the chemical events occurring in the system. This allows for more unbiased simulations of chemical processes, as opposed to classical molecular dynamics, which utilize predefined potentials. In contrast to Born–Oppenheimer molecular dynamics, where the nuclei are propagated under classical equations of motions and where the wave function is calculated in every time step, CPMD maps this two-component classi-

cal/quantum system onto a two-component purely classical system. A short introduction into classical mechanics, which provides the equations of motions, is given. DFT will be briefly outlined because it represents the standard method for obtaining the electronic structure during a CPMD simulation. QM/MM approaches, which combine the CPMD method for a core region with a classical description of its environment, are also reviewed. These schemes provide a more reliable description of biological systems, whose properties are strongly influenced by their surroundings. A brief overview over the application of Car–Parrinello simulations to biological systems is provided. Finally, a case study to illustrate some of the basic possibilities of this method is discussed.

Abbreviations

BOMD	Born–Oppenheimer molecular dynamics
CPMD	Car–Parrinello molecular dynamics
DFT	Density functional theory
GGA	Generalized gradient approximation
HF	Hartree–Fock
LDA	Local density approximation
MD	Molecular dynamics
MM	Molecular mechanics
MP2	Second order Møller–Plesset theory
PP	Pseudopotential
PW	Plane wave
QM/MM	Quantum mechanics/molecular mechanics
TDDFT	Time-dependent density functional theory
WDA	Weighted density approximation

1

Introduction

Chemical and biological systems are far from simple and can encompass a complicated interplay between many different sorts of interactions of many different functional groups. Furthermore, these systems are often not stiff, but flexible and therefore of a dynamical nature. In view of this complicated nature of biological and complex chemical systems the right choice of a method that is able to include all of these effects is necessary. While traditional, static quantum chemical methods are able to resolve, for instance, the geometrical or electronic behavior of a relatively small amount of atoms at a high level of accuracy, traditional molecular dynamics simulation techniques are able to include a large amount of atoms and to temporally resolve the microscopic level. Thus, when interested in the high resolution of a local area within the biomolecule, quantum chemical methods are certainly the right choice. On the other hand, one might be interested in solvent effects (and therefore be in need of a large amount of atoms) or in the dynamical behavior (and be in need of a dynamical description of the nuclei) so the right choice of methods will be given by traditional molecular dynamics

simulations. Situations might occur where both a quantum chemical treatment of the electronic structure and a dynamical description of the nuclei are needed. In 1985, Car and Parrinello introduced with their seminal paper [1] one such method that is able to combine the advantages of both methods. They used as an elegant trick the extended Lagrangian technique to map a two-component classical/quantum mechanical treatment (atoms/electrons) onto a two-component purely classical treatment. As a result the method is suitable:

- To follow the temporal resolution of the atoms, which is not possible with static quantum chemical methods, see Fig. 1
- To describe solvent effects, since an ensemble of molecules with periodic boundary conditions can be treated, see Fig. 2 (right picture)
- To describe spontaneous events, which is impossible for traditional molecular dynamics simulations that rely on predefined potentials, see Fig. 2 (left picture)
- To include polarization or cooperative effects, because the many-body nature is properly treated within this method, see Fig. 3

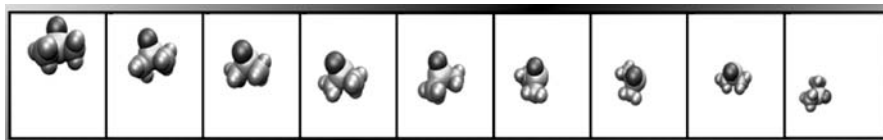


Fig. 1 Diffusion of DMSO molecules in water. The water molecules are dispensed with for clarity. For simulations see [2–4]

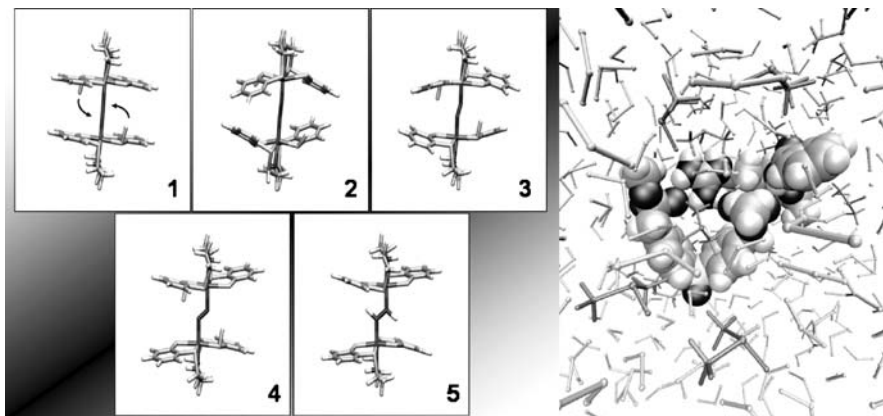


Fig. 2 *Left:* Cartoon of a spontaneous event (a double hydrogen transfer in this case), see Sellmann complexes in Sect. 8 and [5]. *Right:* Illustration of solvent effects at the example of an artificial guanidinium receptor, Kirchner et al. unpublished results

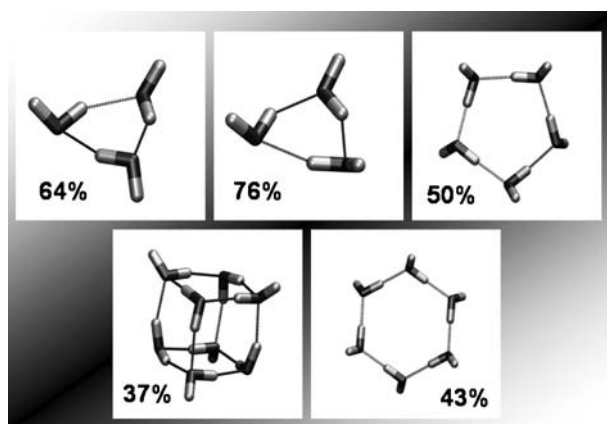


Fig. 3 Various water clusters showing cooperative effects. The numbers give the percentage of energy that is covered by assuming pair additivity, see [6]

Since there is no “free meal”, disadvantages of the Car–Parrinello method, like short simulation times and lower quantum chemical accuracy, are a byproduct of the limited amounts of computer time available today. Those disadvantages will be discussed in the relevant sections below. Car–Parrinello simulations have already been the subject of many review articles. For additional information, the reader is referred to the excellent reviews of Remler and Madden [7], Marx and Hutter [8], and Tuckerman and Martyna [9, 10]. The next few sections describe the Car–Parrinello methodology, which is the central topic of this article. Since the Car–Parrinello method uses molecular dynamics simulations, we will start with a very brief outline of the methodology, however, we will not go into detail because this topic is covered by another chapter (see the review of Schulten et al. in this volume). The molecular dynamics discussion is followed by a section on electronic structure method, where only density functional theory is introduced. The reason for this is that up to now the Car–Parrinello method is mostly implemented on the basis of density functional theory. This is not due to shortcomings of the method, but due to the reasonable amount of computer time available. After the discussion of the electronic structure method we shed some light onto the basis sets and the pseudopotentials used. Next, we turn to the Car–Parrinello method in detail. This is followed by a short introduction of the QM/MM methodology, as most often biological systems are studied using the Car–Parrinello method in the framework of a QM/MM ansatz. Again we do not go into much detail, since this topic is covered by Senn and Thiel in this volume. We continue with an overview of articles that have appeared in the literature and have applied the Car–Parrinello method to biological systems; one selected example will be highlighted. Finally, conclusions and discussions end this review article.

2

Molecular Dynamics Simulations

We will first briefly review the basics of molecular dynamics simulations, which are required to understand the ideas behind first principle molecular dynamics simulations [11]. The aim of molecular dynamics simulations is to study a system by recreating it on the computer in a manner that is as close to nature as possible, i.e., by simulating the dynamics of a system over a physical length of time relevant to properties of interest. Therefore, there are two topics at the heart of molecular dynamics simulations. First, the equations of motion must be defined and second, the forces by which the particles are moved must be described.

To this end, classical mechanics is applied to describe the nuclear motion of the particles, i.e., the equations of motion. This is an approximation. Extensions that include quantum effects [12, 13] have been explored. It turns out that employing classical mechanics is a very good approximation for molecular systems, as long as the properties studied are not related to vibrations with a frequency ν such that $h\nu \gg k_b T$.

2.1

Equations of Motion

The equations of motion may be written in various ways. We assume a system of N particles interacting via a potential U . Each particle I (usually an atom) is described by its position $\mathbf{R}_I$ and conjugated momenta $\mathbf{P}_I = M_I \dot{\mathbf{R}}_I$. Here M_I is the mass of the particle. In our simulations we have a set of all positions $\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N\}$ and of all momenta $\{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_N\}$, which are abbreviated $\mathbf{R}^N$ and $\mathbf{P}^N$, respectively. The equations of motion can be derived from the Euler-Lagrange equations:

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}} - \frac{\partial \mathcal{L}}{\partial \mathbf{R}} = 0, \quad (1)$$

where the Lagrangian function $\mathcal{L}$ is given in terms of kinetic minus potential energy:

$$\mathcal{L}(\mathbf{R}^N, \dot{\mathbf{R}}^N) = \mathcal{K} - U = \sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 - U(\mathbf{R}^N). \quad (2)$$

The Lagrangian formalism allows for an easy introduction of restraints by simply adding the constraining condition. Alternatively, the Hamiltonian for-

malism can be used:

$$\begin{aligned}\mathcal{H}(\mathbf{R}^N, \dot{\mathbf{R}}^N) &= \sum_{I=1}^N \dot{\mathbf{R}}_I^N \dot{\mathbf{P}}_I^N - \mathcal{L}(\mathbf{R}^N, \dot{\mathbf{R}}^N) \\ &= \mathcal{K} + U = \sum_{I=1}^N \frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R}^N).\end{aligned}\quad (3)$$

The force $\mathbf{F}_I(\mathbf{R}^N)$ on particle I is obtained as the change of $U(\mathbf{R}^N)$ due to an infinitesimal displacement of I , i.e., the partial derivative of the potential $U(\mathbf{R}^N)$:

$$\mathbf{F}_I(\mathbf{R}^N) = - \frac{\partial U(\mathbf{R}^N)}{\partial \mathbf{R}_I} . \quad (4)$$

In traditional molecular dynamics simulations an analytical function like the Lennard–Jones potential is used for $U(\mathbf{R}^N)$, such that this analytical form is simple enough to obtain the analytical derivative.

The description of these forces can be very important for the outcome of the simulations [14, 15]. Usually the forces on the particles are operated pairwise additively, as already explained in the introduction. $U(\mathbf{R}^N)$ is then given as a sum over two-body potentials $u(\mathbf{R}_{IJ})$:

$$U(\mathbf{R}^N) = \sum_I^N \sum_{J>I}^N u(\mathbf{R}_{IJ}) = \frac{1}{2} \sum_I^N \sum_{J \neq I}^N u(\mathbf{R}_{IJ}) . \quad (5)$$

This is again an approximation, since many-body or cooperative effects are neglected. As opposed to the quantum nature of the particle (nuclei) dynamics, cooperative effects are obviously more important in complicated liquids. A better description of many-body effects is provided by polarizable forces fields that make up for the neglect of such effects. A second way to circumvent this problem is given by first-principle simulations, which explicitly include cooperative effects in the description of the system under consideration (as we will see in the next sections) because here the complete electronic structure of the system is treated at once.

The equations of motion according to Hamilton's equation read:

$$\dot{\mathbf{R}}_I = \frac{\partial \mathcal{H}}{\partial \mathbf{P}_I} = \frac{\mathbf{P}_I}{M_I} , \quad (6)$$

$$\dot{\mathbf{P}}_I = - \frac{\partial \mathcal{H}}{\partial \mathbf{R}_I} = - \frac{\partial U(\mathbf{R}^N)}{\partial \mathbf{R}_I} = \mathbf{F}_I(\mathbf{R}^N) . \quad (7)$$

Newton's second law can be obtained from this equation by using $\mathbf{P}_I = M_I \dot{\mathbf{R}}_I$, taking the time derivative and equating it to the second part of the Hamilton

equations and inserting it into Eq. 7:

$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_I(\mathbf{R}^N). \quad (8)$$

Having arrived at the equations of motion, numerical integration helps to find the solution. One of the most prominent algorithms was invented by Störmer and Verlet [16, 17].

Before discussing first-principle simulations another approach to simulations should be mentioned, namely the extended system approach. In this, additional degrees of freedom that control a quantity under consideration can be added to the system [16–19]. Thereby, thermostats or barostat can be simulated [20–24].

2.2

Thermostats

The most forward way to keep the temperature (i.e., the kinetic energy) constant in classical molecular dynamics simulations is to scale the velocities of every single particle [16]. However, one can show that the average temperature of a NVT ensemble (i.e., an ensemble where the number of particles N , the volume V , and the temperature T are kept constant) fluctuates, and that therefore the scaling of every single velocity can lead to an unphysical system [16]. A convenient method to keep the NVT ensemble is the Nosé(–Hoover)-thermostat [20, 21, 23, 24], where a heat bath is coupled to the system and energy can flow freely between these two systems. The heat bath is represented by an additional degree of freedom S , while the energy flow is described by the conjugated momentum P_S of S . The complete system is describe by the extended Lagrangian: $\mathcal{L}_{\text{Nosé}}(\dot{\mathbf{R}}, \mathbf{R}, \dot{S}, S)$ [20, 21],

$$\begin{aligned} \mathcal{L}_{\text{Nosé}} &= \mathcal{L}_{\text{Nosé}}(\dot{\mathbf{R}}, \mathbf{R}, \dot{S}, S) \\ &= \sum_{I=1}^N \frac{M_I}{2} \dot{S}^2 \dot{\mathbf{R}}_I^2 - U(\mathbf{R}^N) + \frac{Q}{2} \dot{S}^2 - L k_b T \ln S, \end{aligned} \quad (9)$$

where k_b is the Boltzmann constant and L is a chosen parameter to obtain the canonical average. Q is considered an effective mass, and controls the speed with which energy is transferred between the system and the reservoir. It must be chosen with care, because small values for Q lead to a fast oscillation of the extended variable S , and therefore require a very small time step to accurately describe the extended system. However, this leads to more time steps needed to equilibrate the system, which is to be kept at constant temperature. On the other hand, large values of Q lead to a very slow response of S to changes in this system, and here the convergence of the additional degree of freedom S requires a larger number of time steps [25]. With the conjugated

momenta:

$$P_I \equiv \frac{\partial \mathcal{L}}{\partial \dot{R}_I} = M_I S^2 \dot{R}_I, \quad (10)$$

$$P_S \equiv \frac{\partial \mathcal{L}}{\partial \dot{S}} = Q \dot{S}, \quad (11)$$

one obtains the Hamiltonian $\mathcal{H}_{\text{Nosé}}$ for the complete system:

$$\mathcal{H}_{\text{Nosé}} = \mathcal{H}_{\text{Nosé}}(\dot{\mathbf{R}}, \mathbf{R}, \dot{S}, S) \quad (12)$$

$$= \sum_{I=1}^N \frac{P_I^2}{2M_I S^2} + U(\mathbf{R}^N) + \frac{P_S^2}{2Q} + Lk_b T \ln S. \quad (13)$$

The equations of motion are therefore given by:

$$\frac{d\mathbf{R}_I}{dt} = \frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial \mathbf{P}_I} = \frac{\mathbf{P}_I}{M_I S^2}, \quad (14)$$

$$\frac{d\mathbf{P}_I}{dt} = - \frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial \mathbf{R}_I} = - \frac{\partial U(\mathbf{R}^N)}{\partial \mathbf{R}_I}, \quad (15)$$

$$\frac{dS}{dt} = \frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial P_S} = \frac{P_S}{Q}, \quad (16)$$

$$\frac{dP_S}{dt} = - \frac{\partial \mathcal{H}_{\text{Nosé}}}{\partial S} = \frac{\sum_I \frac{P_I^2}{M_I S^2} - Lk_b T}{S}. \quad (17)$$

For this selection, one must choose $L = 3N + 1$ to obtain a canonical system. More commonly, one introduces the scaled variables V' [22, 23], which are also called real variables because they resemble physical observables most closely. These variables are related to the initial variables by the following equations:

$$R' = R, \quad (18)$$

$$P' = P/S, \quad (19)$$

$$S' = S, \quad (20)$$

$$P'_S = P_S/S, \quad (21)$$

$$t' = t/S. \quad (22)$$

Using relations Eqs. 18–22 and introducing the thermodynamic friction coefficient $\xi = S'P'_S/Q$, we can write the Eqs. 14–17 in the following form:

$$\frac{d\mathbf{R}'_I}{dt'} = S \frac{d\mathbf{R}_I}{dt} = \frac{\mathbf{P}_I}{M_I S} = \frac{\mathbf{P}'_I}{M_I}, \quad (23)$$

$$\begin{aligned} \frac{d\mathbf{P}'_I}{dt'} &= S \frac{d(\mathbf{P}_I/S)}{dt} = \frac{d\mathbf{P}_I}{dt} - \frac{1}{S} \mathbf{P}_I \frac{dS}{dt} \\ &= - \frac{\partial U(\mathbf{R}'^N)}{\partial \mathbf{R}'_I} - \frac{S'P'_S}{Q} \mathbf{P}'_I = - \frac{\partial U(\mathbf{R}'^N)}{\partial \mathbf{R}'_I} - \xi \mathbf{P}'_I, \end{aligned} \quad (24)$$

$$\frac{1}{S} \frac{dS'}{dt'} = \frac{S}{S} \frac{dS}{dt} = \frac{S'P'_S}{Q} = \xi, \quad (25)$$

$$\begin{aligned} \frac{d \frac{S'P'_S}{Q}}{dt'} &= \frac{S}{Q} \frac{dP_S}{dt} \\ &= \frac{\sum_I \frac{P_I^2}{M_I} - Lk_b T}{Q} = \dot{\xi}. \end{aligned} \quad (26)$$

From these equations of motion one can see how the thermostat works: The momentum of each particle depends on the friction coefficient through Eq. 24, i.e., it is lowered if ξ is positive or it rises if ξ is negative. The friction coefficient is related to the desired temperature T through Eq. 26. If the temperature of the system is too large, ξ rises, which scales down the momenta.

3

Electronic Structure: Density Functional Theory

Because first-principle simulations usually employ density functional theory (DFT), we will give a brief outline of DFT. The reason for the extensive use of DFT is its low cost while giving a comparatively good description of the electronic structure.

3.1

Hohenberg–Kohn Theorems

DFT is used as the formulation to describe the electronic structure based on the Hohenberg–Kohn theorems [26]. The first Hohenberg–Kohn theorem states that the ground state density $\rho(\mathbf{r})$:

$$\rho(\mathbf{r}) = \rho(\mathbf{r}_1) = N \int \dots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 d\mathbf{s}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N, \quad (27)$$

with the electronic wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ depending on the spatial and spin coordinates $\mathbf{x}_i$ and with N as the number of electrons in the system, determines the external potential $\hat{V}_{\text{ext}}$. This theorem is only valid for a non-degenerate ground state. With N given by:

$$N = \int \rho(\mathbf{r}) d\mathbf{r}, \quad (28)$$

the complete Hamiltonian is known, which yields the wave function and therefore all properties of the system. As a result, we can write the energy as a functional of the density:

$$E[\rho] = T[\rho] + V_{\text{ext}}[\rho] + V_{\text{ee}}[\rho] \quad (29)$$

$$= T[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + V_{\text{ee}}[\rho], \quad (30)$$

where $T[\rho]$ is the kinetic energy and $V_{\text{ee}}[\rho]$ is the electron–electron interaction. The second Hohenberg–Kohn theorem allows the use of the variational principle so that a trial density $\tilde{\rho}$ provides an approximation $E[\tilde{\rho}]$ to the exact ground state energy $E_0[\rho]$:

$$\langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle = T[\tilde{\rho}] + \int \tilde{\rho}(\mathbf{r})v(\mathbf{r})d\mathbf{r} + V_{\text{ee}}[\tilde{\rho}] = E[\tilde{\rho}] \geq E_0[\rho]. \quad (31)$$

The remaining problem is to determine the functionals $T[\rho]$ and $V_{\text{ee}}[\rho]$.

3.2

Kohn–Sham Equations

The bright idea of Kohn and Sham was to replace the minimization with respect to an unknown ground state density by a minimization with respect to a set of orthonormal one-particle orbitals [27]. This idea comprises a substitute of non-interacting electrons for the fully interacting N -electron system. The auxiliary system is described by a single-determinant wave function consisting of N orbitals ϕ_i for N -independent electrons. Inserting this wave function Ψ^{KS} into Eq. 27 yields:

$$\rho(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2, \quad (32)$$

which equals the density of the fully interacting system. The energy of this auxiliary system is given by:

$$E^{\text{aux}}[\rho] = T_s[\rho] + \int v_s(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \quad (33)$$

where the kinetic energy term can be easily evaluated as:

$$T_s[\rho] = \sum_i^N \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \right\rangle. \quad (34)$$

Turning back to the system with interacting electrons, we can simply add and subtract the energy expression in Eq. 33 from the energy of the interacting system. Rearranging the terms and introducing the exchange–correlation energy functional: $E_{xc}[\rho]$,

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho], \quad (35)$$

where the Coulomb energy J is given by the classical Poisson integral [28]:

$$J[\rho] = \frac{1}{2} \int \int \frac{1}{|\mathbf{r} - \mathbf{r}'|} \rho(\mathbf{r}) \rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' = \frac{1}{2} \int d\mathbf{r} V_H(\mathbf{r}) \rho(\mathbf{r}), \quad (36)$$

finally yields the Kohn–Sham energy $E^{KS}[\rho]$:

$$E^{KS}[\rho] = T_s[\rho] + \int d\mathbf{r} v(\mathbf{r}) \rho(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r} V_H(\mathbf{r}) \rho(\mathbf{r}) + E_{xc}[\rho], \quad (37)$$

which can be varied with respect to the orbitals, subject to an orthonormality constraint. This leads to the Kohn–Sham equations:

$$\left\{ -\frac{1}{2} \nabla^2 + v(\mathbf{r}) + V_H(\mathbf{r}) + \frac{\delta E_{xc}[\rho]}{\delta n(\mathbf{r})} \right\} \phi_i(\mathbf{r}) = \sum_j \Lambda_{ij} \phi_j(\mathbf{r}). \quad (38)$$

These equations are one-particle equations involving an effective one-particle Hamiltonian h^{KS} (see terms in braces of Eq. 38). Λ_{ij} are the Lagrange multipliers for the constraint that orbitals are orthonormal. The “art” of density functional theory is now to derive an accurate exchange–correlation functional empirically. This will be discussed in the next section.

3.3

Exchange Correlation Functional and Dispersion Corrections

The first approximation to the exchange correlation functional is the so called local density approximation (LDA) [27]. To derive the functional, the Kohn–Sham energy is obtained for a homogeneous electron gas in a large box in which a positive charge is spread out to neutralize the system. As a result, the Coulomb interaction of the electrons with the positive charges (nuclei) cancels exactly with the inter-electron interaction and the inter-nuclei interaction. The energy $E^{LDA}[\{\rho\}]$ then finally reads:

$$E^{LDA}[\rho] = T_s[\rho] + E_{xc}[\rho], \quad (39)$$

where $E_{xc}[\rho]$ can be split up into an exchange part $E_x[\rho]$ and a correlation part $E_c[\rho]$:

$$E[\rho] = T_s[\rho] + E_x[\rho] + E_c[\rho]. \quad (40)$$

From this ansatz, E_x is obtained as:

$$E_x[\rho] = -0.7386 \int \rho(\mathbf{r})^{4/3} d\mathbf{r}. \quad (41)$$

To get the correlation part, Nusair et al. [29] fitted exact Monte Carlo calculations and obtained $E_c[\rho]$ as the difference between the exact energy and $T_s[\rho] + E_x[\rho]$. The LDA approximation has two main drawbacks: First, it originates from a homogeneous electron gas, which is normally not found in real systems. Also, the dependence on $\mathbf{r}$ in the limit of $\mathbf{r} \rightarrow \infty$ is not correct. Becke [30] introduced a gradient-corrected exchange functional term E_x^B of the form:

$$E_x^B = - \int \rho(\mathbf{r}) \beta \rho^{1/3} \frac{x^2}{(1 + 6\beta x \sin h^{-1} x)} d\mathbf{r}, \quad (42)$$

with:

$$x = \frac{|\nabla \rho|}{\rho^{4/3}}. \quad (43)$$

The functional is derived by imposing the correct asymptotic behavior as a restraint. The parameter β is adjusted to $\beta = 0.0042$ so that E_x^B reproduces the exchange energy of six noble gases. Therefore, this functional accounts for both problems associated with the LDA functionals. It is one of the so-called generalized gradient approximation (GGA) functionals, which are given by:

$$E_{xc} = \int F(\rho_\alpha, \rho_\beta, \zeta_{\alpha\alpha}, \zeta_{\beta\beta}, \zeta_{\alpha\beta}) d\mathbf{r}, \quad (44)$$

with $\zeta_{\sigma\sigma'} = \nabla \rho_\sigma \nabla \rho_{\sigma'}$. One problem remains with all the above functionals: dispersion interactions are not described sufficiently well.

A wide variety of functionals to address this problem has been developed so far: the GGA in combination with an (empirical) $1/r^6$ van der Waals correction term E_{vdW} , the weighted density approximation (WDA), and methods based on time-dependent density functional theory [31].

GGAs and $1/r^6$ van der Waals corrections terms were originally applied to correct the Hartree–Fock energy [32]. Based on Anderson, Langreth, and Lundqvist [33] the van der Waals term is corrected for both separated uniform electron gas and separated atoms with an explicit $1/r^6$ term. Others [34–38] include an empirical $1/r^6$ term. A cutoff criterion is employed to separate the regions where the length scale for the change of the local Fermi wave-vector is smaller than the electron screening length. Furthermore, because of the divergence of this E_{vdW} expression for short distances, damping functions were introduced. Becke and coworkers were able to develop a correction

scheme based on the idea of parameterizing the $1/r^6$ contribution generated by the instantaneous dipole moment of the exchange hole [39].

The weighted density approximation [40] starts from the exact expression for the exchange correlation functional to obtain a weighted density approximation. An unknown pair-correlation function enters the expression, which leads to an introduction of a weighted density parameter. Several suitable functions were proposed [41], among them functions of Gunnarson-Jones type, Gaussians, fourth-order Gaussians, and Lorentzians.

The strategy for the seamless approach is to approximate the density distribution $n(r)$ by LDA or GGA and to divide the Coulomb interaction $U(r)$ into short- and long-range parts. The contribution for the long-range interaction starts as in WDA from the exact expression. Unlike the WDA approach, a polarization term enters the equation, which is transformed into a time domain and leads to complicated expressions. If pseudopotentials are used, the long-range parts can also be approximated by a suitably chosen non-local form [42]. This approach will be briefly discussed in the next section.

4

Basis Sets and Pseudopotentials

We want to recall here that the basis sets used in quantum chemical calculations are usually built up by atom-centered functions. Within the category of atomic orbitals, there are mostly two types of basis functions used: (1) Slater-type orbitals and (2) Gaussian-type orbitals [43, 44]. In principle, Gaussian-type orbitals are inferior to the Slater-type orbitals, but they are easier to handle, e.g., two-electron integrals can be calculated analytically. Previously, it was common to develop a basis set according to the special problems at hand, for an example see [45]. Nowadays, many special Gaussian basis sets exist and most often only exponents are changed or functions are added instead of developing a completely new basis set. However, the CPMD method utilizes plane wave basis sets. The reason for this is that simulations are often used to obtain insight into bulk materials like liquids or solids, which require periodic boundary conditions to avoid surface effects. While periodic boundary conditions are naturally incorporated into plane waves, this is not the case for Gaussian-type orbitals, and additional problems arise, which need to be handled with care [46–49].

4.1

Plane Waves

Plane waves [7, 10]:

$$f_G^{\text{PW}}(\mathbf{r}) = \exp[i\mathbf{G} \cdot \mathbf{r}], \quad (45)$$

form a complete and orthonormal basis with an inherent periodicity, where $\mathbf{G}$ is the reciprocal space vector and $\mathbf{r}$ is the electronic coordinate. If a simulation takes place in a cubic box, this box is replicated throughout space to form an infinite lattice. Plane waves are not atom-centered and therefore allow for a steady description of the wave functions at all positions. This means that the simulation is not biased by an unequally good description of different spatial regions. A periodic function can be expanded in the basis of the form given in Eq. 45:

$$\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{L}) = \sum_{\mathbf{G}} \psi(\mathbf{G}) \exp[i\mathbf{G} \cdot \mathbf{r}] . \quad (46)$$

The functions $\psi(\mathbf{r})$ and $\psi(\mathbf{G})$ are related by a three-dimensional Fourier transform. The direct lattice vectors $\mathbf{L}$ connect equivalent points in different cells. The Kohn–Sham orbitals can be written in Bloch form:

$$\phi_i(\mathbf{r}) = \phi_i(\mathbf{r}, \mathbf{k}) = \exp[i\mathbf{k} \cdot \mathbf{r}] u_i(\mathbf{r}, \mathbf{k}) , \quad (47)$$

where $\mathbf{k}$ is a vector in the first Brillouin zone. The functions $u_i(\mathbf{r}, \mathbf{k})$ hold the periodicity of the direct lattice:

$$u_i(\mathbf{r}, \mathbf{k}) = u_i(\mathbf{r} + \mathbf{L}, \mathbf{k}) . \quad (48)$$

The index i runs over all one-particle states and these are associated with an occupation. Expanding the periodic functions $u_i(\mathbf{r}, \mathbf{k})$ in the plane wave basis leads to:

$$u_i(\mathbf{r}, \mathbf{k}) = \sum_{\mathbf{G}} c_i(\mathbf{G}, \mathbf{k}) \exp[i\mathbf{G} \cdot \mathbf{r}] , \quad (49)$$

and the Kohn–Sham orbitals are given as:

$$\phi_i(\mathbf{r}, \mathbf{k}) = \sum_{\mathbf{G}} c_i(\mathbf{G}, \mathbf{k}) \exp[i(\mathbf{G} + \mathbf{k}) \cdot \mathbf{r}] . \quad (50)$$

When the density is used in the calculations, the integration over the Brillouin zone has to be approximated by a finite sum over special $\mathbf{k}$ -points. Furthermore, the infinite sum over $\mathbf{G}$ vectors and cells (stems from the prefactor, which is omitted here for brevity) has to be truncated. Since the Kohn–Sham potential converges rapidly with increasing modulus of $\mathbf{G}$, such $\mathbf{G}$ vectors with a kinetic energy higher than a given maximum cutoff E_{cut} are neglected:

$$\frac{1}{2} |\mathbf{G}|^2 \leq E_{\text{cut}} . \quad (51)$$

The drawback of the uniform description is that regions that are situated far apart from the nuclei, and therefore do not contribute significantly to the wave function of the system, are described by the same high accuracy and the therewith connected CPU time as the regions close to the nuclei. This should not be a severe problem as long as the wave function at the nuclei is described properly with a reasonable small number of plane waves, i.e., with a rather

small cutoff. Unfortunately, this is not the case, because the wave function changes rapidly close to the nuclei, and normally shows several nodes in this region. This is the reason why the plane wave description is usually accompanied by pseudopotentials [8].

Another approach uses a mixed auxiliary basis set to describe the density simply by expanding the orbitals in Gaussians [50, 51]. Such approaches are available in standard program packages such as the CP2K code [52].

4.2

Pseudopotentials

The rapid change of the wave function in the inner core region of each atom requires a prohibitively large set of plane waves. This region does not contribute much to the chemical binding so an approximate, but much smoother (and therefore less complicated), description of the wave function in this region is feasible [8, 53]. This can be achieved by altering the potential of the nuclear charges and the core electrons inside an (arbitrarily) chosen cut-off radius r_c , i.e., by replacing the old potential by a pseudopotential. This pseudopotential should fulfill a number of requirements [54]. First it should generate a new orbital $\phi_l = \phi_l^{ps}(\epsilon, r)$ that reproduces the exact valence shell eigenvalues as well as the exact all-electron orbital $\phi_l^{ae} = \phi_l^{ae}(\epsilon, r)$ outside r_c . It should be norm conserving, i.e., the number of electrons in the core region should be identical for both ϕ_l and ϕ_l^{ae} . The pseudoorbital ϕ_l should have the same logarithmic derivative as well as the same first energy derivative of the logarithmic derivative as ϕ_l^{ae} outside r_c . A major breakthrough was achieved by Hamann, Schlüter, and Chiang [54] who showed that norm conservation and the energy derivative of the logarithmic derivative (at the distance R) are related through the identity:

$$-2\pi \left[(r\phi_l)^2 \frac{d}{d\epsilon} \frac{d}{dr} \ln \phi_l \right]_R = 4\pi \int_0^R \phi_l^2(r) r^2 dr. \quad (52)$$

To prove this identity [55], one must consider the radial equation:

$$-\frac{1}{2}\phi_l'' + \left[\frac{l(l+1)}{2r^2} + V_{\text{eff}}(r) - \epsilon \right] \phi_l = 0 \quad (53)$$

for a spherical atom, with $\phi_l' = \frac{d}{dr}\phi_l$. This relation can be reformulated into:

$$2 \left[\frac{l(l+1)}{2r^2} + V_{\text{eff}}(r) - \epsilon \right] = \frac{\phi_l''}{\phi_l} = \frac{d}{dr} \frac{\phi_l'}{\phi_l} + \frac{(\phi_l')^2}{\phi_l^2} = \chi_l' + \chi_l^2, \quad (54)$$

defining the function $\chi_l = \chi_l(\epsilon, r)$ as:

$$\chi_l \equiv \frac{d}{dr} \ln \phi_l = \frac{\phi_l'}{\phi_l}. \quad (55)$$

Differentiating Eq. 54 with respect to ϵ yields:

$$-2 = \frac{\partial}{\partial \epsilon} \chi_l' + 2\chi_l \frac{\partial}{\partial \epsilon} \chi_l. \quad (56)$$

For any arbitrary function $f(r)$ and any l the relation:

$$\frac{1}{\phi_l^2} \frac{\partial}{\partial r} [\phi_l^2 f(r)] = f'(r) + 2 \frac{\phi_l'}{\phi_l} f(r) = f'(r) + 2\chi_l f(r) \quad (57)$$

is valid. Inserting this equation for $f(r) = \frac{\partial}{\partial \epsilon} \chi_l$ into Eq. 56, multiplying by $\phi_l^2(r)$ and integrating over the complete space, one finally obtains the above identity. The pseudopotential is obtained by optimizing ϕ_l^{ae} and fitting ϕ_l with the new pseudopotential under the above-mentioned constraints to the real eigenvalue. Because the pseudopotential must be calculated for every angular momentum, but acts only as a linear operator on r , it is commonly written as a semi-local operator $\hat{V}_{SL}$:

$$\hat{V}_{SL} = \sum_{lm} |Y_{lm}\rangle V_l(r) \langle Y_{lm}|, \quad (58)$$

where $Y_{lm} = Y_{lm}(\theta, \phi_l) = P_{lm}(\cos(\theta))e^{im\phi_l}$ is a spherical harmonic. This approach yields very accurate results [56], but suffers in plane wave calculations from the high cost of evaluating the matrix elements containing the semi-local potential.

Kleinman and Bylander [57] addressed this remaining problem by separating the semi-local potential in a local l -independent and a new semi-local part:

$$\hat{V}_{SL} = \hat{V}_{\text{local}}(r) + \sum_{lm} |Y_{lm}\rangle \delta V_l(r) \langle Y_{lm}|, \quad (59)$$

where the semi-local part is then replaced by a new, computationally less demanding and fully separable non-local term. This term can be chosen arbitrarily, which therefore allows for additional error corrections.

Goedecker et al. [58, 59] finally introduced a new class of analytic separable pseudopotentials, tailor-made for fast integration on the real space grid, but still norm conserving. The fitting procedure was slightly altered by minimizing a penalty function, which takes into account the difference in charge as well as the eigenvalues between ϕ_l^{ae} and ϕ_l .

This approach was extended by Lilienfeld et al. [42] who introduced a second penalty function to account for the long range dispersion interaction represented by MP2 calculations. This second penalty function was minimized

using one single additional projector, i.e., one additional angular momentum quantum number, generating one additional, non-local potential.

5

Car–Parrinello Simulations

The most straightforward way to introduce first-principle molecular dynamics is to propagate the nuclei classically and solve Schrödinger’s equation for every time step “on the fly”, to obtain the new orbitals. This procedure is called Born–Oppenheimer molecular dynamics (BOMD) [60]. Compared to the Born–Oppenheimer approximation, where the atoms do not move during the wave function optimization, and the approximations inherent in the electronic structure method, BOMD is exact. Unfortunately, obtaining the wave functions in every time step results in a prohibitively large effort of computational resources, which gives an upper limit to the accuracy of the wave function. On the other hand, poorly converged wave functions lead to a drift in the total energy [8, 61]. Therefore, energy conservation can be considered as the lower limit to the accuracy of the wave function. In contrast to BOMD, in Car–Parrinello molecular dynamics (CPMD) [1] both systems, i.e., the electrons and the nuclei, evolve at the same timescale, effectively eliminating the explicit time-dependence of the quantum subsystem. The obvious advantage of this procedure is that the wave function is calculated only once, thus saving a lot of computer resources. The classical kinetic energy E_{kin}^e of the electrons was proposed by Car and Parrinello to be:

$$E_{\text{kin}}^e = \sum_i \frac{1}{2} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle, \quad (60)$$

where ϕ_i are the one-electron Kohn–Sham orbitals, $\dot{\phi}_i$ is the time derivate of ϕ_i and μ is an arbitrary parameter, denoted as the fictitious mass of each electron. To keep the orbitals orthonormal one must introduce the constraint $\mathcal{E}^{\text{KS}}[\{\phi_i\}, \mathbf{R}^N] = E^{\text{KS}}[\{\phi_i\}, \mathbf{R}^N] + \sum_{ij} \Lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij})$.

The Lagrangian $\mathcal{L}_{\text{CP}} \equiv \mathcal{L}_{\text{CP}}[\mathbf{R}^N, \dot{\mathbf{R}}^N, \{\phi_i\}, \{\dot{\phi}_i\}]$ postulated by Car and Parrinello [1] finally reads:

$$\begin{aligned} \mathcal{L}_{\text{CP}}[\mathbf{R}^N, \dot{\mathbf{R}}^N, \{\phi_i\}, \{\dot{\phi}_i\}] = & \sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 \\ & + \sum_i \frac{1}{2} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle - \mathcal{E}^{\text{KS}}[\{\phi_i\}, \mathbf{R}^N]. \end{aligned} \quad (61)$$

Note, that the extended Kohn–Sham energy functional $\mathcal{E}^{\text{KS}}$ is dependent on the orbitals $\{\phi_i\}$ and $\mathbf{R}^N$. The corresponding Newtonian equations of motion are derived from the associated Euler–Lagrange equations for the nuclear

positions like in classical mechanics:

$$\frac{d}{dt} \frac{\partial \mathcal{L}_{CP}}{\partial \dot{\mathbf{R}}_I} - \frac{\partial \mathcal{L}_{CP}}{\partial \mathbf{R}_I} = 0, \quad (62)$$

and for the orbitals:

$$\frac{d}{dt} \frac{\partial \mathcal{L}_{CP}}{\partial \langle \dot{\phi}_i |} - \frac{\partial \mathcal{L}_{CP}}{\partial \langle \phi_i |} = 0. \quad (63)$$

This leads to the Car–Parrinello equations of motion:

$$M_I \ddot{\mathbf{R}}_I(t) = - \frac{\partial E^{KS}}{\partial \mathbf{R}_I} + \sum_{ij} \Lambda_{ij} \frac{\partial}{\partial \mathbf{R}_I} \langle \phi_i | \phi_j \rangle, \quad (64)$$

$$\mu \ddot{\phi}_i(t) = - \frac{\delta E^{KS}}{\delta \langle \phi_i |} + \sum_j \Lambda_{ij} | \phi_j \rangle. \quad (65)$$

Because this Lagrangian does not depend explicitly on time, the (physically meaningless) total energy H_{CP} :

$$H_{CP} = \sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle + E^{KS} = E_{kin}^I + E_{kin}^e + E^{KS} \quad (66)$$

of the whole system is conserved. To guarantee that sum of the kinetic energy of the nuclei, E_{kin}^I (and the potential energy E^{KS}) remains nearly constant, the parameter μ is chosen small enough that E_{kin}^e is several orders of magnitude smaller than $E_{kin}^I + E^{KS}$. With this choice of μ , one creates a coupled classical system of two energy reservoirs. The “hot” nuclei, and the “cold” electrons with an unequal distribution of energy [62]. According to the equipartition theorem, which states that the energy is distributed equally between all quadratic degrees of freedom in an equilibrated physical system, one would expect a constant flow of energy from the “hot” nuclei to the “cold” electrons during the course of simulation. This would destroy the conservation of E_{kin}^I . Furthermore, with rising temperature the wave function deviates more and more from a perfectly converged wave function for the given positions of nuclei. In this case, it is commonly said that the electrons leave the Born–Oppenheimer surface [8]. Therefore, one must suppress this energy flow to maintain the adiabatic separation. To do this one can consider adjusting the parameter μ or the introduction of thermostats for both energy reservoirs [63], as shown in Sect. 2.2.

5.1

Fictitious Mass μ

Pastore et al. [62] as well as Tagney [64] pointed out that the adiabatic separation depends crucially on the separation of the spectrum for the ionic

vibrations and the ultrahigh-frequency oscillation of the orbitals. A nice review on the coupling between the electronic excitations and the phonons was written by Chester [65]. Assuming a linear Kohn–Sham Hamiltonian and equal occupation of all occupied orbitals, the frequency for the electronic excitation can be approximated as:

$$\omega_{ij} = (f_j(\varepsilon_i^* - \varepsilon_j)/\mu)^{1/2}, \quad (67)$$

where f_j is the occupation number of the Kohn–Sham orbital ϕ_j , ε_i^* is the eigenvalue of the unoccupied orbital ϕ_i , and ε_j is the eigenvalue of ϕ_j . The lowest electronic excitation frequency $\omega_{\min}$, i.e., the electronic frequency closest to the frequencies of the nuclei, is therefore determined by the Kohn–Sham gap. Obviously, $\omega_{\min}$ can be adjusted by μ . This gives rise to the idea of shifting μ towards zero, giving the best possible adiabatic separation. However, this idea is limited by the fact that the largest possible time step for the simulation depends on the inverse of the highest electronic frequency $\omega_{\max}$ [8]. Therefore, choosing μ means compromising between a sufficiently large frequency gap and a reasonable timescale for the simulation. A second problem arises for systems with a small, or even without any, Kohn–Sham gap, especially metals. In this case, the adiabatic separation breaks down and cannot be saved by a finite μ , because $\omega_{\min}$ goes to zero and the frequency spectra of the ions and the electrons overlap. Here, thermostats for both the electrons and the ions can be used to maintain the adiabatic separation [66]. Tagney et al. [64, 67] could also show that the deviation of the forces of the CPMD simulation compared to BOMD is proportional to the fictitious mass μ . For purely ionic systems, where the electronic structure of one atom is not perturbed by the neighboring atoms, this difference reduces to a simple rescaling of the ionic masses, which can be corrected a priori [67]. For more delocalized systems, this difference was shown to be smaller, but cannot be corrected any further.

6 QM/MM

Despite the success of the Car–Parrinello method, biological systems like proteins or systems where explicit solvent effects are observed are too large to be described by density functional theory. For such systems, several hybrid QM/MM schemes have been developed, rooting back to the original idea proposed by Warshel and Levitt [68]. The basic idea of all the hybrid schemes is that the system is divided into one (chemically active) region described by the quantum mechanical (QM) method and one surrounding region described by classical molecular mechanics (MM). The crucial point of the hybrid methods is the coupling Hamiltonian that describes the interface region between the QM and the MM part. Referring to classical molecular mechanics, the con-

tributions to the coupling Hamiltonian can be divided into bonded and non-bonded interactions. Bonded interactions only arise when the QM/MM partitioning breaks existing bonds. In this case, the related MM atom has to be replaced by a hydrogen atom or a suitable pseudopotential [69–71] to avoid unsaturated valence orbitals on the related QM atom. Another solution to this problem is the so-called frozen orbital method [70, 72–74], where suitable orbitals are located on the related MM atom. These orbitals are not evolved during the simulation, but nevertheless raise the computational cost for plane wave codes significantly [75]. Angles and dihedral terms are completely described by the underlying classical force field [75]. For the discussion of the non-bonded interaction, we restrain ourselves to hybrid schemes involving the CPMD method [75–78]. Non-bonded interactions in classical molecular dynamics consist of the van der Waals contribution and the electrostatic interaction. The van der Waals interaction can be expressed by a Lennard–Jones potential H_{vdW} :

$$H_{\text{vdW}} = \sum_{\substack{i \in \text{MM} \\ j \in \text{QM}}} 4\epsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{|\mathbf{R}_i - \mathbf{R}_j|} \right)^{12} - \left(\frac{\sigma_{ij}}{|\mathbf{R}_i - \mathbf{R}_j|} \right)^6 \right\}, \quad (68)$$

where QM denotes the set of all atoms in the QM region, MM denotes the set of atoms of the MM region, and $\mathbf{R}_k$ denotes the position of one nucleus k in the QM region or one atom k in the MM region. The construction of an appropriate electrostatic interaction is the most challenging point. In general, the electrostatic Hamiltonian H_{el} , describing the interaction between the electrons of the QM system with the charges of the MM system, is given by:

$$H_{\text{el}} = \sum_{j \in \text{MM}} q_j \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_j|}, \quad (69)$$

where q_j is the charge of the MM atom at the position $\mathbf{R}_j$ and $\rho(\mathbf{r})$ is the charge density of the QM part at the electron position $\mathbf{r}$.

Therefore, a partially positively charged MM atom represents a trap for the electron density of the QM part, because the electrostatic interaction goes to infinity if the electron approaches the MM atom. This phenomenon is called the spill-out effect (Fig. 4) and is attributed to the missing repulsion of the core electrons on the MM atoms. All schemes rely on modifying the electrostatic potential of the MM atoms in the vicinity of the QM region, i.e., by altering the short range Coulomb interaction such that H_{el} remains finite for $|\mathbf{r} - \mathbf{R}_j| \rightarrow 0$.

Eichinger et al. [76] replaced the point charges of the MM atoms close to the QM part by a Gaussian distribution of charges. Thus, their electrostatic

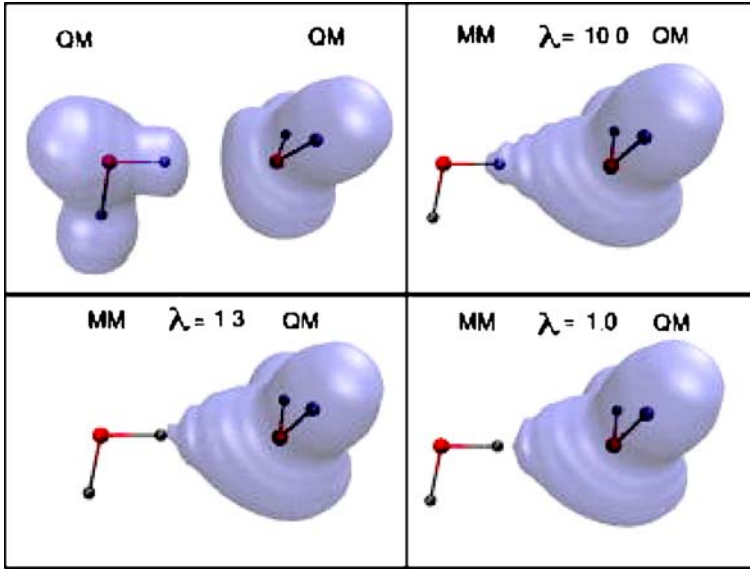


Fig. 4 Plot of electron localization function (ELF) for a water dimer as obtained from a full quantum calculation employing CPMD and a QM/MM calculation employing GROMACS-CPMD. $\lambda = 10$ is equivalent to a pure $1/r$ electrostatic potential, $\lambda = 1$ is equivalent to a modified electrostatic interaction with a finite value for $r \rightarrow 0$. Reprinted with permission from [77]. Copyright 2005, American Institute of Physics

Hamiltonian finally reads:

$$H_{\text{el}} = \sum_{j \in \text{MM}} q_j N_\rho \int d\mathbf{r} \rho(\mathbf{r}) \frac{\text{erf}(|\mathbf{r} - \mathbf{R}_j|/\sigma)}{|\mathbf{r} - \mathbf{R}_j|}, \quad (70)$$

where N_ρ is the normalization constant for the electronic density and erf is the error function.

A second approach was undertaken by Laio et al. [75], where H_{el} was modified to:

$$H_{\text{el}} = \sum_{j \in \text{MM}} q_j \int d\mathbf{r} \rho(\mathbf{r}) v_j(|\mathbf{r} - \mathbf{R}_j|), \quad (71)$$

with:

$$v_j(|\mathbf{r} - \mathbf{R}_j|) = \frac{r_{cj}^n - |\mathbf{r} - \mathbf{R}_j|^n}{r_{cj}^{n+1} - |\mathbf{r} - \mathbf{R}_j|^{n+1}}, \quad (72)$$

where r_{cj} is the covalent radius of the atom j and n is an empirically determined parameter.

One further proposal was given by Biswas et al. [77], who expanded the charge q_j on the MM atom in terms of hydrogen-like wave functions

$\phi(\mathbf{r}' - \mathbf{R}_j)$:

$$\phi(\mathbf{r}' - \mathbf{R}_j) = \left(\frac{q_j \xi^3}{\pi} \right)^{1/2} e^{-\xi |\mathbf{r}' - \mathbf{R}_j|}, \quad (73)$$

where ξ can be freely adjusted. The electrostatic Hamiltonian in this case reads:

$$H_{\text{el}} = \sum_{j \in \text{MM}} N_\rho \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r}) |\phi(\mathbf{r}' - \mathbf{R}_j)|^2}{|\mathbf{r}' - \mathbf{r}|}. \quad (74)$$

All of these Hamiltonians approach a finite value of H_{el} for $|\mathbf{r} - \mathbf{R}_j| \rightarrow 0$ and show a $1/|\mathbf{r} - \mathbf{R}_j|$ behavior for long distances. In order to save CPU time, the explicit evaluation of the Coulomb interaction with every MM atom can be replaced by a multipole expansion [75] of the electrostatic field of the QM region:

$$\begin{aligned} \int d\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_j|} &= \mathcal{C} \frac{1}{|\mathbf{R}_j - \bar{\mathbf{R}}|} + \sum_{\alpha=1}^3 \mathcal{D}^\alpha \frac{(\mathbf{R}_j^\alpha - \bar{\mathbf{R}}^\alpha)}{|\mathbf{R}_j - \bar{\mathbf{R}}|^3} \\ &+ \frac{1}{2} \sum_{\alpha, \beta=1}^3 \mathcal{Q}^{\alpha\beta} \frac{(\mathbf{R}_j^\alpha - \bar{\mathbf{R}}^\alpha)(\mathbf{R}_j^\beta - \bar{\mathbf{R}}^\beta)}{|\mathbf{R}_j - \bar{\mathbf{R}}|^5} + \mathcal{O} \left(\frac{R_q}{|\mathbf{R}_j - \bar{\mathbf{R}}|} \right)^4 \\ &= H_{\text{lr}}(j) + \mathcal{O} \left(\frac{R_q}{|\mathbf{R}_j - \bar{\mathbf{R}}|} \right)^4, \end{aligned} \quad (75)$$

where $\bar{\mathbf{r}}$ is the geometrical center of the quantum subsystem and therefore the origin of the multipolar expansion. $\mathcal{C}$ is the monopole, $\mathcal{D}^\alpha$ is the dipole, and $\mathcal{Q}^{\alpha\beta}$ is the quadrupole. They are given by:

$$\mathcal{C} = \int d\mathbf{r} \rho(\mathbf{r}), \quad (76)$$

$$\mathcal{D}^\alpha = \int d\mathbf{r} \rho(\mathbf{r}) (\mathbf{r}^\alpha - \bar{\mathbf{R}}^\alpha), \quad (77)$$

$$\mathcal{Q}^{\alpha\beta} = \int d\mathbf{r} \rho(\mathbf{r}) \left[3(\mathbf{r}^\alpha - \bar{\mathbf{R}}^\alpha)(\mathbf{r}^\beta - \bar{\mathbf{R}}^\beta) - \delta^{\alpha\beta} |\mathbf{r} - \bar{\mathbf{R}}|^2 \right]. \quad (78)$$

This multipole expansion is only used for the potential seen by atoms outside an arbitrarily chosen cut-off radius R_c around every quantum atom. All atoms inside R_c see the exact, non-expanded potential.

7

Literature on Biological Systems

In the last decade, Car-Parrinello molecular dynamics simulations have developed into a powerful tool for the investigation of biomolecules such as proteins, peptides, enzymes, or even DNA and RNA. We present a brief review of applications of CPMD in this area.

7.1

Proteins, Peptides, and Amino Acids

One of the strengths of CPMD is the possibility of observing spontaneous events. This can be utilized for the investigation of systems where proton transfers or the formation and breakage of hydrogen bonds are expected. Examples of such investigations are: the simulation of the proton transfer in a polyglycine analog of the gramicidin A ion channel, by Klein and coworkers [79]; the study of the dynamics and energetics of a fully hydrated model complex of the pSer133 and the corresponding part of the CREP-binding protein (CBP), by Carloni and coworkers [80]; simulations of the deprotonation of histidine residues in aqueous solution, by Ivanov and Klein [81]; and studies of the role of intramolecular hydrogen bonds for the torsion of arylamide derivatives, by Doerksen et al. [82].

Further examples of protein CPMD simulations are the examination of the binding of copper ions to prion proteins by the R  thlisberger group [83, 84], and a study by Gageot and Sprik where they demonstrated the applicability of CPMD for the computation of infrared absorption spectra in condensed molecular systems based on a study of aqueous uracil [85].

Retinal derivatives as chromophores in many light-detecting proteins, such as the visual pigment rhodopsin and the proton pump bacteriorhodopsin, have been the subject of detailed investigations: First molecular dynamics studies of the retinylidene chromophore of rhodopsin were performed by Buda, Bifone, and coworkers [86–89]. R  hrig et al. studied the structure and dynamics of the retinal chromophore of rhodopsin in vacuum, dissolved in methanol, and in the protein binding pocket [90] and the photoreaction in rhodopsin [91]. Marx and coworkers investigated the protonated water networks on the extracellular side of bacteriorhodopsin [92, 93].

Liquid *N*-methylacetamide was studied by a combination of classical and Car-Parrinello molecular dynamics by Whitfield et al. [94]. Investigation of the hypothetical hydroxide and proton migration along the linear water chain in aquaporins was carried out by Jensen, R  thlisberger, and Rovira [95]. Simulations of the zwitterionic form of the dipeptide glycine-alanine in water were undertaken by Hugosson [96]. The 2D (ϕ, ψ) free energy surface of alanine dipeptide in both gas phase and aqueous solution phase were determined [97] by means of the meta-dynamics approach [98, 99].

7.2

Metal-Containing Biomolecules: Metallo-enzymes

Car-Parrinello molecular dynamics simulations are also established as a widespread tool for the investigation of metal-containing biomolecules such as metallo-enzymes. Heme models belong to the first biologically relevant systems that were studied by Car-Parrinello simulations. Parrinello, Rovira, and coworkers gained insight into the binding of O₂, CO, and NO to iron porphyrin [100–103]. They extended the investigation of these reactions to the natural system, the myoglobin protein [104–106]. Recently Spiro and coworkers studied the influence of a distal histidine residue on O₂ and CO binding in myoglobin [107].

Further early applications of the CPMD method on metal-containing systems were the studies on the role of hydrogen bonds for the enzymatic function of serine protease by De Santis and Carloni [108], and the simulation of model systems for the adduct between the HIV-1 reverse transcriptase and nucleosidic phosphate by Alber and Carloni [109]. HIV-1 protease and its enzymatic reaction are also under intensive study by Piana et al. [110–113].

Molecular dynamics studies of a biomimetic model of galactose oxidase were performed by Röthlisberger and Carloni [114]. Consecutively, they investigated together with Doclo and Parrinello the catalytic properties of the galactose oxidase and the synthetic analog of the active site on the basis of QM/MM Car-Parrinello calculations [115].

Further examples of such systems for investigation using Car-Parrinello simulations are: the enzymatic catalysis of D-alanine oxidation by the flavoenzyme D-amino acid oxidase [116]; the study of the initial and fundamental step in enzymatic GTP hydrolysis, by Cavelli and Carloni [117]; and the investigation of different protomeric states of methionine aminopeptidase active site, by Klein et al. [118].

In addition, Magistrato et al. studied the structural and dynamical properties of the dizinc analog of the synthetic diiron protein Due Ferri 1 [119], and Klein and coworker investigated the Glu₄His₂Zn₂ cofactor that is found in the metallo-protein Duo Ferro 1 [120].

Carloni and coworker carried out QM/MM simulations of the hydrolysis of cefotaxime by the monozinc β -lactamase [121] and of the enzymatic activity of orotidine monophosphate decarboxylase [122].

More recent applications are: the study of the role that magnesium cations play in the cleavage reaction of fully hydrated RNA enzymes, by Boero et al. [123]; simulations to gain insight into the dynamical flexibility of the bridged binuclear structural motif in the active site of arginase, by Ivanov and Klein [124]; investigation of the initial dinitrogen reduction step in Sellmann-type nitrogenase model complexes [125]; and the study of the binding mode of anticancer ruthenium(II) arene complexes to double-stranded DNA, by

Gossens et al. [126]. Alagano et al. compared the results of quantum mechanics/free energy perturbation (QM/FEP) calculations of the second proton transfer step in triosephosphate isomerase with the results of QM/MM Car-Parrinello simulations, in order to estimate the bias of the FEP simulations [127].

Recently, Rovira and coworkers studied the bacillus 1,3-1,4- β -glucanase [128]. The hydrogen-bonding pattern between phosphinate and phosphonate inhibitor to aspartic protease was investigated by Vidossich and Carloni [129].

Giannozzi et al. demonstrated, using the example of porphyrine representatives, that their ultrasoft pseudopotential implementation of first-principle molecular dynamics is suitable for the investigation of biologically relevant metallo-organic systems [130].

7.3

DNA and RNA

Up to now there have been only a few applications of the Car-Parrinello methods to DNA and RNA. One of the first applications of the CPMD code in this context was the calculation of nucleic acids under laboratory-realizable conditions, performed by Hutter et al. for sodium guanylyl-3'-5'-cytidine nonahydrate [131]. Molecular dynamics simulations of such systems were performed by Carloni et al. [132]. They simulated the hydrolysis of *cis*-platin as well as the final step of the binding process to pGpG dinucleotide, as model for the DNA.

Further studies were concerned with the hydration and dynamics of arginine-phosphate adducts in proteins, by Frigyes et al. [133]; with the interaction between DNA and radicals by calculating the reaction mechanisms of the reaction between isolated guanine and an OH radical, by Mundy et al. [134]; with the QM/MM examination of the second step of the enzymatic reaction of caspase-3 (namely the hydrolysis of the acyl-enzyme complex), by Sulpizi et al. [135, 136]; with the investigation of the structure and binding of *cis*-platin to DNA oligomers in aqueous solution, by Spiegel et al. [137]; with the mixed QM/MM study of the DNA influence on the reactivity of the guanine radical cation by the Parrinello group [138]; and with the simulation of the OH radical reaction with guanine or thymine in water in order to obtain a better understanding of the OH radical-induced DNA-base damage, by Car and coworkers [139].

Recently a fully hydrated polyd(GpCp) DNA strand was investigated by Gervasio et al. in order to gain a deeper insight into the charge localization in DNA fibers [140]. Spiegel et al. studied the binding of duocarmycins to DNA [141].

7.4

Miscellaneous, Biomimetics

In this section we will briefly review some further applications of the CPMD method on biological relevant molecules. Early examples are: the application of the projector augmented wave (PAW) method of Blöchl [142] on the investigation of penta-*O*-methyl- α -D-glycopyranose inversion and penta-*O*-methyl- β -D-glycopyranose inversion by Ionescu et al. [143] and the simulation of the anomeric equilibrium of glucose in aqueous solution by Molteni and Parrinello [144].

Later, Rovira et al. studied the protonation state of dimethylglyoxime ligands and the dynamics of $\text{OH} \cdots \text{H}$ units in a B_{12} coenzyme biomimetic system [145]. Schwarz and coworkers could clarify the unusual temperature effects on the dissociation patterns of valeramide radical cations observed in mass spectrometric experiments by simulation of neutral and ionized valeramide [146]. The solvation of 2,2'-dihydroxyl-1,1'-binaphthyl in water was simulated by Entel, Kreth et al. [147].

Recent applications are: studies of the dynamics and NMR properties of a vanadate-glycylglycine complex in aqueous solution [148]; investigation of the role of the aqueous solvent for the acid-catalyzed glycosidic bond formation, by Stubbs and Marx [149]; investigations of the protonation-induced diastereoisomerism in nicotine, which may have decisive influence on the dynamical behavior of the molecules, by Hammond et al. [150]; and simulations of the shape and flexibility of arylamide polymers by classical molecular dynamics and CPMD calculations, which resulted in the development of new torsional parameters for classical MD simulations of these compounds [151].

7.5

Excited States MD

Car-Parrinello molecular dynamics in excited states has become possible in recent years: The Parrinello group introduced a Kohn-Sham-like formalism for the treatment of excited singlet states [152, 153]. The Röthlisberger group implemented this approach in their mixed QM/MM Car-Parrinello code [83].

Bittner and Kosov used the random phase approximation for the description of electronically excited systems with Car-Parrinello molecular dynamics [154]. Doltsinis and Marx applied a method for an efficient treatment of electronically non-adiabatic processes with CPMD by using a surface hopping scheme for the coupling of the S_1 restricted open-shell Kohn-Sham excited state to the S_0 ground state [155].

Hutter's implementation of time-dependent linear response density functional theory (LR-TDDFT) to plane wave basis set methods allowed the simulation of photochemical reactions and fluorescence spectra [156, 157].

Recently, R  thlisberger and coworkers presented implementations of a linear response and time propagation scheme to molecular dynamics simulations [158].

A few examples of applications of these methods to systems of biological interest are: the investigation of rhodopsin by Molteni et al. [159]; hybrid QM/MM simulations of the photoreaction in rhodopsin, by the R  thlisberger group [91]; the calculation of optical properties of molecules in solution, by Sulpizi et al. [160, 161]; simulations of guanine [162, 163] and methylated guanine [164], by Langer et al.; the study of photoactive proteins, by the R  thlisberger group [165]; and the investigation of the photochemical activation scheme of dinitrogen by dinuclear Ru^{II} and Fe^{II} complexes [166].

8

Case Study: Biomimetic Nitrogen Fixation

In this section we briefly review an example of a Car–Parrinello molecular dynamics simulation, namely the initial dinitrogen reduction step in Sellmann-type nitrogenase model complexes [125]. These complexes are of special interest because they are supposed to enable access to artificial nitrogen fixation at ambient conditions. An introduction to biological nitrogen fixation mediated by the enzyme nitrogenase can be found in recent literature, see [167–176].

Sellmann-type complexes are based on the so-called open-side FeMo cofactor (FeMoco) systems, and have only recently been substantiated by some DFT calculations [172, 173]. Sellmann assumed that FeMoco (a transition metal–sulfur cluster, which is the core of the active center of the enzyme nitrogenase) opens so as to bind nitrogen at the iron centers. In fact, the synthesis of $[\mu\text{-N}_2\{\text{Ru}(\text{P}i\text{Pr}_3)(\text{“N}_2\text{Me}_2\text{S}_2\text{”})\}_2]$ [$\text{“N}_2\text{Me}_2\text{S}_2\text{”}^{2-} = 1,2\text{-ethanediamine-}N,N'\text{-dimethyl-}N,N'\text{-bis(2-benzenethiolate)(2-)}]$], which is up to now the only known example of a stable dinuclear metal–sulfur N_2 complex, was accomplished in 2001 [177, 178]. In further experiments, Sellmann and coworkers tried to establish a reduction pathway to the corresponding diazene complex $[\mu\text{-N}_2\text{H}_2\{\text{Ru}(\text{P}i\text{Pr}_3)(\text{“N}_2\text{Me}_2\text{S}_2\text{”})\}_2]$, which had been already fully characterized [179].

Instead of the reduction to diazene, various side reactions were observed during the experiments. This can be explained by the fact that the first endothermic reduction step is already difficult to achieve [180]. However, the results indicate that these model complexes are not suitable for the desired behavior to reduce nitrogen, and not that Sellmann-type complexes are in principle not able to succeed in the reduction of nitrogen.

Car–Parrinello molecular dynamics simulations were performed in order to clarify whether the reduction of nitrogen with Sellmann-type complexes is possible in principle or not. The advantage of Car–Parrinello simulations

is the detailed time-resolved picture, which is needed for the discussed processes.

Two different complexes were investigated: a small $[\text{Ru}''\text{S}_2\text{N}_2'']_2\text{L}$ model complex (1N_2) and a larger $[\text{Fe}''\text{N}_\text{H}\text{S}_4'']_2\text{L}$ complex (2N_2) depicted in Fig. 5.

These systems were chosen because they are similar to the first experimentally observed dinitrogen metal–sulfur complex [178, 179, 181] and because an investigation of complexes with chelate ligands of different rigidity allows elucidation of the general properties of such complexes. The starting points of our investigation are the doubly protonated complexes 1N_2^{2+} and 2N_2^{2+} . A twofold reduction of these systems leads to complex **A1** on the corresponding neutral electronic energy surface, on which the subsequent reaction takes place and finally stops at the diimine **B** (see Fig. 6). **A2** is obtained if the preparation of **B** during the course of the reaction is stopped.

The presented products **A1** and **A2** of the twofold reduction are high-energy structures on the neutral potential energy surface. Hence, it can be

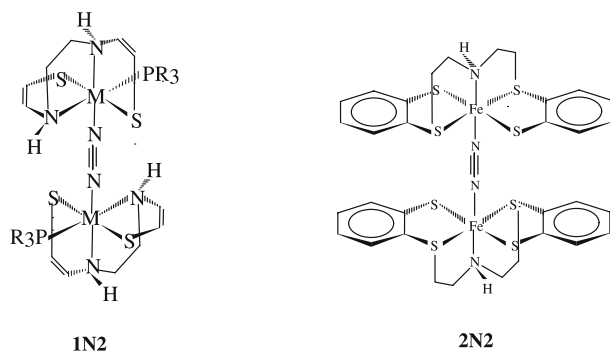


Fig. 5 Sellmann-type $\text{Ru}''\text{S}_2\text{N}_2''$ model complex 1N_2 (left) and the $\text{Fe}''\text{N}_\text{H}\text{S}_4''$ complex (2N_2) (right)

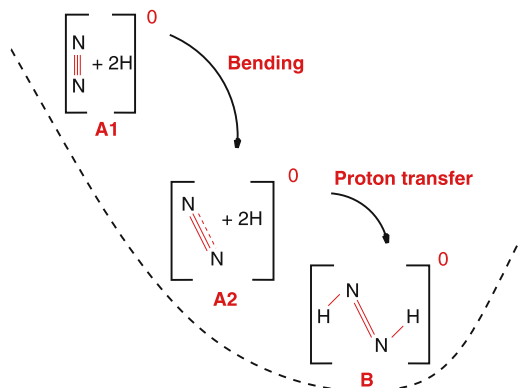


Fig. 6 Schematic drawing of initial reduction pathway

expected that spontaneous events take place and it can be observed how A1 and A2 behave under the effects of temperature, forced bending, and forced decrease of the nitrogen proton distances. This can indicate a potential mechanism and a possible reaction coordinate for this system. It is not possible to reach extended statistics because of the very expensive CPMD calculations.

A detailed description of the methodology is presented in the original publication [125].

Constrained and free simulations of complex $1N_2$ and $2N_2$ were run at different temperatures. In one experiment both hydrogen MNN angles were forced to decrease with time in complex $2N_2$. This was done because it had been shown before that the bending of the nitrogen moiety is an important step for the reaction [166]. The results of these experiments are depicted in Fig. 7. Because the symmetry of the complex is preserved during the reaction, only one curve is shown. Temperature regulation was used to prevent the chelate ligands from decomposing. The initiation of the reaction can be recognized at 600 fs. It involves a sequence of complex processes. The increase of the Kohn-Sham energy and temperature, which can be observed in Fig. 8, points to a small barrier at an initial time step of 150 fs.

This barrier is accompanied by:

1. A small elongation (5 pm) of the N – N distance, see bold black line in the left panel of Fig. 7,
2. An increase (150 pm) of the hydrogen–nitrogen atom distance, see black line in the left panel of Fig. 7,
3. An elongation (15 pm) of the metal–nitrogen distance (dashed line in the left panel of Fig. 7) and an increase of the FeSH angle (20 degrees), see right panel of Fig. 7.

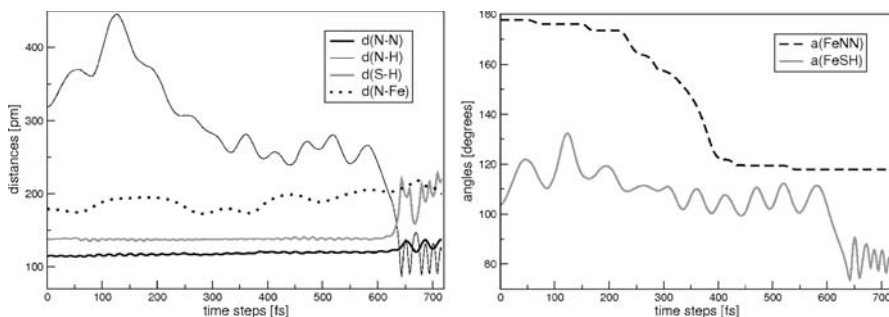


Fig. 7 Results of the simulation of the large Fe complex $2N_2$, where the MNN angles are forced to decrease in time. *Left panel:* Nitrogen–nitrogen distance (**bold black line**), nitrogen–hydrogen distances (**black line**), sulfur–hydrogen distance (**gray line**), and N – M distances (**dashed line**). *Right panel:* MNN angles (**dashed line**) and MSH angles (**gray line**). All data are plotted against the time step in fs. Distances are in pm and angles in degrees. Picture reproduced from data of [125]

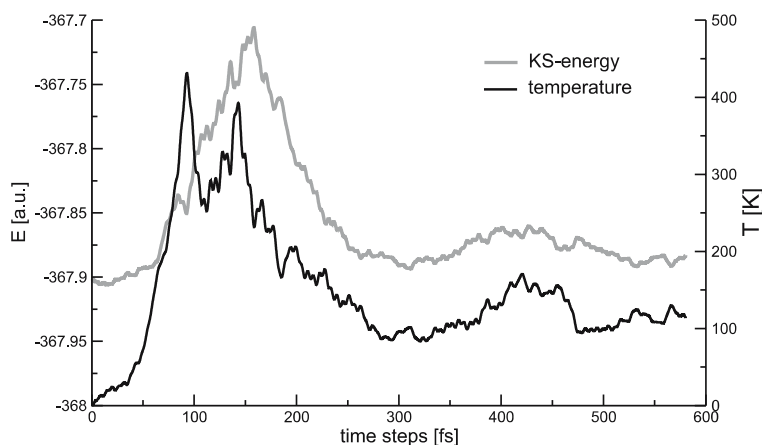


Fig. 8 Forcing both metal–nitrogen–nitrogen angles to decrease: Kohn–Sham energy (gray line) and temperature (black line) plotted against time steps. Picture reproduced from data of [125]

On the basis of these results it can be assumed that in the starting configuration A1 the bending of the N_2 moiety is structurally hindered. At a time step of about 300 fs a further small increase of the Kohn–Sham energy of about 0.025 a.u. appears. It can be explained by a slight barrier for the initiation of the bending of the FeNN angle (see right panel of Fig. 7). The results of these constraint simulations indicate that bending of the $M_1N_1N_2$ and $M_2N_2N_1$ angles is necessary for a successful reaction. It seems apparent that this bending of the N_2 moiety is prohibited unless the hydrogen atoms leave the vicinity of the nitrogen atoms. At this point it is important to note that the transfer of the hydrogen is a concerted process. If only one angle is constrained, a simultaneous reaction takes place nevertheless.

In different experiments, free simulations without any constraints for both Sellmann models $1N_2$ and $2N_2$ were carried out. The results of simulations at 50 K, where the starting point is a pre-bent structure (A2 of Fig. 6) of the small complex $1N_2$, are presented in Fig. 9. The reaction of the N_2 moiety with one hydrogen atom ($H_2 - N_2$) begins after about 450 fs. However, in this case the transfer of the hydrogen atoms is not a simultaneous process.

It can be understood from the left upper panel of Fig. 9 that at about 450 fs the vibration of the dinitrogen bond is enhanced and the NN bond elongates. The vibrations with small amplitudes around 121.5 pm belong to the original triple bond of N_2 , whereas the vibrations with a larger amplitude around an average NN distance of 129.2 pm can be assigned to a NN double bond. The bottom left panel of Fig. 9 shows that the vibrational motion of both NH bonds is anticorrelated. The increase of one NH distance (dashed curve) at the beginning enables the other hydrogen atom to approach its nitrogen atom. Then, an oscillating motion of the hydrogen atoms can be observed

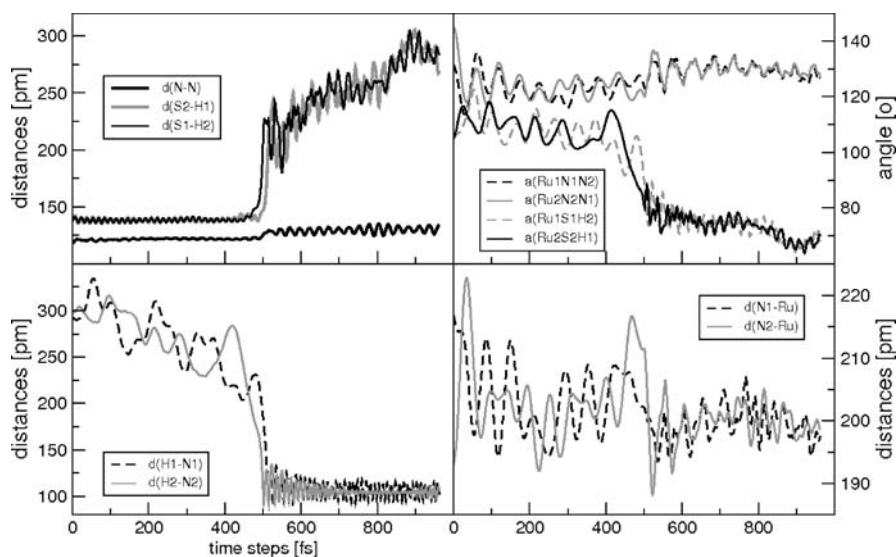


Fig. 9 Free simulation of the bent small Sellmann-type complex $1N_2$: *Left upper panel*: Nitrogen–nitrogen distance and sulfur–hydrogen distances. *Right upper panel*: RuNN and RuSH angles. *Bottom left panel*: Nitrogen–hydrogen distances. *Bottom right panel*: Nitrogen–Ru distances. All data are plotted against the time step in fs. Distances are in pm and angles in degrees. Picture reproduced from data of [125]

until, at approximately 450 fs, the distance between one hydrogen and the corresponding nitrogen atom drops down rapidly to about 100 pm. This shows that an angle as well as a distance constraint is not an appropriate reaction coordinate, because the reaction cannot be described by only one simple reaction coordinate. Only 15 fs later the second hydrogen atom is transferred.

After 450 fs, the distances between the S atoms and the corresponding H atoms, as well as the amplitude of the vibrational motions, increase abruptly. This also affects the curve of the RuSH angles; they decrease to below $\angle 90^\circ$. The bottom right panel of Fig. 9 shows that the N–Ru distances do not change much upon reaction but that they vibrate with smaller amplitude after the hydrogen transfer. Again, the Kohn–Sham energy and the temperature are very good indicators for the course of the reaction (Fig. 10). The proton transfer at 450 fs is connected with a notable drop of the Kohn–Sham energy and a heating-up of the system.

In further experiments [125], the behavior of the Fe complex $2N_2$ at a temperature of 100 K without any constraints was investigated. The results of these simulations in Fig. 11 demonstrate successful events. In this case it takes about 1050 fs until the first proton transfer proceeds, and the delay of the second proton transfer is almost 100 fs. The bottom left in Fig. 11 indicates that the hydrogen atoms move away from the corresponding nitrogen

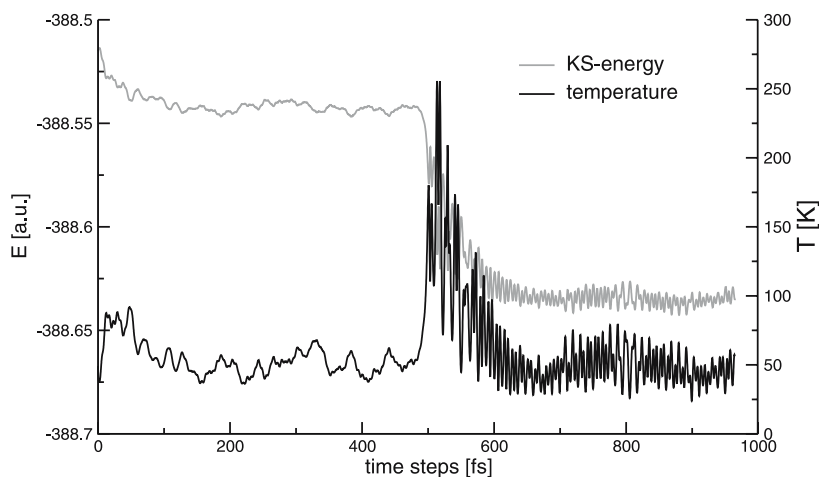


Fig. 10 Kohn–Sham energy (gray line) and temperature (black line) plotted against time steps. Picture reproduced from data of [125]

atom before the bending of the FeNN angles takes place. The vibrational motion of these angles is smooth until proton transfer begins. The S – H bonds are not elongated until the beginning of the proton transfer. In the first 800 fs the N – Fe distance increases by 20 pm only at the side of the complex where the reactions starts (see dashed curve in the upper left panel of Fig. 11).

From the results of the simulations that led to successful events, it can be concluded [125] that the bending motion of the MSH angle plays a crucial role in determining whether the reaction takes place. This motion induces an increase in the N–H distance, which in turn induces the start of the bending motion of the MNN angle. Even for a pre-bent structure the hydrogen nitrogen distance has to grow in order to enable the MNN angle bending mode. A comparison of the results of the free simulations with the constraint simulation shows that we observe for the latter a simultaneous proton transfer whereas the former is a stepwise process. The constraints lead to a somewhat unphysical behavior as is evident from the right panel of Fig. 7, where the curve of the FeNN angles shows no vibrational motion.

The above case study of the initial proton electron transfer step in Sellmann-type nitrogenase model complexes has shown that Car–Parrinello molecular dynamics is a powerful tool for the investigation of complex reactions. On the basis of these simulations, a three-step reaction mechanism was suggested for both investigated complexes: The N–H distance has to first increase. Hereby a bending of the MNN angles is allowed, which leads to an activated diazenoid N₂ species. Vibrations of the MSH angle effect vibrations of the N–H distances, which are in opposite phase to simultaneous vibrations of the MNN angles. This finally leads to the observed proton transfer, because

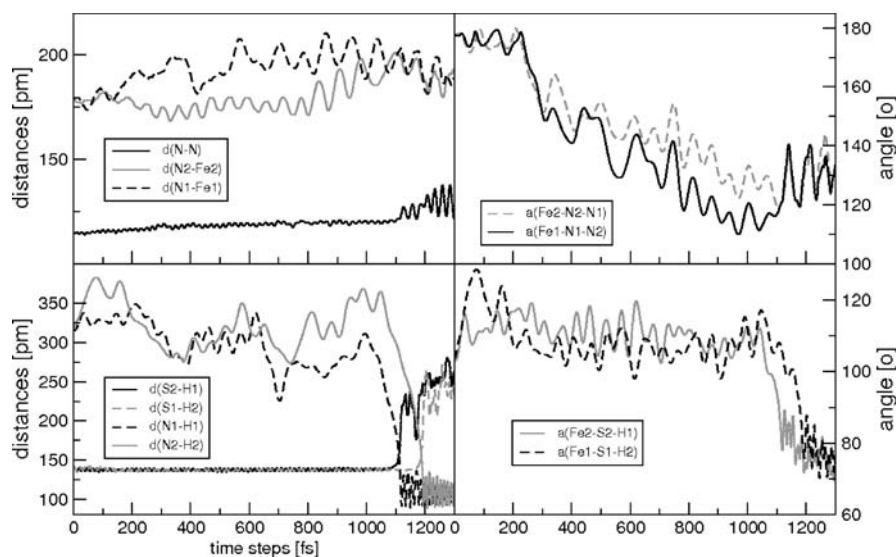


Fig. 11 Free simulation results of the large Fe complex $2N_2$ at 100 K. *Left upper panel:* Nitrogen–nitrogen distance and nitrogen–iron distances. *Right upper panel:* FeNN angles. *Bottom left panel:* Nitrogen–hydrogen distances and sulfur–hydrogen distances. *Bottom right panel:* FeSH angles. All data are plotted against the time step in fs. Distances are displayed in pm and angles in degrees. Picture reproduced from data of [125]

the nitrogen atoms can get close to the hydrogen atoms. Since both complexes with chemically different chelate ligands exhibit an equivalent behavior, it can be concluded that the observed features of the reactions are universal [125]. Particularly important is the fact that the reduction of nitrogen is a spontaneous event, if a two-electron reduction of a Sellmann-type complex leads to a trans-diprotonated equilibrium structure. On the basis of our simulations we can also estimate that the time needed for the twofold proton transfer is in the order of 500–1000 fs in this case.

During the simulations a few side reactions are also observed. One of these side reaction consists of a partial dissociation of the chelate ligands of the metal atom. In another side reaction, a hydrogen transfer to the metal center is observed. These side reactions could be related [125] to side products observed in experiment.

9

Conclusion

We have presented an introduction to the Car-Parrinello simulation technique with a special emphasis on the applications to biomolecular systems.

Although many new developments and improvements are still in progress, this method is now amongst the standard tools of theoretical chemists. This is mostly due to the excellent program packages that were provided by the groups of Parrinello, Car, and Hutter, for example. A large step towards the appropriate description of biological systems in the Car–Parrinello framework is due to the developments of the Röthlisberger group. These developments include QM/MM coupling schemes and methodological developments related to this topic. Pioneering work on the field of Car–Parrinello simulations of biomolecules was also undertaken by Carloni and coworkers. Biological systems are complicated in several aspects. For instance, they can demand a quantum chemical treatment, they are large and extended systems, and a dynamical description is often necessary. Therefore the continuing developments contributed by many different groups are important and will aid the investigation of challenging topics such as those present in biology. We concentrated in this review only on the Car–Parrinello simulation technique, but we want to stress here that many important contributions also come from scientists working with similar techniques (not treated in this review), like for instance the tight-binding density functional theory method (TB-DFT) [182] or Born–Oppenheimer molecular dynamics simulations (BOMD) [142, 183–186].

Acknowledgements We gratefully acknowledge the financial support of the DFG priority program SPP 1191 “Ionic Liquids” and the ERA program, which allows productive collaboration under the project “A Modular Approach to Multi-responsive Surfactant/Peptide (SP) and Surfactant/Peptide/Nanoparticle (SPN) Hybrid Materials”. We would furthermore like to acknowledge the financial support from the collaborative research center SFB 624 “Templates” at the University of Bonn. Fruitful discussion with Jürg Hutter, Ursula Röthlisberger, and Markus Reiher are also gratefully acknowledged. JT would like to thank the Fonds der Chemischen Industrie for a Chemiefonds-Stipendium.

References

1. Car R, Parrinello M (1985) *Phys Rev Lett* 55:2471
2. Kirchner B, Reiher M (2002) *J Am Chem Soc* 124:6206
3. Kirchner B, Hutter J (2002) *Chem Phys Lett* 364:497
4. Kirchner B, Hutter J (2004) *J Chem Phys* 121:5133
5. Reiher M, Sellmann D, Hess BA (2001) *Theor Chim Acta* 106:379
6. Kirchner B (2005) *J Chem Phys* 123:204116
7. Remler DK, Madden PA (1990) *Mol Phys* 70:921
8. Marx D, Hutter J (2000) Ab-initio molecular dynamics: Theory and implementation. Modern methods and algorithms in quantum chemistry, Forschungszentrum Juelich, NIC Series 1:301
9. Tuckerman ME, Martyna GJ (2000) *J Phys Chem B* 104:159
10. Tuckerman ME (2002) *J Phys Condens Matter* 14:1297
11. Alder BJ, Wainwright TE (1957) *J Chem Phys* 27:1208

12. Ermakova E, Solca J, Huber H, Marx D (1995) *Chem Phys Lett* 246:204
13. Kirchner B, Ermakova E, Steinebrunner G, Dyson AJ, Huber H (1998) *Mol Phys* 94:257
14. Huber H, Dyson A, Kirchner B (1999) *Chem Soc Rev* 28:121
15. Kirchner B, Ermakova E, Solca J, Huber H (1998) *Chem Eur J* 4:379
16. Allen MP, Tildesley DJ (1987) *Computer simulations of liquids*. Clarendon, Oxford
17. Frenkel D, Smit B (2002) *Understanding molecular simulations*. Academic, San Diego CA
18. Edelman A, Arias TA, Smith ST (1998) *SIAM J Matrix Anal Appl* 20:303
19. Andersen HC, Allen MP, Bellemans A, Board J, Clarke JHR, Ferrario M, Haile JM, Nosé S, Opheusden JV, Ryckaert JP (1984) Technical report, Rapport d'activité scientifique du CECAM
20. Nosé S (1984) *J Chem Phys* 81:511
21. Nosé S (1984) *Mol Phys* 52:255
22. Hoover WG (1985) *Phys Rev A* 31:1695
23. Hoover WG (1986) *Phys Rev A* 34:2499
24. Martyna GJ, Klein ML, Tuckerman ME (1992) *J Chem Phys* 97:2635
25. Cho K, Joannopoulos JD (1992) *Phys Rev A* 45:7089
26. Hohenberg P, Kohn W (1964) *Phys Rev B* 136:864
27. Kohn W, Sham LJ (1965) *Phys Rev* 140:1133
28. Roos BO, Widmark PO (eds) (2000) *European summerschool in quantum chemistry*. Lund University
29. Vosko SH, Wilk L, Nusair M (1980) *J Phys* 58:1200
30. Becke AD (1988) *J Chem Phys* 88:2547
31. Filatov M, Thiel W (1998) *Phys Rev A* 57:189
32. Ahlrichs R, Penco R, Scoles G (1977) *Chem Phys* 19:119
33. Andersson Y, Langreth DC, Lundqvist BI (1996) *Phys Rev Lett* 76:102
34. Kamiya M, Tsuneda T, Hirao K (2002) *J Chem Phys* 117:6010
35. Grimme S (2004) *J Comput Chem* 25:1463
36. Wu Q, Yang W (2002) *J Chem Phys* 116:515
37. Zimmerli U, Parrinello M, Koumoutsakos P (2004) *J Chem Phys* 120:2693
38. Seifert G, Kohler T, Tenne R (2002) *J Phys Chem B* 106:2497
39. Johnson ER, Becke AD (2005) *J Chem Phys* 123:024101
40. Lundqvist BI, Andersson Y, Shao H, Chan S, Langreth DC (1995) *Int J Quantum Chem* 56:247
41. Rushton PP, Tozer DJ, Clark SJ (2002) *Phys Rev B* 65:23503
42. von Lilienfeld OA, Tavernelli I, Röthlisberger U (2004) *Phys Rev Lett* 93:153004
43. Szabo A, Ostlund NS (1996) *Modern quantum chemistry*. Dover, New York, USA
44. Jensen F (2002) *Introduction to computational chemistry*. Wiley, Chichester
45. Kirchner B, Huber H, Steinebrunner G, Dreizler H, Grabow J-U, Merke I (1997) *Z Naturforsch* 52a:297
46. Feibelman PJ (1987) *Phys Rev B* 35:2626
47. Jaffe JE, Hess AC (1996) *J Chem Phys* 105:10983
48. Kudin KN, Scuseria GE (2000) *Phys Rev B* 61:16440
49. Saunders VR, Dovesi R, Roetti C, Orlando R, Zicovich-Wilson CM, Harrison NM, Doll K, Civalieri B, Bush I, D'Arco Ph, Llunell M (2003) *CRYSTAL2003 user's manual*. University of Torino
50. Lippert G, Hutter J, Parrinello M (1997) *Mol Phys* 92:477
51. Lippert G, Hutter J, Parrinello M (1999) *Theor Chem Acc* 103:124
52. <http://cp2k.berlios.de/>

53. Fuchs M, Scheffler M (1999) *Comput Phys Commun* 119:67
54. Hamann DR, Schlüter M, Chiang C (1979) *Phys Rev Lett* 43:1494
55. Martin RM (2004) *Electronic Structure*. Cornell University Press, Cambridge
56. Bachelet GB, Hamann DR, Schlüter M (1982) *Phys Rev B* 26:4199
57. Kleinman L, Bylander DM (1982) *Phys Rev Lett* 48:1425
58. Goedecker S, Teter M, Hutter J (1996) *Phys Rev B* 54:1703
59. Hartwigsen C, Goedecker S, Hutter J (1998) *Phys Rev B* 58:3641
60. Payne MC, Joannopoulos JD, Allan DC, Teter MP, Vanderbilt DH (1986) *Phys Rev Lett* 56:2656
61. Pulay P, Fogorasi G (2004) *Phys Rev Lett* 386:272
62. Pastore G, Smargiassi E, Buda F (1991) *Phys Rev A* 44:6334
63. Galli G, Parrinello M (1992) *Phys Rev Lett* 69:3547
64. Tagney P (2006) *J Chem Phys* 124:044111
65. Chester GV (1961) *Adv Phys* 10:357
66. Bloechl PE, Parrinello M (1992) *Phys Rev B* 45:9413
67. Tagney P, Scandolo S (2002) *J Chem Phys* 116:14
68. Warshel A, Levitt M (1976) *J Mol Biol* 103:227
69. Antes I, Thiel W (1999) *J Phys Chem A* 103:9290
70. Zhang YK, Lee TS, Yang WT (1999) *J Chem Phys* 110:46
71. Reuter N, Dejaegere A, Maignet B, Karplus M (2000) *J Phys Chem A* 104:1720
72. Monard G, Loos M, Thery V, Baka K (1996) *Int J Quantum Chem* 58:153
73. Kairys V, Jensen JH (2000) *J Phys Chem A* 104:6656
74. Murphy RB, Philipp DM, Friesner RB (2000) *J Comput Chem* 21:1442
75. Laio A, VandeVondele J, Röthlisberger U (2002) *J Chem Phys* 116:6941
76. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) *J Chem Phys* 110:10452
77. Biswas PK, Gogonea V (2005) *J Chem Phys* 123:164114
78. Woo TK, Cavallo L, Ziegler T (1998) *Theor Chem Acc* 100:307
79. Sagnella DE, Laasonen K, Klein ML (1996) *Biophys J* 71:1172
80. Dal M Peraro, Alber F, Carloni P (2001) *Eur Biophys J* 30:75
81. Ivanov I, Klein ML (2002) *J Am Chem Soc* 124:13380
82. Doerksen RJ, Chen B, Klein ML (2003) *Chem Phys Lett* 380:150
83. Colombo MC, Guidoni L, Laio A, Magistrato A, Maurer P, Piana S, Röhrig U, Spiegel K, Sulpizi M, Vande J Vondele, Zumstein M, Röthlisberger U (2002) *Chimia* 56:13
84. Carloni P, Röthlisberger U, Parrinello M (2002) *Acc Chem Res* 35:455
85. Gageot MP, Sprik M (2003) *J Phys Chem B* 107:10344
86. Buda F, de Groot HJM, Bifone A (1996) *Phys Rev Lett* 77:4474
87. Bifone A, de Groot HJM, Buda F (1997) *Pure and Appl Chem* 69:2105
88. La Penna G, Buda F, Bifone A, de Groot HJM (1998) *Chem Phys Lett* 294:447
89. Bifone A, de Groot HJM, Buda F (1997) *J Phys Chem B* 101:2954
90. Röhrig UF, Guidoni L, Röthlisberger U (2005) *Comput Phys Commun* 6:1836
91. Röhrig UF, Guidoni L, Laio A, Frank I, Röthlisberger U (2004) *J Am Chem Soc* 126:15328
92. Rousseau R, Kleinschmidt V, Schmitt UW, Marx D (2004) *Phys Chem Chem Phys* 6:1848
93. Rousseau R, Kleinschmidt V, Schmitt UW, Marx D (2004) *Angew Chem Int Ed Engl* 43:4804
94. Whitfield TW, Martyna GJ, Allison S, Bates SP, Jason C (2005) *Chem Phys Lett* 414:210
95. Jensen MO, Röthlisberger U, Rovira C (2005) *Biophys J* 89:1744

96. Hugosson HW, Laio A, Maurer P, Röthlisberger U (2006) *J Comput Chem* 27:672
97. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML (2006) *Acc Chem Res* 39:73
98. Huber T, Torda AE, van Gunsteren WF (1994) *J Comput-Aided Mol Design* 8:695
99. Iannuzzi M, Laio A, Parrinello M (2003) *Phys Rev Lett* 90:238302
100. Rovira C, Ballone P, Parrinello M (1997) *Chem Phys Lett* 271:247
101. Rovira C, Parrinello M (1998) *Int J Quantum Chem* 70:387
102. Rovira C, Kunc K, Hutter J, Ballone P, Parrinello M (1998) *Int J Quantum Chem* 69:31
103. Rovira C, Parrinello M (1999) *Chem Eur J* 5:250
104. Rovira C, Parrinello M (2000) *Int J Quantum Chem* 80:1172
105. Rovira C, Schulze B, Eichinger M, Evanseck JD, Parrinello M (2001) *J Biol Chem* 276:81:435
106. Rovira C (2003) *J Phys Condens Matter* 15:1809
107. De Angelis F, Jarzecki AA, Car R, Spiro TG (2005) *J Phys Chem B* 109:3065
108. De Santis L, Carloni P (1999) *Proteins* 37:611
109. Alber F, Carloni P (2000) *Protein Sci* 9:2535
110. Piana S, Carloni P (2000) *Proteins* 39:26
111. Piana S, Sebastiani D, Carloni P, Parrinello M (2001) *J Am Chem Soc* 123:8730
112. Piana S, Carloni P, Parrinello M (2002) *J Mol Biol* 319:567
113. Piana S, Bucher D, Carloni P, Röthlisberger U (2004) *J Phys Chem B* 108:11139
114. Röthlisberger U, Carloni P (1999) *Int J Quantum Chem* 73:209
115. Röthlisberger U, Carloni P, Doclo K, Parrinello M (2000) *J Biol Inorg Chem* 5:236
116. Tilocca A, Gamba A, Vanoni MA, Fois E (2002) *Biochemistry* 41:14111
117. Cavalli A, Carloni P (2002) *J Am Chem Soc* 124:3763
118. Klein CDP, Schiffmann R, Folkers G, Piana S, Röthlisberger U (2003) *J Biol Chem* 278:47862
119. Magistrato A, DeGrado WF, Laio A, Röthlisberger U, Vande Vondele J, Klein ML (2003) *J Phys Chem B* 107:4182
120. Papoian GA, DeGrado WF, Klein ML (2003) *J Am Chem Soc* 125:560
121. Dal Peraro M, Llarrull LL, Röthlisberger U, Vila AJ, Carloni P (2004) *J Am Chem Soc* 126:12661
122. Rauegi S, Cascella M, Carloni P (2004) *J Am Chem Soc* 126:15730
123. Boero M, Tateno M, Tenkura K, Oshiyama A (2005) *J Chem Theory Comput* 1:925
124. Ivanov I, Klein ML (2005) *J Am Chem Soc* 127:4010
125. Kirchner B, Reiher M, Hille A, Hutter J, Hess BA. (2005) *Chem Eur J* 11:574
126. Gossens C, Tavernelli I, Röthlisberger U (2005) *Chimia* 59:81
127. Alagona G, Campanile S, Ghio C, Molin D (2005) *J Molec Struct, THEOCHEM* 729:131
128. Biarnes X, Nieto J, Planas A, Rovira C (2006) *J Biol Chem* 281:1432
129. Vidossich P, Carloni P (2006) *J Phys Chem B* 110:1437
130. Giannozzi P, De Angelis F, Car R (2004) *J Phys Chem* 120:5903
131. Hutter J, Carloni P, Parrinello M (1996) *J Am Chem Soc* 118:8710
132. Carloni P, Sprik M, Andreoni W (2000) *J Phys Chem B* 104:823
133. Frigyes D, Alber F, Pomgor S, Carloni P (2001) *J Molec Struct, THEOCHEM* 574:39
134. Mundy CJ, Colvin ME, Quong AA (2002) *J Phys Chem B* 106:10063
135. Sulpizi M, Folkers G, Röthlisberger U, Carloni P, Scapozza L (2002) *Quant Struct Act Relat* 21:173
136. Sulpizi M, Laio A, Vande Vondele J, Cattaneo A, Röthlisberger U, Carloni P (2003) *Proteins* 52:212
137. Spiegel K, Röthlisberger U, Carloni P (2004) *J Phys Chem B* 108:2699
138. Gervasio FL, Laio A, Iannuzzi M, Parrinello M (2004) *Chem Eur J* 10:4846

139. Wu Y, Mundy CJ, Colvin ME, Car R (2004) *J Phys Chem A* 108:2922
140. Gervasio FL, Laio A, Parrinello M, Boero M (2005) *Phys Rev Lett* 94:158103
141. Spiegel K, Röthlisberger U, Carloni P (2006) *J Phys Chem B* 110:3647
142. Blöchl PE (1994) *Phys Rev B* 50:17953
143. Ionescu AR, Berces A, Zgierski MZ, Whitfield DM, Nukada T (2005) *J Phys Chem A* 109:8096
144. Molteni C, Parrinello M (1998) *J Am Chem Soc* 120:2168
145. Rovira C, Kunc K, Parrinello M (2002) *Inorg Chem* 41:4810
146. Semialjac M, Schröder D, Schwarz H (2003) *Chem Eur J* 9:4396
147. Entel P, Kadau K, Adeagbo WA, Sugihara M, Rollmann G, Zayak AT, Kreth M (2004) *Lecture Notes Phys* 642:177
148. Bühl M (2005) *Inorg Chem* 44:6277
149. Stubbs JM, Marx D (2005) *Chem Eur J* 11:2651
150. Hammond PS, Wu Y, Harris R, Minehardt TJ, Car R, Schmitt JD (2005) *J Comput-Aided Mol Des* 19:1
151. Pophristic V, Veparala S, Ivanov I, Liu Z, Klein ML, De Grado WF (2006) *J Phys Chem B* 110:3517
152. Frank I, Hutter J, Marx D, Parrinello M (1998) *J Chem Phys* 108:4060
153. Molteni C, Frank I, Parrinello M (2001) *Comput Mater Sci* 20:311
154. Bittner ER, Kosov DS (1999) *J Chem Phys* 110:6645
155. Doltsinis NL, Marx D (2002) *Phys Rev Lett* 88:166402
156. Hutter J (2003) *J Chem Phys* 118:3928
157. Odelius M, Laikov D, Hutter J (2003) *J Molec Struct, THEOCHEM*, 630:163
158. Tavernelli I, Röhrig UF, Röthlisberger U (2005) *Mol Phys* 103:963
159. Molteni C, Frank I, Parrinello M (1999) *J Am Chem Soc* 121:12177
160. Sulpizi M, Carloni P, Hutter J, Röthlisberger U (2003) *Phys Chem Chem Phys* 5:4798
161. Sulpizi M, Röhrig UF, Hutter J, Röthlisberger U (2005) *Int J Quantum Chem* 101:671
162. Langer H, Doltsinis NL (2003) *J Chem Phys* 118:5400
163. Langer H, Doltsinis NL, Marx D (2005) *Comput Phys Commun* 6:1734
164. Langer H, Doltsinis NL (2004) *Phys Chem Chem Phys* 6:2742
165. Moret ME, Tapavicza E, Guidoni L, Röhrig UF, Sulpizi M, Tavernelli I, Röthlisberger U (2005) *Chimia* 59:493
166. Reiher M, Kirchner B, Hutter J, Sellmann D, Hess BA (2004) *Chem Eur J* 10:4443
167. Einsle O, Tezcan FA, Andrade SLA, Schmid B, Yoshida M, Howard JB, Rees DC (2002) *Science* 297:1696
168. Leigh GJ (ed) (2002) *Nitrogen fixation at the millenium*. Elsevier, Amsterdam
169. Hinnemann B, Nørskov JK (2004) *J Am Chem Soc* 126:3920
170. Hinnemann B, Nørskov JK (2003) *J Am Chem Soc* 125:1466
171. Dance I (2003) *Chem Commun* 3:324
172. Schimpl J, Petrilli HM, Blöchl PE (2003) *J Am Chem Soc* 124:15772
173. Huniar U, Ahlrichs R, Coucouvanis D (2004) *J Am Chem Soc* 126:2588
174. Yandulov DV, Schrock RR (2003) *Science* 301:76
175. Sellmann D, Utz J, Blum N, Heinemann FW (1999) *Coord Chem Rev* 190–192:607
176. Sellmann D, Fürsattel A, Sutter J (2000) *Coord Chem Rev* 200–202:541
177. Reiher M, Hess BA (2004) *Adv Inorg Chem* 56:55
178. Sellmann D, Hille A, Heinemann FW, Moll M, Rösler A, Sutter J, Brehm G, Reiher M, Hess BA, Schneider S (2003) *Inorg Chim Acta* 348:194
179. Sellmann D, Hille A, Rösler A, Heinemann F, Moll M, Brehm G, Schneider S, Reiher M, Hess BA, Bauer W (2004) *Chem Eur J* 10:819
180. Reiher M, Hess BA (2002) *Chem Eur J* 8:5332

181. Sellmann D, Hautsch B, Rösler A, Heinemann FW (2001) *Angew Chem Int Ed Engl* 40:1505
182. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) *Phys Rev B* 58:7260
183. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD (1992) *Rev Mod Phys* 64:1045
184. Bockstedte M, Kley A, Neugebauer J, Scheffler M (1997) *Comput Phys Commun* 107:187
185. NWChem: developed and distributed by Pacific Northwest National Laboratory, USA
186. Kresse G, Furthmüller J (1996) *Phys Rev B* 54:169

QM/MM Methods for Biological Systems

Hans Martin Senn (✉) · Walter Thiel (✉)

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1,
 45470 Mülheim an der Ruhr, Germany
senn@mpi-muelheim.mpg.de, thiel@mpi-muehlheim.mpg.de

1	Overview and Scope	175
2	The QM/MM Method	178
2.1	Terminology	178
2.2	QM/MM Energy Expressions	179
2.2.1	Subtractive Schemes	179
2.2.2	Additive Schemes	181
2.3	Electrostatic Interaction Between Inner and Outer Subsystems	182
2.3.1	Mechanical Embedding	182
2.3.2	Electrostatic Embedding	183
2.3.3	Polarized Embedding	184
2.3.4	Long-Range Electrostatic QM–MM Interactions	186
2.4	Other Non-bonded and Bonded Interactions Between the Subsystems	187
2.4.1	QM–MM van der Waals Interactions	187
2.4.2	QM–MM Bonded Interactions	188
2.5	Covalent Bonds Across the QM–MM Boundary	189
2.5.1	Overview of Boundary Schemes	189
2.5.2	Link Atoms	191
2.5.3	Boundary Atoms	197
2.5.4	Frozen Localized Orbitals	199
2.5.5	Boundary Schemes: Summary	201
3	Choice of QM and MM Models, QM/MM Implementations	202
3.1	Choice of QM Method	202
3.2	Choice of MM Method	204
3.3	QM/MM Implementations	205
3.3.1	Program Architecture and QM/MM Packages	205
3.3.2	QM/MM-Related Approaches	207
4	Optimization and Simulation Techniques Used with QM/MM	208
4.1	General Comments	208
4.2	Stationary Points and Reaction Paths	209
4.2.1	General Optimization Techniques for Large Systems	210
4.2.2	Optimization Techniques Specific to QM/MM	211
4.2.3	Reaction-Path Techniques	213
4.3	Molecular Dynamics and Simulation Techniques	213
4.3.1	QM/MM Molecular-Dynamics and Monte Carlo Simulations	214
4.3.2	QM/MM Free-Energy Perturbation	215
4.3.3	Thermodynamic Integration	217
4.3.4	Transition-Path Sampling	218

4.3.5	Metadynamics	219
4.3.6	Adiabatic Dynamics	219
4.3.7	QM/MM Reaction-Path Potential	219
5	Practical Aspects of Biomolecular Reaction Modelling	220
6	Interpreting the Results: Understanding Enzyme Catalysis	224
7	Survey of Biomolecular QM/MM Studies	227
	References	250

Abstract Thirty years after the seminal contribution by Warshel and Levitt, we review the state of the art of combined quantum-mechanics/molecular-mechanics (QM/MM) methods, with a focus on biomolecular systems. We provide a detailed overview of the methodology of QM/MM calculations and their use within optimization and simulation schemes. A tabular survey of recent applications, mostly to enzymatic reactions, is given.

Keywords QM/MM · Combined quantum mechanics/molecular mechanics · Optimization · Molecular dynamics · Molecular simulations · Free-energy methods · Enzymatic mechanisms

Abbreviations

ADMP	Atom-centred density-matrix propagation
BFGS	Broyden–Fletcher–Goldfarb–Shanno (Hessian update algorithm in minimizations)
CASSCF	Complete active space self-consistent field
CCSD	Coupled-cluster theory including single and double excitations
COSMO	Conductor-like screening model
CP-MD	Car–Parrinello molecular dynamics
DFT	Density-functional theory
DO	Drude oscillator
DTSS	Differential transition-state stabilization
EC	Enzyme class
ECP	Effective core potential
EFP	Effective fragment potential
EGP	Effective group potential
ELMO	Extremely localized molecular orbital
ESP	Electrostatic potential
EVB	Empirical valence bond
FEP	Free-energy perturbation
FQ	Fluctuating charge
GHO	Generalized hybrid orbital
GSBP	Generalized solvent boundary potential
HDLC	Hybrid delocalized coordinates
HF	Hartree–Fock
IMOMM	Integrated molecular orbital/molecular mechanics
KIE	Kinetic isotope effect

L-BFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm
LBHB	Low-barrier hydrogen bond
LSCF	Local self-consistent field
MC	Monte Carlo
MC-VEEP	Multicentred valence-electron effective potential
MD	Molecular dynamics
MM	Molecular mechanics
MECP	Minimum-energy crossing point
MEP	Minimum-energy path
MP2	Second-order Møller–Plesset perturbation theory
NAC	Near-attack configuration
NEB	Nudged elastic band
OECP	Optimized effective core potentials
ONIOM	Our <i>N</i> -layered integrated molecular orbital/molecular mechanics
PBC	Periodic boundary conditions
PES	Potential-energy surface
PPD	Polarized point dipole
P-RFO	Partial rational-function optimizer
QCP	Quantum capping potentials
QTCP	Quantum-mechanical thermodynamic-cycle perturbation
QM	Quantum mechanics
QM/MM	Combined quantum mechanics/molecular mechanics
SCC-DFTB	Self-consistent-charge density-functional tight-binding
RFO	Rational-function optimizer
SCF	Self-consistent field
SLBO	Strictly localized bond orbital
SMD	Steered molecular dynamics
VEP	Variational electrostatic projection
TDDFT	Time-dependent density-functional theory
TDHF	Time-dependent Hartree–Fock
TI	Thermodynamic integration
TPS	Transition-path sampling
TS	Transition state
US	Umbrella sampling
VTST	Variational transition-state theory
ZPE	Zero-point energy

1

Overview and Scope

Combined quantum-mechanics/molecular-mechanics (QM/MM) approaches have become the method of choice for the modelling of reactions in biomolecular systems. On the one hand, the size and conformational complexity of biopolymers, in particular proteins and nucleic acids, call for highly efficient methods capable of treating up to several 100 000 atoms and allowing for extensive sampling or simulations over time scales of hundreds of nanoseconds. Molecular-mechanics (MM) force fields, based on classical empirical poten-

tials, have been proven to provide an effective means for simulating complex biomolecules (see the contribution by K. Schulten and co-workers in this volume). On the other hand, the description of chemical reactions (i.e., bond-forming and bond-breaking) and other processes that involve changes in the electronic structure, such as charge transfer or electronic excitation, require quantum-mechanical (QM) methods. However, their high computational demands still restrict their applicability to systems of several tens up to a few hundred atoms.

A natural solution to this dilemma is to use a QM method for the chemically active region (e.g., substrates and cofactors in an enzymatic reaction) and combine it with an MM treatment for the surroundings (e.g., the full protein and solvent). The resulting schemes are commonly referred to as combined or hybrid QM/MM methods. They enable the modelling of reactive biomolecular systems at reasonable computational cost while providing the necessary accuracy.

The seminal contribution in the field is due to Warshel and Levitt [1], who presented in 1976, exactly thirty years ago, a method that took into account essentially all aspects of the QM/MM approach and applied it to an enzymatic reaction. Based on an earlier formulation [2], which had been developed to treat conjugated hydrocarbons by the combination of a semi-empirical QM method for the π -electrons with classical MM terms for the σ -framework, their method was characterized by a most remarkable combination of features. The energy expression included the usual MM terms; a semi-empirical QM Hamiltonian that accounted for the polarization of the QM density by the MM point charges as well as by induced dipoles placed on all MM atoms of the protein and by the permanent dipoles of the water molecules; the interaction of the point charges, induced dipoles, and permanent dipoles with each other; and classical bonded and van der Waals QM–MM coupling terms. Within certain approximations, the elaborate description of the electrostatic interactions was treated self-consistently. Covalent bonds across the QM–MM boundary were saturated by single hybrid orbitals placed on the frontier MM atom. Structure optimizations were possible owing to the availability of analytical derivatives with respect to the nuclear positions.

Ten years after this pioneering effort, Singh and Kollman [3] took a major step forward by combining an *ab initio* QM method (Hartree–Fock) with a force field. While a few others had done this before, they were the first to report coupled *ab initio* QM/MM structure optimization. They used link atoms (see Sect. 2.1) to cap the covalent bonds across the QM–MM boundary. Polarization effects were only included as an *a posteriori* correction at fixed geometry; however, they allowed for mutual polarization of the QM and MM regions. The contribution by Field, Bash, and Karplus in 1990 [4] described the coupling of a semi-empirical (AM1 or MNDO) QM method with the CHARMM force field in considerable detail. Their formulation again used

link atoms and accounted for the polarization of the QM density by the MM point charges. The paper carefully evaluated the accuracy and effectiveness of the QM/MM treatment against *ab initio* and experimental data.

Over the last 10 years, numerous reviews [5–37] have documented the development of the QM/MM approach as well as its application to biomolecular systems. The use of the QM/MM method as an explicit-solvent approach to model organic reactions in solution (QM solute in MM solvent, calculation of solvation free energies by a Monte Carlo free-energy perturbation technique) was thoroughly reviewed by Gao [5, 6] in 1996, who considered various methodological issues, in particular the treatment of polarization, in detail. Cunningham and Bash [7] described the development and calibration of semi-empirical QM and MM van der Waals parameters for QM/MM simulations of an enzymatic reaction. Several articles in the *Encyclopedia of computational chemistry* [10–13] and an *ACS symposium series* volume [8] give a comprehensive account of the state of the art in 1998. A complete and succinct overview of the QM/MM method was provided by Sherwood in 2000 [21]. Lin and Truhlar [37] have very recently given an astute report of current methodological aspects.

A number of articles have combined, with varying accents, an overview of QM/MM and other computational methods for biomolecular systems with application surveys from this area [14, 15, 17–20, 22–26, 28–36]. Among these, we highlight the contributions by Field (1999, 2002) [19, 24], Mulholland (2001, 2003) [23, 28], and Friesner (2005) [36].

The current review provides a detailed overview of the QM/MM method and its use within optimization and simulation schemes, and surveys recent applications. Throughout, we keep the focus on biomolecular systems, leaving aside QM/MM treatments geared towards inorganic, organometallic, or solid-state systems, nor covering applications from any of these areas.

We start with an introduction into the formalism of the QM/MM method. The use of different computational models for different regions of space immediately raises issues about how to define the regions and how to treat their mutual interaction, and we will give an overview of different approaches that address these issues. Although the choice of QM and MM methods being combined is in principle arbitrary, we comment on some special aspects and list commonly used combinations, some of which are available in commercial programs. Similarly, we highlight issues pertaining to the use of QM/MM geometry optimization, molecular dynamics (MD), and free-energy simulation techniques. Moreover, practical aspects concerning the setup of QM/MM calculations on biomolecular systems are discussed, as well as different ways of analysing and interpreting the results from such calculations. We conclude with a tabular survey of biomolecular QM/MM studies that have appeared since 2000.

2

The QM/MM Method

2.1

Terminology

A very general sketch of the division of the system into QM and MM parts is shown in Fig. 1. The entire system ($\mathbb{S}$) is partitioned into the inner region ($\mathbb{I}$) to be treated quantum-mechanically and the outer region ($\mathbb{O}$) described by a force field. Inner and outer regions are therefore also frequently referred to as QM and MM regions, respectively. Each atom of the entire system is assigned to either of the subsystems. Because the two regions generally (strongly) interact, it is not possible to write the total energy of the entire system simply as the sum of the energies of the subsystems. As detailed below, coupling terms have to be considered, and it will be necessary to take precautions at the boundary between the subsystems, especially if it cuts through covalent bonds. The term boundary region is used here rather loosely to designate the region where the standard QM and MM procedures are modified or augmented in any way. Depending on the type of QM/MM scheme employed, the boundary region may contain additional atoms (link atoms) used to cap the QM subsystem that are not part of the entire system, or it may consist of atoms with special features that appear both in the QM and the MM calculation. Note that the assignment of each atom to either subsystem is no longer unique in this latter case.

Anticipating the discussion of boundary schemes (Sect. 2.5), we introduce here some labelling conventions, illustrated in Fig. 2, that apply to covalent bonds across the QM–MM boundary. The QM and MM atoms directly connected are designated Q^1 and M^1 , respectively, and are sometimes referred to as boundary, frontier, or junction atoms. The first shell of MM atoms (i.e., those directly bonded to M^1) is labelled M^2 . The next shell, separated from M^1 by two bonds, is labelled M^3 ; and so on, following the molecular graph outwards from M^1 . The same naming procedure applies to the QM side; atoms Q^2

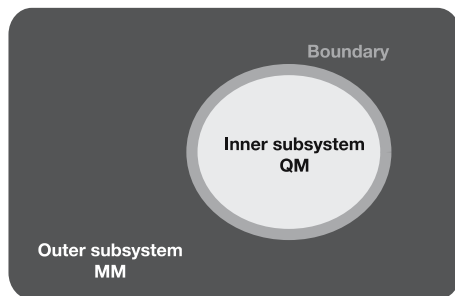


Fig. 1 Partitioning of the entire system $\mathbb{S}$ into inner ($\mathbb{I}$) and outer ($\mathbb{O}$) subsystems

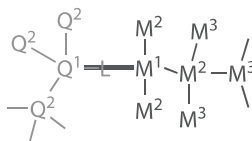


Fig. 2 Labelling of atoms at the boundary between QM and MM regions

are one bond away from Q¹, Q³ are two bonds away, etc. If a link-atom scheme is applied, the dangling bond of Q¹ is saturated by the link atom L.

As a caveat, we emphasize that the classification of QM/MM schemes and the definition of terms such as link, capping, boundary, junction, or frontier atom are not unique and their usage varies between authors. Moreover, a given QM/MM method can incorporate aspects from different schemes, making its classification ambiguous.

If not stated otherwise, for the remainder of this review the classical potential-energy function (the “force field”) of the MM region is assumed to include bonded terms (bond stretching, angle bending, torsions, out-of-plane deformations or improper torsions), a Lennard–Jones-type van der Waals term, and the Coulomb interaction between rigid point charges. A simple, prototypical MM energy expression of this type, sometimes called a “class I” force field, reads:

$$E_{\text{MM}} = \sum_{\text{bonds}} k_b (d - d_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi + \delta)] \\ + \sum_{\text{non-bonded pairs AB}} \left\{ \varepsilon_{\text{AB}} \left[\left(\frac{\sigma_{\text{AB}}}{r_{\text{AB}}} \right)^{12} - \left(\frac{\sigma_{\text{AB}}}{r_{\text{AB}}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_A q_B}{r_{\text{AB}}} \right\}, \quad (1)$$

where d , θ , and ϕ designate bond distances, angles, and torsions, respectively; d_0 and θ_0 are the corresponding equilibrium values; and n and δ are the torsional multiplicity and phase, respectively. The bonded force constants are k_b , k_θ , and k_ϕ . r_{AB} is the non-bonded distance, and ε_{AB} and σ_{AB} are the van der Waals parameters between atoms A and B. q_A , q_B are atomic partial charges and ε_0 is the vacuum permittivity. We refer to the literature [38–42] for pertinent details and extensions or variations of this general form.

2.2

QM/MM Energy Expressions

2.2.1

Subtractive Schemes

In subtractive QM/MM schemes, three calculations are performed: (i) an MM calculation on the entire system, $\mathbb{S}$; (ii) a QM calculation on the inner sub-

system, $\mathbb{I}$; and (iii) an MM calculation on the inner subsystem. The QM/MM energy of the entire system is then obtained by summing (i) and (ii) and subtracting (iii) to correct for double counting:

$$E_{\text{QM/MM}}(\mathbb{S}) = E_{\text{MM}}(\mathbb{S}) + E_{\text{QM}}(\mathbb{I} + \mathbb{L}) - E_{\text{MM}}(\mathbb{I} + \mathbb{L}). \quad (2)$$

Here, as in the following, the subscript indicates the level of the calculation while the system on which it is performed is given in parentheses. As written, Eq. 2 holds for a link-atom scheme, the calculations on the inner subsystem being performed not on the bare $\mathbb{I}$ but on $\mathbb{I}$ capped with link atoms, $\mathbb{I} + \mathbb{L}$. For a scheme with special MM boundary atoms (rather than link atoms) that carry certain features appearing also in the calculations on the inner subsystem, $\mathbb{L}$ is understood to refer to these atoms. If no covalent bond is cut by the QM–MM boundary, $\mathbb{I} + \mathbb{L}$ reduces simply to $\mathbb{I}$.

Conceptually, the subtractive QM/MM scheme can be seen as an MM approach in which a certain region has been cut out and replaced by a higher-level treatment. Its main advantage is simplicity. No explicit coupling terms are needed, avoiding any modification of the standard QM and MM procedures, and the subtraction implicitly corrects for any artefacts caused by the link atoms, provided that the MM force terms referring to the link atoms reproduce the QM potential reasonably well. These features make a subtractive scheme fairly straightforward to implement.

On the downside, a subtractive scheme also requires a complete set of MM parameters for the inner subsystem. These may be difficult or cumbersome to obtain. Moreover, and more severely, the coupling between the subsystems is treated entirely at the MM level. This is particularly problematic for the electrostatic interaction, which is typically described by fixed atomic charges at the MM level. Hence, in a subtractive scheme the electrostatic interaction between the subsystems is treated within a simple point-charge model, which is often a rather severe approximation: First, the charge distribution in the inner subsystem can change (e.g., during a reaction), which cannot be reflected by rigid point charges. Second, the QM calculation does not incorporate the charges in the outer region, that is, the QM charge density is not polarized by the environment. A subtractive scheme is therefore not suitable if the electron density is significantly influenced by electrostatic interactions with the outer region.

Within the classification of QM–MM coupling schemes (Sect. 2.3), a strictly subtractive QM–MM method necessarily implies mechanical embedding (i.e., the QM density is not polarized by the environment). However, mixed formulations are conceivable that are in principle subtractive, but treat the electrostatic interaction separately, allowing for a more elaborate coupling scheme.

As an example for a subtractive QM/MM scheme, we mention the IMOMM method (integrated molecular orbital/molecular mechanics) by Morokuma and co-workers [43]. It has subsequently been extended to enable the combi-

nation of two QM methods (IMOMO [44]) and further generalized to N layers (typically, $N = 3$), each of which can be treated at an arbitrary QM or MM level (ONIOM, our N -layered integrated molecular orbital and molecular mechanics [45–47]). Recent improvements of the ONIOM approach [48–50] that enable the inclusion of MM charges into the QM Hamiltonian (electrostatic embedding, see Sect. 2.3.2) take it beyond a purely subtractive scheme.

2.2.2

Additive Schemes

The basic energy expression for an additive QM/MM scheme is:

$$E_{\text{QM/MM}}(\textcircled{\text{S}}) = E_{\text{MM}}(\textcircled{\text{O}}) + E_{\text{QM}}(\textcircled{\text{I}} + \textcircled{\text{L}}) + E_{\text{QM-MM}}(\textcircled{\text{I}}, \textcircled{\text{O}}). \quad (3)$$

In contrast to the subtractive scheme of Eq. 2, the MM calculation is now performed on the outer subsystem only. In addition, there appears an explicit coupling term, $E_{\text{QM-MM}}(\textcircled{\text{I}}, \textcircled{\text{O}})$, which collects the interaction terms between the two subsystems. The capped inner subsystem, $\textcircled{\text{I}} + \textcircled{\text{L}}$, is treated at the QM level as before.

Assuming a link-atom-based scheme with mechanical embedding, it is possible to derive the additive energy expression from the subtractive one [51]. Using the fact that the MM energy is unambiguously decomposable into contributions depending on exclusive sets of atoms, we can split the MM terms of Eq. 2 as:

$$E_{\text{MM}}(\textcircled{\text{S}}) = E_{\text{MM}}(\textcircled{\text{O}}) + E_{\text{MM}}(\textcircled{\text{I}}) + E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{O}}), \quad (4)$$

$$E_{\text{MM}}(\textcircled{\text{I}} + \textcircled{\text{L}}) = E_{\text{MM}}(\textcircled{\text{I}}) + E_{\text{MM}}(\textcircled{\text{L}}) + E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{L}}). \quad (5)$$

Substituting these into Eq. 2, the MM contribution from the inner subsystem, $E_{\text{MM}}(\textcircled{\text{I}})$, cancels, and we obtain the full QM/MM energy as:

$$E_{\text{QM/MM}}(\textcircled{\text{S}}) = E_{\text{MM}}(\textcircled{\text{O}}) + E_{\text{QM}}(\textcircled{\text{I}} + \textcircled{\text{L}}) + E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{O}}) - [E_{\text{MM}}(\textcircled{\text{L}}) + E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{L}})]. \quad (6)$$

For a mechanical-embedding scheme with link atoms, $E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{O}})$ can be identified with the QM–MM coupling term $E_{\text{QM-MM}}(\textcircled{\text{I}}, \textcircled{\text{O}})$ of Eq. 3 as it contains for this case all the interactions between the subsystems.

The subtractive terms in Eq. 6 are referred to as the “link-atom correction”:

$$E_{\text{link}} = - [E_{\text{MM}}(\textcircled{\text{L}}) + E_{\text{MM}}(\textcircled{\text{I}}, \textcircled{\text{L}})]. \quad (7)$$

While the link atoms are not part of the entire (i.e., physical or real) system being modelled, their interaction with one another and the atoms of the inner subsystem is contained in the term $E_{\text{QM}}(\textcircled{\text{I}} + \textcircled{\text{L}})$. A correction is thus formally justified. However, E_{link} is in practice often omitted, which can be motivated by pragmatic arguments: (i) The accuracy and validity of a correction at the

MM level for QM interactions is questionable. It is therefore unclear if the correction would actually improve the overall model. (ii) The interaction among the link atoms, $E_{\text{MM}}(\mathbb{L})$ (which for a typical force field consists of electrostatic and van der Waals contributions), is expected to be small. (iii) $E_{\text{MM}}(\mathbb{I}, \mathbb{L})$, the interaction between the link atoms and the inner subsystem, is not small. However, in many common link-atom schemes, this term is a constant or depends only weakly on the structure because the position of the link atom, in particular the distance $Q^1\text{--L}$, is constrained.

The working equation adopted in the majority of QM/MM schemes is thus Eq. 3. The exact form of the QM–MM coupling term $E_{\text{QM-MM}}$ defines a particular QM/MM method. In accordance with the interactions considered in the force field, it includes electrostatic, van der Waals, and bonded interactions between QM and MM atoms:

$$E_{\text{QM-MM}}(\mathbb{I}, \mathbb{O}) = E_{\text{QM-MM}}^{\text{el}} + E_{\text{QM-MM}}^{\text{vdW}} + E_{\text{QM-MM}}^{\text{b}}. \quad (8)$$

The following sections deal in more detail with the individual contributions to $E_{\text{QM-MM}}$. The electrostatic coupling term (Sect. 2.3.2) arguably has the largest impact and is also the most technically involved one. The van der Waals interaction and the bonded terms are discussed in Sect. 2.4. Section 2.5 presents various ways that have been devised to treat covalent bonds across the QM–MM boundary.

2.3

Electrostatic Interaction Between Inner and Outer Subsystems

The electrostatic coupling between the QM charge density and the charge model used in the MM region can be handled at different levels of sophistication, characterized essentially by the extent of mutual polarization and classified [51, 52] accordingly as mechanical embedding (model A), electrostatic embedding (model B), and polarized embedding (models C and D).

2.3.1

Mechanical Embedding

In a mechanical-embedding scheme, the QM–MM electrostatic interaction is treated on the same footing as the MM–MM electrostatics, that is, at the MM level. The charge model of the MM method used – typically rigid atomic point charges, but other approaches, e.g., bond dipoles, are also possible – is simply applied to the QM region as well. This is conceptually straightforward and computationally efficient.

However, there are major disadvantages and limitations: (i) The charges in the outer region do not interact with the QM density, which is thus not directly influenced by the electrostatic environment. Hence, the QM density is

not polarized. (ii) As the charge distribution in the QM region changes, for instance during a reaction, the charge model needs to be updated. However, this is problematic because it leads to discontinuities in the potential-energy surface. (iii) The derivation of, e.g., MM point charges for the inner region is often not trivial. The procedures vary widely between force fields and can require considerable effort. Moreover, they may not be general but geared towards the class of compounds for which the force field was developed. In this case, their applicability to the inner region, which is often treated at the QM level exactly because it is outside the chemical domain of the force field, is questionable. (iv) The MM charge model is dependent on, and inter-linked with, the other force-field parameters. Together with these, it is mainly intended to yield a balanced description of conformational or structural preferences, rather than to reproduce accurately the true charge distribution. It is therefore not justifiable to use charges for the QM part derived from a model different to the one applied in the force field.

2.3.2

Electrostatic Embedding

The major shortcomings of mechanical embedding can be eliminated by performing the QM calculation in presence of the MM charge model. For instance, by incorporating the MM point charges as one-electron terms in the QM Hamiltonian, which is thus augmented by an additional term (using atomic units):

$$\hat{H}_{\text{QM-MM}}^{\text{el}} = - \sum_i^{\text{electrons}} \sum_{M \in \mathbb{O}} \frac{q_M}{|\mathbf{r}_i - \mathbf{R}_M|} + \sum_{\alpha \in \mathbb{I} + \mathbb{L}} \sum_{M \in \mathbb{O}} \frac{q_M Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{R}_M|}, \quad (9)$$

where q_M are the MM point charges and Z_α the nuclear charge of the QM atoms; the index i runs over all electrons, M over the point charges, and α over the QM nuclei.

In such a scheme (referred to as electrostatic or electronic embedding) the electronic structure of the inner region can adapt to changes in the charge distribution of the environment and is automatically polarized by it. No charge model needs to be derived for the inner region. The QM–MM electrostatic interaction is treated at the QM level, which obviously provides a more advanced and more accurate description than a mechanical-embedding scheme. Naturally, electrostatic embedding also increases the computational cost, especially for the calculation of the electrostatic force due to the QM density acting on the (many) MM point charges.

Special care is required at the QM–MM boundary, where the MM charges are placed in immediate proximity to the QM electron density, which can lead to overpolarization. This problem is especially pronounced when the boundary runs through a covalent bond, and is therefore discussed later in Sect. 2.5.

Note that because the QM-MM electrostatic interaction term, $E_{\text{QM-MM}}^{\text{el}}$, is now calculated by the QM code, it is sometimes considered a contribution to E_{QM} and included therein. However, in the present review we will strictly adhere to the energy partitioning given by Eqs. 3 and 8, that is, E_{QM} is the pure QM energy, while $E_{\text{QM-MM}}^{\text{el}}$ is part of $E_{\text{QM-MM}}$.

There remains the issue that the MM charge model is not necessarily well-suited to interact with the QM electron density. As mentioned above, the electrostatic MM parameters are not primarily designed to provide a faithful representation of the real charge distribution. It is, in principle, not legitimate to stitch a true charge distribution, as provided by the QM calculation, into the carefully parameterized MM charge model. Nevertheless, this has become common practice, and experience shows that it generally yields reasonable results, at least for the combination of a QM density with one of the widely used biomolecular force fields. The obvious appeal of this approach is that the MM atomic partial charges are readily available from the force field and their inclusion in the QM Hamiltonian is efficient. Electrostatic embedding is the most popular embedding scheme in use today, certainly for biomolecular applications.

2.3.3

Polarized Embedding

As electrostatic embedding accounts for the interaction of the polarizable QM density with rigid MM charges, the next logical step is to introduce a flexible MM charge model that is polarized by the QM charge distribution. One can further divide these polarized-embedding schemes into approaches that apply a polarizable-charge model in the MM region, which is polarized by the QM electric field but does not itself act on the QM density (model C), and fully self-consistent formulations that include the dipoles into the QM Hamiltonian and therefore allow for mutual polarization (model D).

There exist various models used to treat polarization in the MM part, which can broadly be classified as follows. We provide here only a very brief overview and refer to [40, 53, 54] for more detailed treatments and pertaining references:

- *Polarized point dipoles (PPD)*: Polarizabilities are assigned to atoms (or other distinguished sites, e.g., centres of mass), which interact with the electric field at that site, thus inducing point dipoles. The sources of the electric field are the point charges, the other induced dipoles, and possibly the QM charge distribution. Since the dipoles interact with each other, an iterative procedure must be applied to generate a self-consistent polarization. Alternatives are a full-matrix direct solution or extended Lagrangean schemes with the dipoles as fictitious degrees of freedom. The free parameters of the model are the (atomic) polarizabilities. Sometimes,

a dipole–dipole interaction model is applied that damps the interaction between close-lying dipoles. In principle, PPD methods can be extended to include higher-order multipoles.

- *Drude oscillators (DO)*: A mobile point charge of opposite sign is connected to a charge site by a harmonic spring, thus forming a dipole. These dipoles then interact with the local electric field, as outlined above. DO models are usually implemented within an extended Lagrangean scheme. The fit parameters of the model are the magnitude of the mobile charge and the spring constant. In the context of solid-state simulations, the DO approach is often referred to as shell model; it is also known as charge-on-spring model.
- *Fluctuating charges (FQ)*: Based on the principle of electronegativity equalization, the atomic partial charges are optimized with respect to the total electrostatic energy. The practical advantage of the FQ as well as the DO approaches is that the description of polarizability is achieved without explicitly introducing additional (i.e., dipole–dipole) interactions. In FQ methods, even the number of charge–charge interactions is unchanged from the non-polarizable case.

Although the very first QM/MM approaches were in fact polarized-embedding schemes [1, 3], they have remained scarce. The main obstacle is the lack of well-established polarizable biomolecular MM force fields. A variety of polarizable solvent models is available, most prominently for the simulation of liquid water (recent examples include [55–61]). The development of polarizable protein force fields, however, is very much a work in progress. We mention contributions from the developers of the CHARMM [62–65] and AMBER [66–68] force fields, Friesner and co-workers [69–72], and Ren and Ponder [73], and refer to [40] for a recent review.

Apart from the availability of polarizable force fields, there are also some computational and technical issues that need to be considered in polarized-embedding schemes. Model D requires the coupling of the self-consistency cycles for the QM charge density and the MM polarizable-charge model, which increases the computational effort and may create convergence problems. Additional complications also arise at the boundary between the subsystems, where the QM density and the MM charge model interact in close proximity, see Sect. 2.5.2.

As far as the use of polarized-embedding schemes in QM/MM calculations is concerned, there is only limited experience. A PPD model D at the semi-empirical QM level was applied to a fairly large biomolecular system [74], and tests on small organic molecules with a PPD model C at the semi-empirical, Hartree–Fock, and DFT QM levels have been reported [51, 52]. Otherwise, polarized-embedding QM/MM calculations were restricted to explicit-solvation (in particular, hydration) studies, where the solute is treated at the QM level and the solvent by a polarizable force field [18, 75–77].

2.3.4

Long-Range Electrostatic QM–MM Interactions

An accurate description of the electrostatic forces on the QM subsystem due to the environment is essential for a reliable modelling of the structure and function of biomolecules. Including all the electrostatic interactions explicitly is computationally challenging, and QM/MM electrostatic cutoffs are problematic because of the long-range nature of the Coulomb interaction. Several recent studies have shown that cutoffs can introduce significant artefacts [78–80]. While the reliable and efficient treatment of the electrostatic interactions is a well-established topic in the area of classical MD simulations, it has only recently found increased attention in the context of QM/MM methods; we highlight here some recent developments:

- *Ewald methods*: For simulations done under periodic boundary conditions (PBC), Ewald methods provide an accurate treatment of long-range electrostatics. A linear-scaling particle-mesh Ewald scheme for QM/MM simulations has recently been presented by York and co-workers [78]. Although accurate, the PBC/Ewald approach generally suffers from high computational demands because of the large number of explicit solvent molecules that need to be included. The biomolecule of interest is immersed in a box of explicit solvent, whose size must be chosen large enough to minimize artefacts caused by the artificially imposed periodicity. The large number of degrees of freedom further increases the simulation cost because it prolongs the required equilibration times. Different approaches have therefore been proposed that include the electrostatic interactions explicitly only from an active region around the QM part.
- *Charge scaling* [81]: Karplus and co-workers have proposed a procedure for QM/MM free-energy simulations where only a limited number of explicit solvent molecules is considered and the charges are scaled to mimic the shielding effect of the solvent. The energies obtained are then corrected using continuum-electrostatics (linearized Poisson–Boltzmann or finite-difference Poisson) calculations.
- *Variational electrostatic projection (VEP)* [82, 83]: In the popular stochastic-boundary method [84–87] for MD simulations, the spherical “active zone” (treated by standard Hamiltonian dynamics) is surrounded by a “stochastic buffer” shell governed by Langevin dynamics; the solute atoms in this buffer are positionally restrained, and the solvent molecules are subject to a boundary potential. The remaining parts of the system are held fixed and form the external environment.

Proposed by Gregersen and York, the VEP method aims at reducing the cost of calculating the electrostatic forces due to the external environment on the atoms in the two inner regions. It uses Gaussians to expand the charge distribution of the environment on a discretized spher-

ical surface enclosing the moving part of the system. The procedure is related to the continuum-solvent models of the COSMO (conductor-like screening model) type. An improved variant of the VEP method is the VEP-RVM (reverse variational mapping) method. A charge-scaling implementation of the VEP and VEP-RVM approaches has also been presented [83].

- *Generalized solvent boundary potential (GSBP)* [80]: The spherical solvent boundary potential (SSBP) [88,89] includes a small number of solvent molecules explicitly while the surrounding ones are represented by a (spherical) effective boundary potential. The GSBP method generalized this scheme to boundaries of arbitrary shape [90]. All atoms in the inner region are treated by explicit dynamics, while the fixed environment is included in terms of a solvent-shielded static field and a Poisson–Boltzmann reaction field.

The GSBP approach has recently been extended and adapted for QM/MM simulations by Cui and co-workers [80]. They successfully validated and applied it in pK_a calculations [79], proton-transfer processes [91], and other biomolecular simulations [92]. They stress in particular the need to treat the QM–MM and MM–MM electrostatics in a balanced manner to prevent artefacts.

2.4

Other Non-bonded and Bonded Interactions Between the Subsystems

In addition to the electrostatic interaction discussed in the previous section, there are also van der Waals and bonded contributions to the QM–MM coupling term, Eq. 8. Their treatment is considerably simpler than for the electrostatic coupling as they are handled purely at the MM level, irrespective of the class (subtractive or additive) of QM/MM scheme.

2.4.1

QM–MM van der Waals Interactions

The van der Waals interaction is typically described by a Lennard–Jones potential, as shown in Eq. 1; alternative functional forms, e.g., with an exponential repulsive term, are sometimes used instead. However, the present discussion is unaffected by the exact form of this term. The first issue that arises here is the same as discussed above in the context of mechanical embedding: the availability and suitability of MM parameters for the inner region. It is not uncommon that certain QM atoms are not covered by any of the atom types and assignment rules of the force field. Secondly, even if suitable van der Waals parameters exist for a given configuration, QM atoms can change their character, e.g., during a reaction. This then raises the question of whether one should switch the parameter set, say, from a “re-

actant description” to a “product description”; and if so, at which point along the reaction path. Finally, there is the overarching problem that, strictly speaking, MM parameters are not separable and transferable, but only valid within the parameterization for which they were derived, that is, for MM–MM interactions.

In practice, however, all these complications are very much alleviated by the short-range nature of the van der Waals interaction. While every atom of the inner region is involved in van der Waals interactions with all the atoms of the outer region, only the closest-lying ones contribute significantly. Unoptimized van der Waals parameters therefore affect only the QM atoms close to MM atoms, that is, those at the boundary. If one is concerned that this might influence the result, one solution is to move the QM–MM boundary further away from the incriminated QM atom. Similar considerations apply to the ambiguity of choosing a fixed set of van der Waals parameters, where switching between parameter sets would introduce additional problems rather than increase the quality of the model. The effect can simply be checked by comparing the results obtained with different parameters (e.g., derived for the reactants and the products).

Friesner and co-workers [93] in their QM/MM scheme have re-optimized the QM van der Waals parameters against structures and bonding energies of hydrogen-bonded pairs of small models for amino acids. The van der Waals radii thus obtained are 5–10% larger than those of the force field used (OPLS-AA); the van der Waals well depths were left unchanged. The increased van der Waals repulsion compensates for the too-strong QM–MM electrostatic attraction caused by the MM point charges overpolarizing the QM density. Recently, a set of van der Waals parameters optimized for B3LYP/AMBER was presented by a different group [94]. Cui and co-workers [95] showed that thermodynamic quantities in the condensed phase (e.g., free energies), calculated from QM/MM simulations, are rather insensitive towards the QM–MM van der Waals parameters. As expected, they do, however, influence the detailed structure around the QM region.

With respect to the QM–MM van der Waals coupling, subtractive and additive schemes are identical. In an additive scheme, the simple rule is that only pairs consisting of one atom from the inner and one atom from the outer subsystem are considered in $E_{\text{QM-MM}}^{\text{vdW}}$. This yields exactly the same van der Waals terms as a subtractive scheme, where the QM–QM van der Waals pairs are subtracted out.

2.4.2

QM–MM Bonded Interactions

The formal reservations raised above against using MM parameters to describe QM–MM interactions also apply of course in the case of the bonded (bond stretching, angle bending, torsional, etc.) interactions. And again, the

solution is entirely pragmatic. One usually retains the standard MM parameter set and complements it as necessary with additional bonded terms not covered by the default assignment rules of the force field. As the bonded interactions are by definition strictly localized to the boundary, one can validate the results by extending the inner region, shifting the boundary and, hence, potentially less reliable interaction terms away from the chemically active region.

For the bonded QM–MM interaction, there is an operational difference between subtractive and additive schemes with respect to the treatment of link atoms, which leads to different terms being included in the final energy expression. A subtractive scheme removes by construction the QM–QM bonded interactions (i.e., those involving atoms from the capped inner region $\mathbb{I} + \mathbb{L}$ only) and retains all mixed QM–MM bonded terms. It thus implicitly corrects for the link atoms. For instance (see Fig. 2), the stretching terms Q^2-Q^1 and Q^1-L and the bending term Q^2-Q^1-L are removed, while Q^1-M^1 and $Q^2-Q^1-M^1$ are included.

By contrast, an additive scheme requires an explicit set of rules that govern which bonded contributions are to be included in $E_{\text{QM-MM}}^b$, thereby avoiding double-counting of (possibly implicit) interactions. The general rule is that every bonded term that depends on atoms from both the inner and the outer subsystem is included (note that the link atoms do *not* belong to either region). However, terms like $Q^2-Q^1-M^1$ or $Q^3-Q^2-Q^1-M^1$ are excluded to prevent double-counting. For example, when the angle $Q^2-Q^1-M^1$ is distorted, the link atom placed along Q^1-M^1 needs to move as well. This leads to restoring forces on Q^1 and M^1 , as discussed in Sect. 2.5.2. Hence, the angular distortion is implicitly accounted for, and the bending term $Q^2-Q^1-M^1$ is omitted. Commonly, only angle terms of the form $M^1-Q^1-M^1$ and torsion terms where at least one of the two central atoms is QM are retained [96]. However, the exact rules by which bonded interactions between QM and MM atoms are included depend on the details of the boundary scheme employed.

2.5

Covalent Bonds Across the QM–MM Boundary

2.5.1

Overview of Boundary Schemes

This section is concerned with the various approaches that have been devised to treat covalent bonds cut by the QM–MM boundary. The simplest solution is of course to circumvent the problem by defining the subsystems such that the boundary does not pass through a covalent bond. This is trivially fulfilled for explicit-solvation studies, where the solute and maybe the first solvation shell are described at the QM level, surrounded by MM solvent molecules. It is sometimes possible also for biomolecular systems; for

instance, if the reactants of an enzymatic reaction (substrates, cofactors) are not covalently bound to the enzyme and no protein residue is directly involved in the chemical transformation. In many cases, however, it is unavoidable that the QM–MM boundary cuts through a covalent bond. Such situations arise when one needs to include certain protein residues in the inner region or would like to treat chemically inactive parts of the substrate or cofactor at the MM level to reduce the computational cost. One then has to deal mainly with three issues (see Sect. 2.1 for atom labelling conventions): (i) The dangling bond of the QM atom Q^1 must be capped; simply assuming a truncated QM region (i.e., treating the bond as being homolytically or heterolytically cleaved) would be entirely unrealistic. (ii) For electrostatic or polarized embedding, one has to prevent overpolarization of the QM density, in particular, by the partial charge on M^1 . This is problematic especially when link atoms are used. (iii) The bonded MM terms involving atoms from both subsystems have to be selected such that double-counting of interactions is avoided (see Sect. 2.4.2). Overall, the boundary scheme should provide a balanced description of the QM–MM interaction at the border between the two subsystems.

The different boundary schemes can be categorized into three groups, examined in more detail in the following sections:

- Link-atom schemes introduce an additional atomic centre L (usually a hydrogen atom) that is not part of the entire, real system. It is covalently bound to Q^1 and saturates its free valency.
- In boundary-atom schemes, the MM atom M^1 is replaced by a special “Janus” boundary atom that appears in both the QM and the MM calculation. On the QM side, it mimics the cut bond and possibly also the electronic character of the MM moiety attached to Q^1 . In the MM calculation, it behaves as a normal MM atom.
- Localized-orbital schemes place hybrid orbitals at the boundary and keep some of them frozen. They serve to cap the QM region, replacing the cut bond.

Where to Cut

While cutting through covalent bonds can often not be avoided, as discussed above, one can minimize its ramifications by an appropriate choice of the boundary, that is, of the bonds being cut. Apart from the obvious rule that the QM–MM frontier should be as distant from the chemically active region as the size of the QM part (i.e., the computational demand) allows, one can give some additional guidelines. A minimum requirement is that QM atoms participating in bond making or breaking should not be involved in any bonded coupling term [50]. Since the dihedral terms extend at most two bonds into the inner region (depending on the details of the boundary scheme in use,

Sect. 2.4.2), one is on the safe side if such atoms are at least three bonds away from the boundary.

The bond being cut should be non-polar and not involved in any conjugative interaction (multiple bonding, hyperconjugation, stereoelectronic interaction). A good place to cut is thus an aliphatic, “innocent” C–C bond, whereas cutting through an amide bond, which has partial double-bond character, is less desirable.

Another restriction is introduced by MM charge groups. It is common practice in biomolecular force fields to collect several connected atoms into a group with integer, normally zero, charge. This is advantageous since the electrostatic interaction between neutral charge groups can be neglected to first order, thus enabling the construction of a molecule from these neutral units without reparameterization of the partial charges. Cutting through a charge group is to be avoided because it creates an artificial net charge in the immediate vicinity of the QM density. It may also interfere with certain algorithms that calculate the MM electrostatic interactions based on charge groups. Finally, it is desirable, but not compulsory, that the total charge of the MM atoms being replaced by the QM part is zero. In other words, the hole created in the MM region that surrounds the QM part (and therefore the QM part itself) should be neutral, such that the leading electrostatic interaction between QM and MM subsystems is the dipole contribution.

2.5.2

Link Atoms

The appeal of the link-atom method, adopted already by early QM/MM studies [3, 4], lies in its conceptual simplicity: the free valency at Q^1 created by the QM–MM separation is capped by introducing an additional atom that is covalently bonded to Q^1 . This link atom L is in most cases a hydrogen atom, but any monovalent atom or group is in principle conceivable. One thus performs the QM calculations on an electronically saturated system consisting of the inner subsystem and the link atom(s), $\mathbb{I} + \mathbb{L}$. The bond Q^1-M^1 is described at the MM level.

Although simple, the introduction of an additional atomic centre, which is not part of the real system, entails consequences that need to be addressed:

- Each link atom introduces three artificial (structural) degrees of freedom not present in the real system. This causes complications during structure optimizations and raises the question of how the position of the link atom is to be determined (discussed in more detail below).
- The link atom, and with it the QM electron density, is spatially very close to the MM frontier atom M^1 . If M^1 bears a partial charge and the QM density is allowed to be polarized (i.e., electrostatic or polarized embedding), the point charge on M^1 will overpolarize the density. Different

approaches to alleviate or eliminate this unphysical effect are presented below.

- The link atom introduces artificial interactions with other link atoms and the inner region. This has already been treated in Sect. 2.2.

Another issue of the link-atom method is that the link atom is generally chemically and electronically different from the group it replaces. Attempts to overcome this have led to the more elaborate boundary schemes discussed in the following sections below. In the context of link atoms, Morokuma and co-workers suggested shifting the energy level of a selected orbital by means of an additional one-electron operator in the Hamiltonian [97]. They proposed mimicking the electronic effect of a substituent by a hydrogen link atom with appropriately shifted energy levels [98]. A similar idea was explored using an angular-momentum-dependent localized potential within the projector-augmented waves method [99].

Despite their shortcomings, link atoms are the most popular and most widely used boundary method. Correspondingly, a large variety of link-atom schemes has evolved. One of the first is due to Singh and Kollman [3], followed by Field, Bash, and Karplus [4]. Modified or extended formulations were reported by several groups [96, 100–104]. Another line of developments derives from Morokuma's IMOMM method [43, 46, 47, 105, 106]. A double-link-atom method has also been proposed [107], in which a second link atom is introduced to also saturate the MM region.

2.5.2.1

Placement of the Link Atom

In some schemes [3, 4, 102], the link atoms are treated as independent atomic centres, thus introducing three additional structural degrees of freedom per link atom. The link atoms are initially positioned at a certain distance along the Q^1-M^1 bond vector, but are completely free during structure optimization. To mitigate the inconsistencies that arise when the artificial bond Q^1-L is no longer collinear with the real Q^1-M^1 bond, it was suggested [96] that a classical angle term $L-Q^1-M^1$ be introduced, with an equilibrium value of 0° that keeps the bonds aligned.

To remove the excess degrees of freedom altogether, one can eliminate them by the use of constraints. This was first realized by Maseras and Morokuma [43], who constrained the MM frontier atom M^1 to lie along Q^1-L and fixed both the Q^1-L and the Q^1-M^1 distances using a formulation in internal coordinates. As this eliminates four, rather than three, degrees of freedom per link atom, the procedure was subsequently modified [47] by defining the position of the link atom as a function of the positions of Q^1 and M^1 in Cartesian coordinates:

$$R_L(R_{Q^1}, R_{M^1}) = R_{Q^1} + g(R_{M^1} - R_{Q^1}) . \quad (10)$$

This definition places L along Q^1-M^1 , and the distance Q^1-L is related to the distance Q^1-M^1 by the scaling factor g . Exactly three degrees of freedom are thus removed. Most current link-atom schemes are based on Eq. 10, using different definitions for g .

Woo et al. [105] eliminate the coordinates of M^1 , rather than L, by making R_{M^1} a function of R_{Q^1} and R_L :

$$R_{M^1}(R_{Q^1}, R_L) = R_{Q^1} + g'(R_L - R_{Q^1}). \quad (11)$$

This formulation is equivalent to Eq. 10 in that it also removes three degrees of freedom. However, the link bond Q^1-M^1 is not described at the MM level in this case, its length being determined according to Eq. 11 from the Q^1-L distance, which is calculated at the QM level. The Q^1-M^1 bond is therefore in principle allowed to break if Q^1-L breaks. The factor g' is related to g by simply exchanging M^1 and L in the definitions of g below.

If g is chosen to be a constant [47, 105], the distance Q^1-L varies with the length of the Q^1-M^1 bond (or the other way round in Woo's scheme [105]). A suitable choice for g is the ratio of the equilibrium bond lengths for Q^1-L and Q^1-M^1 [47]:

$$g = \frac{d_0(Q^1-L)}{d_0(Q^1-M^1)}; \quad (12)$$

where the values of d_0 can be taken, e.g., from the force field.

Alternatively, the link atom can be positioned at a constant distance from Q^1 [100, 103, 104, 106] by defining g as:

$$g = \frac{d_0(Q^1-L)}{|R_{M^1} - R_{Q^1}|}. \quad (13)$$

The Q^1-L bond length is determined by the constant $d_0(Q^1-L)$, which is typically assigned different values for different types of QM-MM bonds. Note that g now depends on the positions R_{Q^1} and R_{M^1} .

Tavan and co-workers [101] supplement Eq. 13 with a term reflecting the deviation of Q^1-M^1 from its equilibrium value, making Q^1-L again variable. Their correction also accounts for the different stiffness of the Q^1-M^1 and Q^1-L bonds.

In principle, Eq. 10 or Eq. 11 can be extended to include other atoms in the definition of the link-atom position. We are, however, not aware of any link-atom scheme that makes use of this generalization.

2.5.2.2

Link-Atom Forces

If the position of the link atom is not independent but expressed as a function of other atomic positions, Eqs. 10, 12 and 13, its coordinates are eliminated

from the set of coordinates used to describe the entire system. In other words, the link atoms appear only in the internal description of the QM/MM coupling scheme and are transparent to geometry optimization or molecular dynamics algorithms, which handle the entire coordinate set. However, as the QM calculation treats the link atoms generally on the same footing as the atoms of the inner subsystem, there exist forces acting on the link atoms. These forces are relayed onto the atoms appearing in the definition of the link atom coordinates. The link atoms are then effectively force-free, and their coordinates in the next geometry or time step are fully determined by the positioning rule, rather than being propagated according to the forces acting on them.

The distribution of the forces acting on a link atom onto the atoms used in its definition is formulated in terms of the chain rule. The dependence of the total QM/MM energy on the coordinates of the entire system, $\{\mathbf{R}_I\}$, $I \in \mathbb{S}$, and the link-atom coordinates, $\mathbf{R}_L$, $L \in \mathbb{L}$, can be expressed as:

$$\tilde{E}_{\text{QM/MM}}(\{\mathbf{R}_I\}) = E_{\text{QM/MM}}[\{\mathbf{R}_I\}, \mathbf{R}_L(\{\mathbf{R}_I\})] . \quad (14)$$

The derivative of the energy with respect to an atomic position $\mathbf{R}_K$, $K \in \mathbb{S}$, is then obtained from the chain rule (dropping the QM/MM subscript for clarity):

$$\frac{\partial \tilde{E}}{\partial \mathbf{R}_K} = \frac{\partial E}{\partial \mathbf{R}_K} + \frac{\partial E}{\partial \mathbf{R}_L} \frac{\partial \mathbf{R}_L}{\partial \mathbf{R}_K} . \quad (15)$$

The link-atom contribution to the force, the second term of Eq. 15, vanishes if atom K is not involved in the definition of link atom L . There is a corresponding force contribution on K for each link atom in which K is involved.

The notation $\frac{\partial \mathbf{R}_L}{\partial \mathbf{R}_K}$ designates the Jacobian matrix $\mathbf{J}_K$ constructed from the partial derivatives of $\mathbf{R}_L$ with respect to $\mathbf{R}_K$. The elements of this 3×3 matrix depend on the definition used to determine the position of the link atom. If Eq. 10 is used together with the constant g of Eq. 12, the Jacobians $\mathbf{J}_{Q^1}$ and $\mathbf{J}_{M^1}$ take a particularly simple, diagonal form [47]:

$$\mathbf{J}_{Q^1}^{\alpha\beta} = \frac{\partial R_L^\alpha}{\partial R_{Q^1}^\beta} = (1 - g)\delta_{\alpha\beta} , \quad (16a)$$

$$\mathbf{J}_{M^1}^{\alpha\beta} = \frac{\partial R_L^\alpha}{\partial R_{M^1}^\beta} = g\delta_{\alpha\beta} , \quad (16b)$$

where $\delta_{\alpha\beta}$ is the Kronecker symbol and α, β are Cartesian components.

If the bond length Q^1 – L is kept constant by applying Eq. 10 with the g of Eq. 13, off-diagonal elements also appear in the Jacobians because of the dependence of g on the positions of the two frontier atoms [100]. They are,

however, still symmetric:

$$J_{Q^1}^{\alpha\beta} = (1 - g)\delta_{\alpha\beta} + g\widehat{R}_{Q^1,M^1}^\alpha\widehat{R}_{Q^1,M^1}^\beta, \quad (17a)$$

$$J_{M^1}^{\alpha\beta} = g\delta_{\alpha\beta} - g\widehat{R}_{Q^1,M^1}^\alpha\widehat{R}_{Q^1,M^1}^\beta, \quad (17b)$$

where $\widehat{R}_{Q^1,M^1}$ designates the unit vector pointing from Q^1 to M^1 , $\widehat{R}_{Q^1,M^1} = (\mathbf{R}_{M^1} - \mathbf{R}_{Q^1})/|\mathbf{R}_{M^1} - \mathbf{R}_{Q^1}|$.

The corresponding expressions for the forces in Woo's scheme, Eq. 11, are obtained by replacing g by g' and exchanging M^1 and L in Eqs. 16 and 17.

2.5.2.3

Electrostatic Interactions at the Boundary

For the embedding schemes that allow the QM electron density to be polarized by the environment (i.e., electrostatic and polarized embedding, see Sect. 2.3), there exists the problem that the QM density is overpolarized by the rigid point charges of the MM charge model. While this artefact is always present to some extent when a point charge interacts with a polarizable charge distribution, it is the more pronounced (i) the closer the point charge approaches the QM density, and (ii) the more spatially flexible the density is.

The problem is therefore especially critical at the QM–MM boundary in the presence of link atoms. If there are no covalent bonds across the boundary, the van der Waals interaction prevents the atoms from approaching each other too closely. At a link, however, the link atom, which is part of the QM region, is positioned in immediate proximity to the frontier MM atom, typically at a distance of about 0.5 Å. Different approaches to alleviate the resulting spurious polarization effect are discussed below. One should also keep in mind the possibility of other, non-bonded close contacts. In the context of biomolecules, hydrogen bonds across the QM–MM boundary can lead to non-bonded distances between the hydrogen and the acceptor atom of around 1.6–1.7 Å.

Overpolarization is less severe when small, atom-centred basis sets are used in the QM calculation, e.g., a semi-empirical method with a minimal basis. Larger basis sets, which include polarization and diffuse functions, provide more flexibility to place electron density further away from the nuclei and are therefore more prone to overpolarization. Especially affected are methods using plane waves (see Sect. 3.1).

We continue by describing different approaches that have been put forward to mitigate overpolarization within link-atom schemes:

- *Deleting one-electron integral terms:* The spurious electrostatic interaction between the QM density and the MM point charges is dominated on the QM side by the one-electron terms in the Hamiltonian associated with the basis functions and the nuclear charge of the link atoms. It has been sug-

gested that these terms be deleted [4, 51, 52, 104, 108], which effectively removes the interaction of the link atoms with the MM charge environment. While this may be acceptable for semi-empirical QM methods, it becomes problematic when used with higher-level *ab initio* or DFT methods and larger basis sets because it leads to an unbalanced representation of the polarization and electrostatic potential of the QM density [52, 108, 109]. A variant of this procedure is to delete only those terms that involve the link atom and the charge of the MM frontier atom M^1 , which was, however, found to yield inconsistent results [52].

- *Deleting point charges:* The complementary approach to excluding the link atom from interacting with the environment is to delete from the Hamiltonian one or more of the MM point charges at the boundary. These charges then do not interact any more with the QM density as a whole. Different implementations of this idea have been proposed: (i) deletion of only the charge on M^1 [96, 110–114]; (ii) deletion of the charges on M^1 and M^2 [114]; (iii) deletion of those on M^1 , M^2 , and M^3 [3, 114]; or (iv) deletion of those on the atoms belonging to the same charge group as M^1 [52, 96, 102]. With the exception of the latter approach, these schemes suffer from the creation of a net partial charge near the QM region, which leads to severe artefacts [114, 115], and they do not conserve the total charge of the system. Excluding the charge group to which M^1 belongs from interaction with the QM density avoids the most serious problems, at least if the charge group is neutral. However, all these deletion schemes distort the electric field of the environment in the vicinity of the QM region, where it affects the QM density the most, which is not satisfactory.
- *Shifting point charges:* To cure the problems of deleted-charge schemes, different charge-shifting formulations have been introduced. They share the common feature of preserving the charge and sometimes also the dipole in the boundary region, while still removing the overpolarizing partial charge from M^1 . In the charge-shift scheme of Sherwood and co-workers [21, 103, 116, 117], the charge of M^1 is distributed evenly over the M^2 atoms. The dipole created by shifting the charges is compensated for by a pair of point charges placed near each M^2 atom, which generate dipoles of the same magnitude and opposite direction. In a variant of this approach [115], the charge of M^1 is distributed over the other atoms of the charge group, with (“charge shift”) or without (“divided frontier charge”) dipole correction.

In Lin and Truhlar’s redistributed-charge scheme [114], the charge of M^1 is replaced by charges at the midpoints of the M^1 – M^2 bonds. They also propose modification of the values of the M^2 and the redistributed charges such that the M^1 – M^2 bond dipoles are conserved (“redistributed charges and dipoles” scheme). These formulations can be seen as classical substitutes of the generalized-hybrid-orbital (GHO) scheme (see below).

Recent evaluation studies [114, 115, 118] have demonstrated the importance of preserving charges and dipoles in the link region.

- *Charge smearing*: Another alternative is to replace the point charge on M^1 (and possibly other MM atoms near the QM region) by a charge distribution, which significantly reduces the spurious overpolarization. A simple choice for the form of the charge distribution is a spherical Gaussian, $\varrho_{MM}(\mathbf{r}) = q_{MM}(\sqrt{\pi}\sigma_{MM})^{-3} \exp\{-(|\mathbf{r} - \mathbf{R}_{MM}|/\sigma_{MM})^2\}$, where q_{MM} , σ_{MM} , and $\mathbf{R}_{MM}$ are the charge, width, and centre of the charge distribution, respectively. Such a scheme has been applied by Eichinger et al. [101] in their QM/MM implementation in the CPMD code [119], which uses plane waves as the basis set. It was also used within conventional QM/MM methods based on atom-centred basis functions [107, 118]. The delocalized charge representation is applied based on a distance criterion (e.g., to all MM atoms within 5 Å of a QM atom) [101] or according to the connectivity at the link (e.g., to M^1 or M^1 and M^2) [107, 118]. The value chosen for the width parameter σ_{MM} varies between 0.8 and 4 Å.

2.5.3

Boundary Atoms

Boundary-atom schemes replace the MM frontier atom M^1 by a special “Janus” boundary atom that participates as an ordinary MM atom in the MM calculation but also carries QM features to saturate the free valency of Q^1 . They avoid the complications of link-atom approaches related to the introduction of additional atoms, such as ambiguous placement and artificial interactions, and enable one to mimic the electronic properties of the MM moiety at the link.

Most of the boundary-atom schemes proposed are based on a type of monovalent pseudopotential (or effective potential) that is parameterized to reproduce certain desired properties and that is located at the position of M^1 :

- *Adjusted connection atoms* [120]: Defined within semi-empirical QM methods (MNDO, AM1, PM3), adjusted connection atoms feature one valence electron in an s-orbital. They were fitted using a set of 30 molecules to mimic the structural and electronic (charges, dipole moment, formation energy) features of a methyl group by adjusting the atomic parameters of the respective method. The scheme is intended to saturate a cut C–C single bond.
- *Pseudobonds* [121, 122, 738]: The pseudobond approach, developed for ab initio and DFT methods, uses a monovalent, fluorine-like boundary atom with seven valence electrons, $Z = 7$, and an angular-momentum-dependent effective core potential (ECP); in the original formulation [121], it carries a fluorine 3–21G or 6–31G* basis set. With the aim of capping a $C(sp^3)$ – $C(sp^3)$ bond, the six ECP parameters (for 6–31G*) were

determined from six structural and electronic properties of ethane, as calculated with B3LYP. A given pseudobond is therefore specific to the bond type and to the basis set used in the parameterization. It is, however, independent of the MM force field and only weakly dependent on the QM method. Although developed within B3LYP, the pseudobond parameters are transferable to other DFT or HF calculations. The MM point charges of the charge group to which M^1 belongs are deleted.

In a recent modification of the scheme [122], $C(sp^3)$ –carbonyl $C(sp^2)$ and $C(sp^3)$ – $N(sp^3)$ pseudobonds were presented in addition to $C(sp^3)$ – $C(sp^3)$. A simpler, angular-momentum-independent form of the ECP was adopted, together with an STO-2G basis set on the boundary atom, which is determined by six parameters.

- *Effective group potentials (EGP)* [123–127]: Aimed at replacing ligands such as CO, NH_3 , CH_3 , or cyclopentadienyl in transition-metal complexes, effective group potentials were also proposed [126] for use as boundary atoms in QM/MM schemes. We are, however, not aware of any applications so far.

The EGP can be regarded as a type of generalized ECP, expressed as an expansion over Gaussian projectors that may be located on different centres. The EGP is determined such that a reduced representation of the system mimics as closely as possible a suitably chosen reference system. Only the electrons involved in bonding are described explicitly, e.g., one in the case of CH_3 .

- *Quantum capping potentials (QCP)* [128, 129]: These combine the form of shape-consistent ECPs with the idea of a one-electron potential. By adding additional spherical terms to a standard carbon ECP with four valence electrons and fitting to molecular properties of ethane, effective potentials with one explicit electron were obtained. These quantum capping potentials (QCPs) mimic a methyl group at the QM/MM boundary.
- *Effective Hamiltonians from a minimum principle* [130]: A formal framework was proposed in which an effective QM/MM Hamiltonian is defined that provides the best approximation in a least-squares sense to the full QM Hamiltonian under the condition that no basis functions are located in the outer region. ECPs or point charges, however, may be present. This effective Hamiltonian is transferable, that is, it is independent of the inner subsystem.

From this formalism, both a one-electron ECP and a classical potential term were obtained that describe a methyl group in ethane and were applied to several small test systems. Moreover, the LSCF (see below), pseudobond, QCP, and EGP schemes were shown to be derivable from within the formalism.

- *Optimized effective core potentials (OECP)* [131]: QM methods using plane waves as basis set are often implemented within the pseudopotential approach to eliminate the core electrons from the calculation. It seems

therefore natural to exploit the same formalism to describe boundary atoms in plane-wave-based QM/MM calculations. R  thlisberger and co-workers [132] have used a one-electron ECP, empirically optimized to reproduce the C–C distance in ethane.

Recently, the same group [131] proposed a scheme that provides a systematic way of deriving optimized effective core potentials for use as boundary atoms. It is based on a form for the pseudopotential frequently employed in plane-wave calculations. The parameters are obtained by minimizing a penalty function depending on the electron density. The scheme was successfully tested on the methyl group of acetic acid, which was replaced by a seven-electron OECP.

- *Multicentred valence-electron effective potentials (MC-VEEP)* [133]: These effective potentials are able to treat both ground and excited electronic states correctly. They exploit the established quantum-chemical ECP approach and build on ideas similar to those used for the QCP method and the minimum effective Hamiltonian. At the QM/MM boundary, these one-electron potentials replace a methyl group.

2.5.4

Frozen Localized Orbitals

The approach of using a frozen hybrid orbital to saturate the dangling bond at the QM–MM boundary dates back to Warshel and Levitt [1]. Different schemes have been elaborated that share the idea of placing a set of suitably oriented localized orbitals on one of the frontier atoms and keeping some of these orbitals frozen, that is, they do not participate in the SCF iterations:

- *Local self-consistent field (LSCF)* [134–138]: In the LSCF method, developed by Rivail and co-workers, one starts out with a QM calculation on a model system that contains the frontier bond to be described. Applying a localization scheme, one constructs a strictly localized bond orbital (SLBO) for this bond. The SLBO has contributions from the frontier atoms only and is assumed to be transferable. In the QM/MM calculation, it is excluded from the SCF optimization and does therefore not mix with other orbitals. Its orientation is always kept along the Q^1 – M^1 vector. It may be described as a kind of frozen lone pair on Q^1 pointing towards M^1 . To compensate for the additional electron introduced with the doubly occupied SLBO, an extra charge of $1e$ is placed on M^1 , which interacts with all other MM charges. On the model compound, a special classical bond potential with five parameters is fitted, which is used in the QM/MM calculation together with the SLBO (parameters for common bonds are listed in [138]). The MM charges on M^1 and M^2 are adjusted as necessary to obtain a balanced description of the frontier bond and the polarization of the QM region, while maintaining the overall charge [138, 139]. It has

also been suggested that the total point charge on M^1 (i.e., compensation charge + MM partial charge) be replaced by a Gaussian charge distribution [140].

Very recently, the performance of various localization schemes used in the construction of the SLBOs was assessed and compared to extremely localized molecular orbitals (ELMOs) [141, 142]. The latter were found to be superior because of their better transferability. They avoid the somewhat arbitrary deletion of orbital contributions not localized on the frontier atoms.

- *Frozen orbitals* [93, 143, 144]: Friesner and co-workers have presented a formulation of the LSCF procedure that differs in some technical details from the original one; for instance, the compensation charge is placed at the midpoint of the Q^1-M^1 bond. Furthermore, there is a major conceptual difference as compared to most other QM/MM schemes in that the QM–MM interactions at the boundary are heavily parameterized: (i) Several electrostatic correction terms are introduced that reduce the short-range electrostatic interactions at the interface, following the spirit of 1–2, 1–3, and 1–4 electrostatic exclusion and scaling rules used in many force fields. These corrections also require the assignment of point charges to the atoms of the inner subsystem and involve the optimization of the MM, QM, and bond partial charges in the boundary region. (ii) As mentioned in Sect. 2.4.1, the van der Waals parameters of the QM atoms are re-optimized. (iii) Certain classes of hydrogen bonds across the boundary are described by an additional repulsive term. (iv) The QM–MM bonded terms are re-optimized, rather than taken directly from the force field.

The goal of this extensive parameterization is to reproduce as closely as possible the conformational and reaction energetics in the boundary region. A database of parameters has been derived for QM–MM bonds in the backbone and on the side chains of proteins. The parameterization is specific for the basis set and the QM method.

- *Generalized hybrid orbitals (GHO)* [145–150]: The GHO method of Gao and co-workers is closely related to the LSCF and frozen-orbital approaches in that it constructs localized hybrid orbitals and freezes some of them. However, it places the set of localized hybrid orbitals on M^1 , rather than Q^1 . M^1 thus becomes a boundary atom. (The classification of boundary methods into boundary-atom and frozen-orbital schemes is therefore somewhat arbitrary.) The orbital pointing towards Q^1 is active and participates in the SCF iterations, while the remaining “auxiliary” hybrids are kept frozen and are not allowed to mix with the other orbitals.

The standard $C(sp^3)$ boundary atom in a HF or DFT calculation has four electrons, $Z = 4$, and a minimal basis set, from which four localized

sp^3 hybrid orbitals are constructed. The MM point charge of M^1 is distributed equally over the three frozen auxiliary hybrids. They thus provide a type of pseudopotential that mimics the electronic character at the link. The hybridization is completely determined by the local geometry at the boundary, that is, by the relative positions of Q^1 , M^1 , and M^2 . In contrast to the LSCF and frozen-orbital methods, there is thus no need for parameterization calculations on model compounds to derive the localized hybrids.

However, to improve the structure at the boundary, in particular the Q^1-M^1 distance and $Q^1-M^1-M^2$ angles, some additional parameters are introduced. Depending on the QM level at which the GHQ scheme is implemented, certain classical bonded terms involving M^1 are modified or added and/or certain integrals are scaled. For instance, in the HF and DFT implementations, the one-electron, one-centre kinetic-energy integrals involving orbitals on M^1 or on M^1 and Q^1 are scaled, which introduces seven parameters.

2.5.5

Boundary Schemes: Summary

Several studies have evaluated the merits and drawbacks of different boundary methods. As link atoms are the most widely used boundary scheme, most of these assessments [52, 109, 114, 115, 118, 139, 140, 151] compare link-atom approaches, which differ in particular in the way that the charges at the boundary are handled. Some have also compared link-atom to localized-orbital schemes [109, 139, 140].

Approaches based on hybrid orbitals are certainly more fundamental from the theoretical point of view, providing a boundary description essentially at the QM level. They also avoid some of the complications inherent in the link-atom method arising from introducing additional atoms. However, they are technically considerably more complicated, not least because of the orthogonality constraints required to prevent the mixing of frozen and active orbitals. In addition, the localized orbitals themselves, or specific parameter sets related to them, must be determined beforehand, involving calculations on model compounds. These parameters are usually not transferable and need to be reconsidered upon changing the MM force field or the QM method or basis set.

The conclusion from the available evaluations is that localized-orbital approaches can sometimes be tweaked more specifically towards a given application, but that the performance of link-atom schemes is generally on par. They both provide reasonable accuracy when used with care; in particular, one should minimize distortions of the charge distribution at the boundary.

3

Choice of QM and MM Models, QM/MM Implementations

3.1

Choice of QM Method

The QM/MM formalism as such is sufficiently flexible to accommodate almost any QM method. The particular choice thus follows the same criteria as in pure QM studies and is not further elaborated on here. Especially with link-atom schemes, only minimal changes to the QM code are required. Essentially, the QM code must be able to perform the SCF treatment in the presence of the external point-charge field, representing the MM charge model in the case of electronic or polarized embedding. Other boundary schemes can require somewhat larger modifications.

In practice, many current biomolecular QM/MM applications use DFT as the QM method due to its favourable cost/accuracy ratio. Traditionally, semi-empirical QM methods have been the most popular, and they remain important for QM/MM molecular dynamics, where the aspect of computational cost is paramount. They are also very useful in method evaluation studies because they can be expected to enlighten the same problems as would occur with more costly *ab initio* methods. Especially in the context of biomolecular QM/MM studies [115, 152, 153, 739], the semi-empirical, DFT-inspired SCC-DFTB (self-consistent-charge density-functional tight-binding [154]) method due to Elstner and co-workers appears promising because it approaches, within the validity domain of the parameterization, the accuracy of DFT at the cost of a semi-empirical treatment.

At the other end of the spectrum are the post-Hartree-Fock *ab initio* electron-correlation methods, such as those based on Møller-Plesset perturbation theory (e.g., to second order, MP2) or coupled-cluster theory (e.g., CCSD including single and double excitations or CCSD(T) adding a perturbative treatment of triple excitations). Recent developments (exemplified by the work of Schütz, Werner, and collaborators [155–162]) have extended the size of systems that can be treated with these methods by almost an order of magnitude to several tens of atoms. They take advantage of the short-ranged nature of electron correlation and are commonly referred to as local methods (e.g., LMP2, LCCSD); their computational effort scales linearly with system size. The superior accuracy of high-level *ab initio* methods can therefore now also be exploited for biomolecular QM/MM studies [740], certainly at the level of energy calculations at fixed geometries (i.e., single points).

Plane-Wave Methods

We highlight here specific issues of QM methods employing plane waves as the basis set within a QM/MM framework. The application of such methods is intimately linked to Car–Parrinello first-principles molecular dynamics [163, 164], which almost always treats the electronic-structure problem within DFT in a plane-wave basis. There are two main issues, both concerning the electrostatic QM–MM coupling used in electrostatic or polarized embedding.

The first is a fundamental issue, the severe overpolarization of an electron density expanded in plane waves by a bare point charge, sometimes referred to as the “electron spill-out” problem. While any charge distribution is overpolarized to some extent when interacting with a point charge (see Sect. 2.5.2) the effect is completely deleterious when plane waves are used. Because they form an origin-less basis set, the density is pulled away from the nuclei and localizes in the purely attractive potential around the point charges. The problem has been dealt with in different ways:

- By substituting the point charges within a certain distance from the QM region by Gaussian charge distributions (see Sect. 2.5.2) [101].
- By coupling the point charges to a model density of atom-centred Gaussians [165], which reproduces the multipoles of the true density and is variational with respect to the true density [166].
- By smoothly replacing the Coulomb potential at short range by a form which goes to a constant at zero distance [132].
- By representing the point charges as Slater-type *s*-functions (or, more generally, partial-wave expansions), which also provides a finite potential at zero distance [167].

The second problem is of a more technical or algorithmic nature. Directly evaluating the (possibly modified) Coulomb interaction between the MM point charges and the QM density presents a considerable computational effort. In a plane-wave scheme, the electrostatic potential due to the total density is represented on a real-space grid in the unit cell. Hence, computation of the Coulomb energy requires $N_q N_r$ evaluations, where the number of charges, N_q , is of the order of 10^4 and the number of grid points, N_r , is of the order of 100^3 . Several techniques have been proposed to reduce the cost:

- A hierarchical multipole expansion is used to represent the electrostatic potential at the grid points due to the point charges [101]. However, this scheme is not symmetrical (i.e., not Hamiltonian and therefore not energy-conserving) in that the electrostatic potential acting on the MM charges due to the density is derived from a point-charge model of the density.
- Beyond a chosen distance, the Coulomb interaction is evaluated from a multipole expansion of the density [132] and directly otherwise. An

intermediate layer can be defined, where the density is represented by variational electrostatic-potential-derived (ESP) charges [168, 169] to calculate the interaction with the MM charges.

- The point charges are represented by Gaussian charge distributions and their potentials by sums of Gaussians with different widths (Gaussian expansion of the electrostatic potential, GEEP) [170]. These Gaussians are then mapped onto the suitable grid level of a multigrid scheme. The potential on the finest grid, which couples to the density, is obtained by sequentially interpolating starting from the coarsest grid level. This procedure reduces the computational cost by up to two orders of magnitude without introducing cutoff parameters that need to be adjusted. It scales linearly for systems as small as a few hundred atoms. Recently, an extension to periodic boundary conditions has appeared [741].

3.2

Choice of MM Method

The QM/MM formalism is also largely independent of the choice of the MM method. Subtractive QM/MM schemes are generally more easily compatible with wider classes of force fields than additive ones. As far as force fields for (bio)molecular systems are concerned [171–175], they are all valence force fields using point-charge models. As discussed in Sect. 2.3.3, there is no established polarizable biomolecular force field available as yet. One can further differentiate between all-atom and united-atom force fields. The latter describe explicitly only selected hydrogen atoms (typically the polar ones) and replace the remaining ones by a suitably parameterized “united atom” representing, e.g., a CH₂ unit.

We list here the most widely used biomolecular force fields as well as a few more general ones:

- Biomolecular force fields: AMBER [66, 176–178], CHARMM [39, 179–182], GROMOS [183, 184], OPLS-AA [185–187]
- General-purpose force fields: MM3 [188–194], MM4 [195–205], MMFF [206–213], UFF [214–217]

We give only the “family names”, without detailing specific variants. We emphasize that it is important to distinguish very clearly the force field proper from the force-field engine (i.e., the program in which it is implemented), especially when they bear the same name, and to specify precisely which parameter set was used (as characterized, e.g., by the exact designation and citation). We also note that “biomolecular” typically includes proteins [171, 172] and in many cases also nucleic acids [173, 174], but less frequently carbohydrates [175] or lipids. See [40–42] for recent surveys on available force fields.

3.3

QM/MM Implementations

3.3.1

Program Architecture and QM/MM Packages

There are essentially three main approaches in which QM/MM implementations have been realized: (i) by adding QM capabilities to an MM package; (ii) by adding MM functionality to a QM package; or (iii) in a modular manner by coupling existing QM and MM programs to a central QM/MM engine. Approaches (i) and (ii) take advantage of the inherent strengths of the respective base program. MM packages are designed to handle large, complex systems and offer the corresponding simulation and analysis tools, whereas quantum-chemistry programs traditionally provide, e.g., efficient algorithms to locate stationary points on the potential-energy surface.

The modular approach (iii) minimizes as far as possible modifications of the standard QM and MM codes and offers more flexibility. The external QM and MM packages are linked via interfaces to a central core that supplies the QM/MM coupling as well as routines for structure optimization, molecular dynamics, etc. The system is relatively easily extended by interfacing to additional QM or MM programs. When updated versions of the external programs become available, they can normally be used immediately or with only minimal changes to the interface routines. The core also provides a common user interface to the external programs, at least for the most common tasks.

There are also drawbacks to the modular architecture: (i) the increased complexity of the program because of the need to create a unifying framework that is able to accommodate external programs with possibly very different designs; (ii) the considerable amount of data being transferred between the core and the external programs and between different core modules, which is often done by writing and reading files on disk; and (iii) the repeated start-up, initialization, and close-down of the external programs on each call. The latter two points have implications for the efficiency of the whole program system.

We list here examples for each of the three types of QM/MM implementations and describe in somewhat more detail the modular QM/MM package CHEMSHELL that is co-developed in our laboratory:

- MM packages with QM: AMBER [218, 702], BOSS (MC simulations, semi-empirical QM only) [219, 703], CHARMM [180, 704]
- QM packages with MM: ADF [220, 705], GAMESS-UK [221, 706], GAUSSIAN [707], NWCHEM [222, 708], QSITE/JAGUAR [709]; Car-Parrinello MD codes with QM/MM capabilities: CPMD [101, 119, 131, 132, 168, 169, 223, 224], CP-PAW [99, 105, 165, 225–227, 710].
- Modular QM/MM packages: CHEMSHELL [117, 711], QMMM [712]

The ChemShell package

Following the modular design approach sketched above, CHEMSHELL is presently interfaced with the following QM packages (see Fig. 3): GAMESS-UK [221, 706], GAUSSIAN03 [707], MOLPRO [713], MNDO99 [228], ORCA [714], and TURBOMOLE [715]. On the MM side, interfaces are available to CHARMM [180, 704], GROMOS96 [184, 716], and GULP [229, 230, 717]. These external programs are supplied with input data (coordinates of atoms and point charges, command file) by the core and return energy and gradient data.

The core modules provide the following key functionalities:

- QM/MM coupling; mechanical and electrostatic embedding, shell model for solid-state embedding; link atoms with charge shifting; double link atoms with charge smearing
- Structure optimization using different types of coordinates (Cartesian, internal, hybrid delocalized coordinates)
- Molecular-dynamics driver for *NVE*, *NVT*, and *NPT* simulations with rigid-body motion (quaternions) or SHAKE constraints
- Vibrational frequencies from finite-difference second derivatives
- A general force-field engine compatible with the functional form of various force fields (e.g., AMBER, CHARMM, UFF)
- Utility routines for data management, coordinate manipulations, etc.

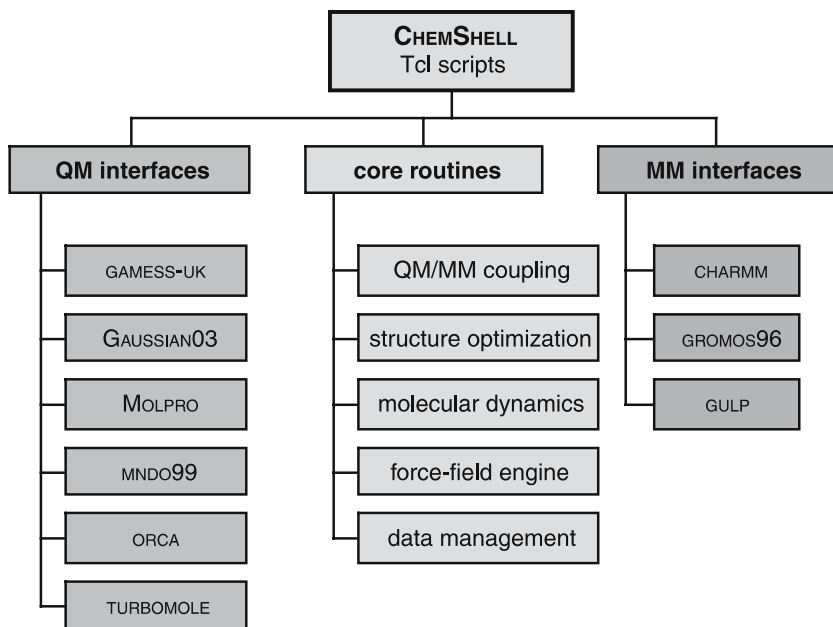


Fig. 3 Schematic design of the modular CHEMSHELL QM/MM package

The MD driver and the force-field engine are built on routines from DL_POLY [718] that are tightly integrated into the core.

The controlling framework and the user interface are implemented within the tool command language (Tcl). The user input is in the form of Tcl scripts. The various functions are implemented either directly in Tcl or as Tcl wrappers to lower-level routines written in C or Fortran.

Current development projects include the incorporation of a polarizable water model due to van Gunsteren and co-workers [61] within the GROMOS interface as well as algorithms for QM/MM calculations on excited states (localization of conical intersections, surface hopping).

3.3.2

QM/MM-Related Approaches

Some QM/MM or QM/MM-like schemes, which have been applied to biomolecular systems, are sufficiently conceptually distinct from the “standard” QM/MM implementations to be mentioned separately:

- *Empirical valence-bond (EVB) method* [231–234]: Within a valence-bond description of the reactive part of the system, the energies of the interacting diabatic (resonance) states are represented by simple empirical potential terms. They incorporate the interaction of the charges of the states with the environment. No QM (electronic-structure) calculations are involved. Two-state models are the most common. The EVB energies are calibrated to reproduce experimental or ab initio data. The method has been applied notably by Warshel and collaborators to model the influence of the solvent or protein environment on reactions. It offers the possibility of driving a reaction from reactants to products by means of a mapping potential that gradually transforms the diabatic state describing the reactant into the product state, which can be considered a kind of electronic reaction coordinate.
- *Effective fragment potentials (EFP)* [235–238]: From separate calculations on model systems, one derives a set of one-electron terms (the EFP) that account for the electrostatic, inductive, and repulsive interactions of a moiety. For instance, the electrostatic potential is represented by a distributed multipole expansion. The EFP is then incorporated into the Hamiltonian of a QM calculation, where it describes the effects of the environment on the QM part. The method was originally designed to model the solvent environment in QM calculations by means of EFPs. These fragments may be considered the MM part of a QM/MM scheme. The scope was extended to biomolecular systems, using an LSCF-type procedure (see Sect. 2.5.4) to treat covalent bonds across the boundary [237, 238]. Recent developments have lifted the limitation that the fragments be fixed in space [239–241]. By defining relatively small fragments (a few

atoms) and treating the fragment–fragment interaction by an empirical potential, the MM subsystem can be represented as a chain of effective fragments, which are internally rigid but are allowed to move relative to each other.

- *QM/MM structure refinement* [242–244]: Ryde and co-workers recently introduced an approach that integrates QM calculations into the refinement of X-ray crystallographic or NMR-derived structural data of biological macromolecules, in particular proteins.

Conventionally [245], an MM model of the structure is refined against the experimental electron-density data. Whereas these models have been carefully tuned for the macromolecular part, they are clearly less reliable when it comes to substrates, products, or other “ligands” bound in the active site or to cofactors and prosthetic groups, that is, compounds of other chemical classes also present in the structure. This poses the dilemma that the structural information in and around the active site, the focus of most mechanistic studies, is usually the least dependable [246].

The new approach alleviates this problem by describing, e.g., the non-protein components in the active site of an enzyme at the QM level, while the remainder is treated by the force field. In contrast to standard QM/MM schemes, but in line with established refinement procedures, no electrostatic interactions are considered, neither between QM and MM regions nor within the MM part. The program, named ComQUM-X, is derived from a QM/MM implementation by Ryde and co-workers. Merz and co-workers have recently presented a similar approach [742].

4

Optimization and Simulation Techniques Used with QM/MM

4.1

General Comments

The QM/MM method as presented in Sect. 2 is a potential-energy scheme yielding the energy (and the forces) at a given structure. As such it can be combined with any approach that makes use of this information to update the structure, that is, with any optimization, molecular-dynamics, or Monte Carlo technique. Moreover, a single QM/MM energy and gradient evaluation is not significantly more costly than the corresponding QM calculation alone.

However, it is an almost intrinsic feature of the QM/MM method that it is applied to “large” systems with at least hundreds, but often several 10 000 structural degrees of freedom. Two types of problems hence arise: (i) There are algorithmic complications due to the mere system size. Algorithms for

manipulating the coordinates (e.g., in optimization schemes) whose computational effort per step scales as $\mathcal{O}(N^2)$ or $\mathcal{O}(N^3)$, where N is the number of degrees of freedom, are not computationally viable any more. At the same time, procedures with more favourable scaling, well-established from pure MM calculations, are often not suitable either because they need much more (however less costly) steps, which is not affordable with QM/MM methods. (ii) Conformational complexity is an issue. The number of steps required to converge an optimization or to obtain sufficient sampling increases with the number of degrees of freedom.

The optimization and simulation techniques presented in the following sections follow two strategies to address these problems: (i) Some are generally suitable for large systems, e.g., by scaling as $\mathcal{O}(N)$. (ii) Others take advantage specifically of the partitioning of the system into the QM region, where evaluating energy and gradients is computationally demanding, and the MM region, where these calculations are almost for free.

Related to the determination of stationary points (minima, saddle points) on the potential-energy surface (PES) for biomolecular reactions, a comment is in order that projects beyond the purely algorithmic issues discussed below. The vast size of the configuration space available to a reacting biomolecular system as compared to a “small-molecule” reaction drastically reduces the significance of single stationary points. As opposed to typical QM studies on smaller molecular systems, knowledge of the optimized reactant, transition state, and product is not “everything”, even if a reaction path connecting them has been determined and even if we disregard finite-temperature effects for the moment. As recently re-emphasized by Warshel and co-workers [247], one should in QM/MM optimization studies consider at least several “representative” transition states with their corresponding minima. While this does not fully substitute an extensive configurational sampling, it reflects at least partly the influence of the conformational diversity of the environment on the reaction. An expedient way to generate a selection of environmental configurations is, for instance, to take snapshots from a classical MD trajectory and use them as starting structures in subsequent QM/MM optimizations, as demonstrated, e.g., in [248].

4.2

Stationary Points and Reaction Paths

For general reviews on algorithms for locating minima and (first-order) saddle points on the PES, we refer to [27, 249–253]. We first briefly consider optimization methods treating the complete system on equal footing, which thus are generally applicable to large systems and are not specific to the QM/MM approach. Secondly, we describe techniques specifically designed for, and taking advantage of, the QM–MM partitioning.

4.2.1

General Optimization Techniques for Large Systems

Among the most efficient algorithms developed to locate stationary points on the PES are quasi-Newton methods using some form of internal coordinates. Efficiency here means that they take a minimal number of steps to convergence, that is, a minimal number of energy and gradient calculations. They require an (approximate) Hessian, which is updated, e.g., according to the Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula for minima. The matrices involved in the conversions between Cartesian and internal coordinates and the Hessian itself are $N \times N$; their diagonalization is an $\mathcal{O}(N^3)$ operation, storage scales as $\mathcal{O}(N^2)$, which becomes impractical for large systems.

A popular optimization algorithm that scales linearly in both CPU and memory usage is the limited-memory BFGS (L-BFGS), which stores only the diagonal of the Hessian and uses information only from a limited number of previous steps. However, it does not resolve the scaling problem of the coordinate-conversion steps. Optimization methods that reduce the overall scaling to $\mathcal{O}(N^2)$ or even $\mathcal{O}(N)$ by various algorithmic enhancements have been developed, e.g., by Farkas and Schlegel [254–257], Pulay and co-workers [258–260], and others [261–263].

The scheme by Billeter et al. [264] divides the system into fragments and performs expensive coordinate manipulations only within these fragments (“divide-and-conquer” approach). Each fragment is described by a set of non-redundant delocalized coordinates [265], built from redundant primitive internals [266]. However, to preserve the information about the fragment’s absolute position in space, its Cartesian coordinates are included together with the internals in the delocalization procedure, affording hybrid delocalized coordinates (HDLs). Using them within the L-BFGS algorithm provides an overall $\mathcal{O}(N)$ scheme.

For the optimization of transition states, the idea of micro-iterations [43, 267, 268] has been combined with HDLs [264]. A core fragment is defined that contains the atoms immediately involved in the reaction. Within this core, a partial rational-function optimizer (P-RFO) with Powell updates of the approximate Hessian is applied; the initial core Hessian is calculated exactly. With the core kept frozen, the environment fragments are fully relaxed using the L-BFGS algorithm in HDLs. One P-RFO step is then taken in the core, and the procedure is iterated to convergence. This scheme thus searches for a first-order saddle point, controlled by the (low-dimensional) core Hessian, while minimizing the energy with respect to the environmental degrees of freedom. The partitioning into core and environment is exact in the limit of a completely relaxed environment (i.e., exactly vanishing gradients). The environment then adiabatically follows the core degrees of freedom. This is approximated by applying a stricter convergence criterion to the environment

than to the core. These procedures have been implemented in the optimizer HDLCOpt [264] and are part of the CHEMShell package.

A combination of an RFO with an L-BFGS algorithm for core and environment, respectively, has been described that uses Cartesian coordinates [269]. The core and environment optimizations are alternately iterated to convergence.

4.2.2

Optimization Techniques Specific to QM/MM

The principal idea here is to exploit the division of the system into QM and MM regions in the spirit of the micro-iterative scheme [43, 267, 268]. The core/environment partitioning thus reflects the QM/MM partitioning. Different coordinates and optimization algorithms can be applied in the two regions; for instance, plain Cartesians with a conjugate-gradient or truncated-Newton algorithm for the MM degrees of freedom (which avoids costly coordinate or Hessian manipulations) and an efficient quasi-Newton algorithm in some form of internal coordinates for the QM degrees of freedom. If desired, additional atoms beyond the QM part may also be included in the core region.

There are essentially two possibilities to achieve mutual convergence of the optimizations in the two parts: (i) the adiabatic approach, in which the MM system is fully relaxed in each QM step [43, 93, 139, 270, 271]; and (ii) the alternating scheme, where the QM and MM optimizations are alternately iterated to convergence [3, 272, 273]. The QM atoms are always kept fixed during the MM optimization and vice versa. Different options within the micro-iterative scheme (including the frequency of environment minimizations, the size of the core, and approximations to the electrostatic QM-MM interaction) are assessed in [274] for an enzymatic reaction.

The use of different coordinate systems and optimizers can compromise the convergence behaviour and lead to numerical instabilities. The reasons are that the approximate QM Hessian contains no information about the MM atoms (i.e., the QM and MM degrees of freedom are treated as decoupled) and that the Hessian-update procedures may become unstable as a result of spurious force contributions on QM atoms due to the environment [271]. An adiabatic optimization scheme for minima that remedies these complications for mechanical embedding has very recently been presented [271]. It uses Cartesians for the MM atoms and redundant internal coordinates in an RFO-type algorithm for the QM atoms. QM and MM degrees of freedom are explicitly coupled in the QM optimization by virtue of MM contributions to the Hessian, which are analytically evaluated. The scheme takes advantage of $\mathcal{O}(N)$ techniques and employs fast-multipole methods to evaluate the electrostatic interactions.

Separating the QM and MM optimizations is particularly straightforward for mechanical embedding and has been used, for instance, by Maseras and

Morokuma [43] with their IMOMM scheme. For electrostatic embedding, however, the difficulty arises that the electrostatic QM-MM interaction is evaluated at the QM level. That is, a QM calculation is required in each MM step to let the density adapt to the new MM configuration and to obtain the forces on the MM atoms due to the QM density. To restore the decoupling of QM and MM calculations, one can represent the QM density in the MM calculations by a point-charge model. Schemes based on atomic charges fitted to the electrostatic potential (ESP) have been proposed, the basic assumptions being that the QM charge distribution is sufficiently accurately described by the charges and that its relaxation during the MM optimization can be neglected.

Yang and co-workers [273] implemented an alternating scheme. They derive the ESP charges from the optimized QM part, use them to optimize the MM part, then re-optimize the QM part, and so on. The drawback of this procedure is that the final structure and energy are not consistent because the structure is obtained using the ESP charges, whereas the energy is calculated from the full QM density.

An adiabatic approach has been presented by Friesner and co-workers [93], who use the electrostatic force calculated with the ESP charges as a correction during the MM optimization. In each step of the QM optimization, the true force $F_{\text{QM-MM}}^{\text{el, QM}}(0)$ and its approximation $F_{\text{QM-MM}}^{\text{el, ESP}}(0)$ are obtained. Their difference is added to the ESP-derived force in each step k of the subsequent full MM optimization:

$$F_{\text{QM-MM}}^{\text{el}}(k) = F_{\text{QM-MM}}^{\text{el, ESP}}(k) + F_{\text{QM-MM}}^{\text{el, QM}}(0) - F_{\text{QM-MM}}^{\text{el, ESP}}(0). \quad (18)$$

When the MM optimization has converged, the next QM step is taken. Once mutual convergence has been achieved, no more MM steps are needed at a given QM configuration, hence the correction $F_{\text{QM-MM}}^{\text{el, ESP}}(k) - F_{\text{QM-MM}}^{\text{el, ESP}}(0)$ vanishes, yielding a fully consistent structure and energy.

A modification of this procedure has been suggested [270], in which the density and the MM configuration (i.e., SCF calculations and MM optimizations) are iterated to convergence before the next QM step is taken. This eliminates the approximation that the change in the density due to the changed MM configuration is negligible and guarantees that the QM optimizer always obtains fully consistent forces. It was also proposed that fast multipole methods be used instead of the ESP charge model to calculate the QM-MM electrostatic interaction. However, we are not aware of a practical implementation of these suggestions as of this writing.

Martí et al. proposed use of a lower-level QM scheme to calculate the QM-MM electrostatic interaction during the MM optimization, rather than a point-charge model [275–277]. This leads to smoother convergence behaviour because the QM density can adapt to changes in the MM environment, and is not much costlier than a point-charge description.

4.2.3

Reaction-Path Techniques

The nudged-elastic-band (NEB) method [278–282] represents the minimum-energy path (MEP) by a chain of states in configuration space (also called replicas) linked by springs. The optimized MEP provides a continuous path connecting reactant and product states without assuming a specific reaction coordinate. Yang and co-workers have extended the NEB method for large systems [283]. In particular, they include in the path definition only selected degrees of freedom involved in the reaction and make sure that the environment follows the reaction smoothly.

The same group also adapted the path-optimization procedure by Ayala and Schlegel [284] for the use within a QM/MM approach [285]. They make use of their alternating QM/MM optimization scheme [273] and again restrict the degrees of freedom considered in the definition of the distance between configurations to the relevant coordinates. Recently, they have combined their NEB and path-optimization schemes into a two-step procedure [286]. The NEB method is used to obtain an approximate reaction path, which is subsequently refined with the path optimizer.

A “chain-of-replicas” method has been proposed [287] that replicates only a selected subsets of atoms and restrains them based on the root-mean-square rotational and translational distance to the adjacent steps. Implemented in CHARMM, it is suitable for use with QM/MM approaches and is efficiently parallelized.

4.3

Molecular Dynamics and Simulation Techniques

As mentioned above, the QM/MM energy and forces can in principle be used within any molecular dynamics (MD) or Monte Carlo (MC) scheme. In most cases, the rationale behind such simulations is the sampling of phase (or rather configuration) space to calculate statistical-thermodynamical ensemble averages. Typical examples include free-energy differences, such as reaction or activation free energies or free energies of solvation.

As the amount of sampling necessary to obtain converged averages is considerable, especially for complex systems such as biomolecules, the computational demands of these simulations are extremely high, even considering the fast increase in available computing power. Similarly to the optimization schemes presented in the previous section, simulation approaches have therefore been developed that reduce the computational cost by taking advantage of the QM–MM partitioning; in particular, one tries to avoid the expensive direct sampling of the QM contribution.

We highlight in this section a number of simulation techniques that have been applied in QM/MM calculations on biomolecular systems.

4.3.1

QM/MM Molecular-Dynamics and Monte Carlo Simulations

Historically, full QM/MM simulations (i.e., with freely moving QM atoms) were first applied as explicit-solvent models to calculate solvation free energies or reaction free energies in solution. Both molecular dynamics and Monte Carlo were used as sampling methods, along with standard free-energy methods such as free-energy perturbation (FEP), umbrella sampling, and thermodynamic integration (see [288–292] for general overviews on MD, MC, and free-energy methods). In most cases, a semi-empirical QM method was used. We note here contributions by Bash, Field, and Karplus [293], Gao and co-workers [5, 6, 294, 295], and Merz and co-workers [296–299]; the latter group also utilized first-principles QM methods (DFT, HF).

Probably the first full QM/MM MD simulation on a biomolecular system (a metalloenzyme) was reported by Hartsough and Merz in 1995 [300] with a semi-empirical QM method. Whereas calculations at this level are now well affordable (e.g., [301]), QM/MM MD with a first-principles QM method has remained very computationally demanding also by today's standards and is not routinely performed.

In the above-mentioned studies, the QM energy and forces are taken from a converged SCF calculation in each step of the dynamics, that is, they employ a Born–Oppenheimer MD scheme. An alternative is Car–Parrinello MD (CP-MD) [163, 164], where the wave functions are treated as fictitious dynamic variables within a Lagrangean scheme and follow the nuclei “on the fly”. QM/MM approaches based on CP-MD have been developed by Woo et al. [105, 165, 225–227], Tavan and co-workers [101], and Röthlisberger and collaborators [131, 132, 168, 169, 223, 224] (see also Sects. 3.1 and 3.3.1). Notably, the latter group has been active in the area of biomolecular simulations.

While the Car–Parrinello scheme is most commonly implemented with DFT and plane waves, a very much related extended-Lagrangean formulation using atom-centred Gaussians has been presented, the atom-centred density-matrix propagation (ADMP) method [302–306]. It has recently been combined with the ONIOM QM/MM scheme [307]. We are, however, not aware of any biomolecular applications so far.

Also within QM/MM MD, one can apply the adiabatic micro-iterative scheme by optimizing the MM degrees of freedom in each step of the QM dynamics. In a sense, this can be considered a limiting case of the multiple-time-step technique, where the MM region is oversampled with respect to the QM region by performing multiple time steps in the MM part per QM time step. Oversampling the MM environment generates an effective free-energy surface, on which the QM degrees of freedom move. The adiabatic and multiple-time-step schemes have been implemented within QM/MM CP-MD by Woo et al. [226, 227]. They are, however, generally applicable to QM/MM MD approaches.

Because of the considerable computational cost of full QM/MM MD simulations, semi-empirical QM methods are still a popular choice in this context. Several approaches have been suggested to improve the accuracy of the results. These include tuning the semi-empirical Hamiltonian with specific reaction parameters (SRP) [308–311]; adding a valence-bond correction term [312, 313]; and correcting the potential of mean force obtained at a lower QM level with an interpolative scheme based on few higher-level calculations [314–316].

4.3.2

QM/MM Free-Energy Perturbation

A QM/MM free-energy perturbation (FEP) scheme to calculate free-energy differences along a predefined reaction path has been proposed by Yang and co-workers [273]; Rod and Ryde recently presented a similar formulation [317, 318], dubbed QTCP (quantum-mechanical thermodynamic-cycle perturbation). The basic idea is to sample only the MM degrees of freedom, thus significantly reducing the computational effort compared to full QM/MM sampling.

One determines first a series of QM/MM-optimized structures along the chosen reaction path. The Helmholtz free-energy difference between two states (i.e., points on the path) a and b is then written as:

$$\Delta A_{\text{QM/MM}}^{a \rightarrow b} = \Delta A_{\text{QM}}^{a \rightarrow b} + \Delta A_{\text{env}}^{a \rightarrow b}. \quad (19)$$

The contribution from the QM region ($\Delta A_{\text{QM}}^{a \rightarrow b}$) is expressed as the difference of the QM energies (ΔE_{QM}) plus zero-point energy ($\Delta E_{\text{QM}}^{\text{ZPE}}$), finite-temperature ($\Delta U_{\text{QM}}^{\text{th}}$), and entropic (ΔS_{QM}) contributions obtained within the harmonic approximation [319]:

$$\Delta A_{\text{QM}}^{a \rightarrow b} = \Delta E_{\text{QM}} + \Delta E_{\text{QM}}^{\text{ZPE}} + \Delta U_{\text{QM}}^{\text{th}} - T \Delta S_{\text{QM}}. \quad (20)$$

The harmonic frequencies for the QM part are calculated at the end points of the path (assuming that they are stationary points), in the presence of the fixed MM region. Note that, for N QM degrees of freedom, all N ($N - 1$ in the case of a transition state) vibrational degrees of freedom are considered in the calculation of the partition function; there are no rotational and translational contributions.

The free-energy contribution of the environment, $\Delta A_{\text{env}}^{a \rightarrow b}$, is obtained from Zwanzig's perturbation formula [320] (which, despite its name, has nothing to do with a perturbational treatment in the usual sense). It incorporates the free-energy contributions from the MM region and the QM–MM coupling. One samples the MM degrees of freedom in state a and considers

the perturbation due to the QM region of state b :

$$\Delta A_{\text{env}}^{a \rightarrow b} = -RT \ln \left\langle \exp \left\{ -\Delta E_{\text{pert}}^{a \rightarrow b} / (k_B T) \right\} \right\rangle_{\text{MM}}^a, \quad (21)$$

where

$$\Delta E_{\text{pert}}^{a \rightarrow b} = E_{\text{QM-MM}} \left(\mathbf{R}_{\text{MM}}(t); \mathbf{R}_{\text{QM}}^b \right) - E_{\text{QM-MM}} \left(\mathbf{R}_{\text{MM}}(t); \mathbf{R}_{\text{QM}}^a \right). \quad (22)$$

Here, R is the universal gas constant, k_B is Boltzmann's constant, and the QM-MM coupling term, $E_{\text{QM-MM}}$, is defined as in Eq. 8. $\langle \cdot \rangle_{\text{MM}}^a$ designates the canonical ensemble average obtained by sampling over only the MM degrees of freedom in presence of the QM atoms fixed in state a ; that is, the forces governing the dynamics are those of state a . In each MD step, that is, for each configuration of the MM atoms $\mathbf{R}_{\text{MM}}(t)$, the QM-MM coupling energy is evaluated once with the QM atoms in state b ($\mathbf{R}_{\text{QM}}^b$) and once with the QM atoms in state a ($\mathbf{R}_{\text{QM}}^a$). Although the free-energy contribution due to the MM degrees of freedom is incorporated in ΔA_{env} , the MM energy, E_{MM} , is not explicit in ΔE_{pert} ; $E_{\text{MM}}(\mathbf{R}_{\text{MM}})$ does not depend on the QM coordinates and therefore cancels out.

The sampling could also be performed by MC, rather than MD, methods, which is, however, less common in practice. If one samples from the (isobaric-isothermal) NPT , instead of the (canonical) NVT , ensemble, one obtains a Gibbs, rather than a Helmholtz, free energy. The difference is negligible for most practical purposes.

To calculate $E_{\text{QM-MM}}$ and the forces on the MM atoms, Yang and co-workers proposed use of an ESP charge model of the QM density, as they do in their QM/MM optimization scheme (see Sect. 4.2.2) [273]. This avoids the costly evaluation of the electrostatic QM-MM interaction at the QM level, but approximates the continuous density by point charges and implies a frozen density (i.e., the density cannot adapt to the changing MM environment). We have recently confirmed that the use of the ESP charge model is a justified approximation and that the QM/MM FEP approach yields results comparable to thermodynamic integration or umbrella sampling at a fraction of the cost [321]. The most time-consuming parts of the procedure are the optimizations to define the reaction path.

The reaction path used in QM/MM FEP has usually been defined in terms of a preconceived reaction coordinate. However, one could also use a minimum-energy path (MEP) determined by a reaction-path optimization method (see Sect. 4.2.3).

QM/MM FEP (sometimes also called QM/MM-FE) is related to older techniques, with which it shares the common feature that sampling of the QM degrees of freedom is avoided. In the mid-1980s, Jorgensen and co-workers proposed a scheme for calculating the reaction free energy in solution along a reaction coordinate defined by a series of structures optimized at the (ab

initio) QM level in the gas phase [322, 323]. They then derived classical potential terms describing the solute–solvent interactions at each point of the path and calculated the free energy along the reaction coordinate by MC umbrella sampling. Similar techniques were used for reactions in enzymes by Merz and co-workers (thermodynamic integration with MD sampling) [324] and Kollman and co-workers (FEP with MD sampling) [325]. In all these QM free-energy (QM-FE) approaches, the reaction path is determined in the gas phase and the QM and MM potentials are not coupled. This amounts to assuming that the influence of the environment on the reaction path as well as polarization effects are negligible; both approximations are absent in QM/MM FEP.

To calculate reaction and activation free energies in enzymes, Warshel and co-workers [326–329] developed a scheme in which the sampling is performed on an EVB reference potential fitted to *ab initio* data. A linear-response approximation is then applied to evaluate the free energy of transfer from the EVB to the *ab initio* surface.

Different formulations have also been proposed for solving the problems that arise when performing “alchemical” FEP simulations with a QM/MM potential [330–333].

4.3.3

Thermodynamic Integration

In standard thermodynamic-integration (TI) simulations, the system is taken along a reaction coordinate ξ , which parameterizes the Hamiltonian, and the mean generalized force acting on this coordinate is integrated to give the free-energy difference:

$$\Delta A^{a \rightarrow b} = \int_a^b d\xi \frac{dA(\xi)}{d\xi} = \int_a^b d\xi \left\langle \frac{\partial \mathcal{H}(\xi)}{\partial \xi} \right\rangle_{\xi}, \quad (23)$$

where $\langle \cdot \rangle_{\xi}$ denotes the canonical ensemble average at fixed ξ . In the pointwise TI approach, one constrains ξ to a series of values, runs *NVT* MD simulations to sample the average force at each location, and integrates these average forces numerically. (Note that if ξ is coordinate-dependent, i.e., a structural or geometric constraint, the correct general expression for $\langle \partial \mathcal{H}(\xi) / \partial \xi \rangle_{\xi}$ must correct for the fact that in a constrained simulation the conjugate momentum of ξ is also constrained [334–342].)

Alternatively, one can move the reaction coordinate continuously from a to b over a time interval τ :

$$\Delta A^{a \rightarrow b} = \int_0^{\tau} dt \frac{\partial \mathcal{H}(\xi)}{\partial \xi} \dot{\xi}, \quad (24)$$

with $\dot{\xi} = \partial\xi/\partial t$. If a constraint is used, this is commonly referred to as “slow-growth” or “adiabatic switching” TI. If one uses a guiding potential, that is, a moving restraint, it is known as “steered MD” (SMD) [343, 344]. A crucial assumption underlying all these methods is that the system is always in equilibrium. The switching process from a to b is assumed to be reversible, that is, infinitely slow.

In real simulations, however, ξ is changed at a finite rate. It then follows from the second law of thermodynamics that the average work $\overline{W}_\tau$ expended in switching from a to b in a finite time τ is an upper bound to the free energy, $\Delta A^{a \rightarrow b} \leq \overline{W}_\tau$; $\Delta A^{a \rightarrow b} = \overline{W}_\tau$ only for $\tau \rightarrow \infty$.

However, Jarzynski discovered a remarkable equality [345–350, 743], valid under very general conditions, that connects the irreversible work for arbitrary switching times τ to the equilibrium free energy:

$$\exp\{-\Delta A/(k_B T)\} = \langle \exp\{-W_\tau/(k_B T)\} \rangle_0, \quad (25)$$

where $\langle \cdot \rangle_0$ denotes an average of W_τ values over initial states taken from an equilibrium (canonical) ensemble.

This relation has become the foundation for “fast-growth” [351] or “multiple-steering” [349, 352, 353] simulations, in which irreversible-work values from deliberately non-adiabatic switching processes are accumulated and exponentially averaged according to Eq. 25; it has also been applied in the analysis of single-molecule pulling experiments [354–357]. Jarzynski’s equality immediately raises questions about the dependence of the free-energy error on the switching time and about the efficiency of performing many fast switches versus fewer slow ones. These issues have been amply discussed in the literature, comparing fast-growth to other free-energy simulations [358–368]. Applications of the fast-growth method in the context of enzymatic QM(DFT)/MM simulations are reported in [369, 744, 755].

4.3.4

Transition-Path Sampling

Developed by Chandler and collaborators [370–380], transition-path sampling (TPS) uses Monte Carlo importance sampling in the space of trajectories connecting the reactant with the product basin and yields an ensemble of reactive trajectories (see the contribution by C. Dellago and P.G. Bolhuis in this volume). It does not involve any preconceived knowledge about the reaction path or the transition state. All that is required are low-dimensional order parameters (e.g., combinations of bond lengths) that are able to unambiguously separate reactant from product configurations. TPS has recently been applied to an enzymatic reaction [381]; semi-empirical QM/MM MD was used to generate the trajectories.

4.3.5

Metadynamics

The metadynamics approach by Laio, Parrinello, and co-workers [382–390] is capable of exploring the free-energy surface without prior information about the location of minima or transition states. It is related to earlier techniques such as the local-elevation method [391], conformational flooding [392, 393], or the flat-histogram method for MC [394–396]. Metadynamics follows the dynamics in the space of a set of collective coordinates that characterize the process of interest, and is coupled to the real dynamics of the system via a history-dependent bias potential. Controlled by the metadynamics, the bias potential drives the system away from regions of the free-energy surface it has already visited by “filling them up”. The bias potential is constructed from Gaussians of a specified width that are deposited in the space of the collective coordinates at a chosen frequency. Once the free-energy landscape has been filled to a certain level (related to the temperature) with Gaussians, their sum represents a cast of the free-energy surface, with all minima and transition regions. This technique has recently been applied to an enzyme reaction within Car–Parrinello QM/MM MD [397].

4.3.6

Adiabatic Dynamics

The phase-space sampling can be enhanced in constant-temperature MD by coupling selected degrees of freedom to a separate thermostat and keeping them at a higher temperature than the remainder of the system. To minimize the heat flow from the hot degrees of freedom into the cooler surroundings, the masses of the hot atoms are scaled up, thus creating an adiabatic separation between the hot, slow and the cool, fast degrees of freedom. Such techniques have been presented by Tuckerman and co-workers (adiabatic free-energy dynamics, AFED) [398, 399] and VandeVondele and Röthlisberger (canonical adiabatic free-energy sampling, CAFES) [400]. The dynamics of the slow degrees of freedom is effectively performed on the free-energy surface generated by the environment. CAFES has been used in a Car–Parrinello QM/MM MD simulation [401].

4.3.7

QM/MM Reaction-Path Potential

Lu and Yang [402] extended the idea of the reaction-path Hamiltonian by Miller et al. [403] to large systems described at the QM/MM level. Using the energies, vibrational frequencies, and electronic response properties of the QM region along a QM/MM reaction path (e.g., the minimum-energy path), they constructed a harmonic reaction-path potential. It provides an analytical

expression for the QM/MM potential energy along the path, accounting for the coupled dynamics of QM and MM parts. This potential can subsequently be used for sampling, for instance, to perform thermodynamic integration with constrained MD.

5

Practical Aspects of Biomolecular Reaction Modelling

When performing a QM/MM study of a biomolecular reaction, particularly an enzymatic reaction, one generally has to invest a considerable amount of work into the setup and preparation of the system prior to the actual QM/MM calculations. We provide here a brief list of possible issues that need to be addressed in this process. Many of them are not exclusive to the QM/MM approach, but arise also in classical MD studies:

- *Developing MM parameters:* As it is useful to be able to perform classical MD simulations during the preparation stage (as described below), one needs a set of MM parameters for the entire system, including the prospective QM part. While one can restrain, e.g., substrates or cofactors to their experimental position, thus avoiding the need for accurate bonded parameters, it is highly desirable to develop at least reasonable non-bonded parameters, in particular atomic charges. They largely control the arrangement of polar or charged residues and water molecules around these moieties.

Different biomolecular force fields follow different procedures to derive partial charges, which may involve QM calculations on the molecule in question or representative fragments thereof. The effort required to obtain or generate MM parameters can be a decisive factor for the choice of force field.

- *Structure validation:* The starting point of a biomolecular QM/MM study is almost always an experimental structure, most commonly determined by single-crystal X-ray diffraction. To check the structure for inconsistencies [404], it is generally advisable to use one or several of the existing validation tools, for example WHAT CHECK [405, 719]; a web interface to several such tools is available [720]. One should also bear in mind that the structure is usually least reliable in and around the active site, where non-protein molecules (substrates, cofactors) are present [246].

It is often not possible to distinguish between C, N, and O based on crystallographic data alone. A common error are therefore flipped side-chain amide groups of asparagine and glutamine as well as flipped histidine imidazole rings. These are detected by, e.g., WHAT CHECK [405, 719] and REDUCE [406, 407].

Besides the “technical” correctness of a specific structure, there is the more fundamental question of the sensitivity of the computational results

to the initial crystal structure [408]. Structures of the same enzyme family or of the same enzyme crystallized with different inhibitors or under different conditions can differ in subtle details, which may, however, impact the results. There is no guarantee that the crystal structure actually corresponds to the productive form of the enzyme.

- *Adding hydrogen atoms:* The positions of the hydrogen atoms are rarely resolved in X-ray structures, and generally no hydrogen coordinates are provided. For the simulations, and certainly all hydrogen atoms in the QM region as well as the polar and aromatic MM hydrogen atoms need to be added; if an all-atom force field is used, all hydrogen positions are required. For aliphatic and aromatic hydrogen atoms, this is a routine procedure; most biomolecular simulation packages are capable of reliably placing these hydrogen atoms according to standard bond distances and angles.

The situation is less clear for polar hydrogen atoms, especially if they cap a rotatable bond, e.g., the serine side-chain hydroxy group. The orientation of these bonds, and hence the position of the corresponding hydrogen atoms, is controlled by the local hydrogen-bond network as well as steric factors. Several programs and algorithms are available (either as part of a larger modelling package or as stand-alone programs) that attempt to determine the optimum position for these hydrogen atoms. Examples include CHARMM [180, 704], WHAT IF [409, 721], or REDUCE [406, 407]; an assessment of such procedures has recently appeared [410].

Most difficult are the acidic hydrogen atoms, whose presence or absence is ambiguous. Depending on the local pK_a value, ionizable groups can be protonated or not. Most prominent are the imidazole side chain of histidine and the side-chain carboxylates of aspartate and glutamate, but it can also be necessary to consider the phenolic hydroxy group of tyrosine, the thiol of cysteine, as well as the lysine ammonium and the arginine guanidino group. The protonation state controls the availability of acidic or basic sites, and it directly influences the charge and thus the long-range electrostatic field. The decision where to protonate is therefore a rather delicate one, which may well have mechanistic implications.

The most rigorous solution, calculating the local pK_a values using Poisson–Boltzmann or even QM/MM methods [331, 411], is not routinely used because of the computational effort involved. The other extreme, namely to choose the protonation states of the free amino acids in water at pH 7, is not satisfactory. A compromise are methods that assign the protonation state based on the local hydrogen-bonding network and structural considerations; for instance WHAT IF [409, 721] or REDUCE [406, 407]. An alternative is the calculation of local pK_a values with a fast increment method, such as PROPKA [412, 722]. One should bear in mind that these empirical approaches neglect the non-protein part of the enzyme and are therefore less reliable in and around the active site.

The protonation states depend of course on the pH chosen for the simulation. It should be noted that the optimal pH, where the enzyme activity is highest, is often quite different from the pH at which the structure was determined because a pH close to the isoelectric point tends to be favourable for crystallization. The pH for kinetic measurements, from which, e.g., rate constants or activation barriers are derived, can again differ from the optimum pH.

Through the hydrogen-bonding network, the assignment of protonation states is linked to the amide and histidine flips (vide supra) and to the hydration of the structure (vide infra). These three steps are therefore interdependent and may require an iterative treatment. One should also note that the results of different empirical algorithms, e.g., to detect flipped residues or to assign protonation states, may well be at variance. In these cases, the only practical solution is to visually inspect the structure and to take a decision based on chemical insight.

- *Choice of simulation system:* Apart from the definition of the QM region, which should incorporate the “electronically active” part of the structure, one has also to decide on how much of the environment is to be included in the simulation and in which way. The cleanest way is certainly to include the whole protein plus a hydration shell (see below). The next choice is then the size of the active region, where the atoms are free to move during optimization or MD, while the remaining parts are fixed or positionally restrained. In MD simulations, the number of degrees of freedom is not critical. Optimizations, however, tend to become problematic if more than a few thousand (about 2000) atoms are included. Not only can convergence become an issue, but it is also more difficult to ensure that all the many environment degrees of freedom remain consistent among changing active site configurations, e.g., along a reaction path. For instance, if a hydrogen bond in the environment rearranges from one step to the next, the continuity of the energy profile is broken. A reasonable selection for the active region therefore includes all residues within, e.g., 10 Å of the QM part.
- *Solvation/hydration:* Only water molecules occupying a well-defined position are crystallographically resolved; mobile ones are not visible. At least those parts of the system that are allowed to move during the MD or in optimizations (plus a buffer region around them) must therefore be hydrated with additional water molecules. There are two main options: one can either place the whole system in a box of water under periodic boundary conditions; or one superimposes a sufficiently sized sphere of water molecules (cut out from an MD simulation of liquid water) onto the active region of the structure. In the latter case, a spherical boundary potential prevents the water molecules from diffusing away and simulates the interface to the bulk. In either procedure, water molecules too close to an existing atom are deleted, and the remaining ones are subjected to repeated minimization and MD runs.

This protocol should ensure that the surface of the protein as well as cavities and channels connected to the surface are adequately hydrated. However, there may exist small internal cavities without connection to the surface that cannot be hydrated in this manner. Apart from inserting water molecules manually, one can use programs like DOWSER [413, 723] that identify such cavities and attempt to determine whether water molecules should be placed there.

- *Classical MD simulations:* Once the system setup has been completed, it is generally advisable to run classical MD simulations, either of the whole system (under periodic boundary conditions) or of a suitably chosen active region. The purpose of these MD runs is twofold: (i) By gradually releasing the positional restraints applied to the protein part during the hydration phase and performing free MD, one can validate the setup. Significant differences between the X-ray structure and the average positions hint at problems with the protonation or hydration procedures. (ii) Snapshots from the MD simulation can serve as initial structures for subsequent optimizations.

At this stage, one can introduce or modify moieties not present in the experimental structure. For instance, an inhibitor bound in the active site is replaced by the substrate; a cofactor is changed into its chemically active form; or an amino acid is exchanged to generate a mutant. Depending on the manipulations, it may be appropriate to check the protonation states and/or rehydrate and to let the system adapt to the changes with another MD run.

- *Counter ions, charge neutralization:* There is no general consensus on whether it is necessary or desirable to neutralize the total charge of the system by the addition of counter ions. For charged surface residues (e.g., carboxylate or ammonium side chains), there will always be a corresponding counter ion (typically sodium or chloride, respectively) nearby in solution. One can thus argue that counter ions should also be present in the simulation to balance and screen these charges. It has also been suggested that they be neutralized by (de)protonation, especially if they are fixed and outside of the hydration sphere.

Even if one chooses to compensate for the surface charges, there normally remains a net charge due to buried charged groups. To produce an overall neutral system, one may add additional counter ions in the hydrated part (raising the question of where exactly to place them) or remove some of the “surface” counter ions added previously (raising the question of which ones to eliminate).

Because of these ambiguities, a common choice is to leave the total charge as it is after assigning the protonation states according to the chosen pH. Exceptions are simulations of the relative stability of differently charged QM regions (e.g., redox processes). In this case, the environment should be neutral as a non-zero net charge would create an electrostatic field, thus artificially stabilizing one charge state over the other [330, 414, 415].

6

Interpreting the Results: Understanding Enzyme Catalysis

The results from QM/MM calculations and simulations (potential and free-energy profiles, structures, dynamics, etc.) are, of course, on the first level amenable to all the same interpretations as those obtained with any other potential-energy method. However, because of the inherent separation between the QM reactive part and the MM environment, and the possibility of easily dissecting the total energy according to various decomposition schemes (e.g., QM/MM/QM-MM, electrostatic/steric, by residue), QM/MM methods lend themselves to investigations into the sources of the catalytic power of biocatalysts. Simulations and theoretical tools provide the means for a detailed understanding of how enzymes achieve their sometimes spectacular rate enhancements, in particular of the factors contributing to the barrier lowering and the role of enzyme structure and its flexibility.

Since the proposal by Linus Pauling [416] that enzymes work by binding the transition state more strongly than the substrate, thus lowering the activation barrier, a variety of concepts and ideas have been advanced to explain the origin of enzymatic catalysis; we refer to several recent reviews on this topic [233, 234, 417–421, 745–750], including a thematic issue of *Chemical Reviews* [751]. (Note that these articles usually deal not only with interpretive concepts but also with computational methods for the investigation of enzyme catalysis, including, but not restricted to, QM/MM approaches. They thus complement the reviews cited in Sect. 1.)

No single model should be expected to encompass in its entirety a phenomenon as complex as enzyme catalysis. Some proposals are more qualitative notions rather than quantitative models, and many are subject to critical, sometimes controversial, discussion in the literature. One difficulty is the loose definition of certain terms and concepts. To mention just one basic example, there is no general agreement on the reference system to which the enzymatic reaction should be compared and, therefore, how precisely the rate acceleration due to the enzyme (its catalytic effect) is to be defined. As advocated by Warshel [746], the most natural choice seems to be the corresponding reaction in water that proceeds by the same mechanism. We give here a brief, partial list of catalytic models and concepts:

- *Transition-state stabilization*: The environment of the active site is organized such that the TS is preferentially stabilized. In particular, hydrogen-bonding interactions and the electrostatic field in the active site are considered [233, 234, 417, 746].
- *Substrate destabilization*: Under this category, a number of effects related to the preparation or preorganization of the substrate can be subsumed; including steric confinement [422–426], the entatic state [427–430], induced fit [431], orbital steering [432, 433], the spatial-temporal hy-

pothesis [434], cratic free energy [435], and near-attack conformations (NAC) [436–447].

- *Desolvation*: The substrate is less well solvated in the active site than in solution [448, 449].
- *Entropic effects*: The motion of the substrate is restricted in the active site (“entropic trap”) [325, 435, 450–452].
- *Low-barrier hydrogen bonds (LBHB)*: Special hydrogen bonds with a significant covalent contribution stabilize the TS [453–458].
- *Dynamic effects*: Specific protein motions or vibrations couple to the substrate to lower the barrier [417, 459–463, 748].
- *Nuclear quantum effects*: The enzyme increases the transmission coefficient through vibrationally enhanced tunnelling and other nuclear quantum effects [419, 421, 463–467, 745, 750].
- *Chemical effects*: All chemical factors in which the reaction in solution and in the enzyme differ can be collected under this heading; for instance, changes in the mechanism or the participation of a generalized acid or base, of a cofactor, or a metal centre. The covalent hypothesis proposes that these effects are at work in the most proficient enzymes [468]. One may argue, however, that they are not catalytic effects proper and should be considered separately.

From the current literature, electrostatic transition-state stabilization appears to emerge as the dominant factor in the actual chemical step of enzymatically catalysed transformations. The electrostatic basis of enzyme catalysis has been stressed in a recent review that evaluates a number of concepts and proposals in great detail [746].

Besides the plethora of catalytic proposals, one can more generally characterize the role of the enzyme and the microscopic steps involved in an enzymatic reaction [24]:

- The enzyme is prepared for the reaction. This can include changes in the tertiary or quaternary structure, binding a cofactor, and possibly activating the cofactor. (The latter constitutes a, possibly catalytic, reaction in its own right but we consider it here as a preparatory step preceding the reaction step under study.)
- The substrate binds to the enzyme. While this may be a simple association, it can also entail a more complicated mechanism, involving enzyme structural changes or guiding of the substrate to the active site by specific interactions. In this case, the binding step can contribute to the selectivity of the enzymatic reaction.
- The substrate is prepared in the active site, both chemically and structurally. Chemical preparation involves, e.g., (de)protonating the substrate to activate it towards the reaction. Structural preparation (or preorganization) includes both enthalpic (adopting a specific conformation different from the one in solution) and entropic (loss of translational and rotational degrees of freedom) aspects.

- The actual chemical reaction takes place in the active site. The chemical step is in the focus of most of the models listed above that attempt to identify and explain the specific catalytic effect of the enzyme.

We also mention some analysis procedures that have proven useful in obtaining insight into enzymatic reactions:

- *Energy decomposition by residue*: To probe the effect of a selected MM residue on an energy difference (e.g., the activation barrier), certain interactions of this residue are switched off, and the energy difference is re-evaluated. Typically, this procedure is applied to the electrostatic interaction by deleting the residue's point charges. A common way is to remove the electrostatic interactions sequentially residue-by-residue in order of decreasing distance from the active site and to plot the energy difference as a function of the distance. Also known as "perturbation analysis", this approach was established early on [469] and has remained popular since; see, e.g., [310, 470–474].
- *Decomposition of the electrostatic energy*: The electrostatic QM–MM interaction energy can be decomposed into the permanent interaction energy (the interaction between the environment and the unpolarized QM density) and the polarization energy. The latter can be further divided into the polarization-stabilization and the electronic-distortion terms. Such a dissection gives insight into the role of electronic polarization; see, e.g., [5, 294, 475].
- *Differential transition-state stabilization (DTSS)*: This analysis focuses on the specific barrier-lowering effect of the enzyme relative to the gas phase, considering the substrate(s) in their active-site reactant and TS structures [476, 477]. The DTSS can be further decomposed into long-range multipole, short-range penetration, exchange, delocalization, and correlation components; the multipole and penetration terms together form the electrostatic contribution. The procedure can be expanded by performing the analysis on a per-residue basis.
- *General energy decompositions*: A generic energy decomposition scheme can be applied to dissect the contributions related to changing the environment from a reference (typically aqueous solution) to the enzyme [301, 420, 478]. For the substrates, the process of transferring them from solution into the enzyme can be subdivided into three steps: (i) The free reactants are individually distorted in solution from their minimum structure into the conformation that they adopt in the binding site. This preparation energy is strictly positive. (ii) The distorted reactants, still in solution, are brought into the relative position and orientation they adopt in the enzyme; this complexation energy can be either stabilizing or destabilizing. (iii) The distorted and oriented substrates are placed into the active site. The corresponding energy is the interaction energy of the prepared substrate complex.

The analogous procedure can be applied to the transition state, affording a decomposition of the activation barrier. The energy difference in solution between the prepared substrate complex and the enzyme-optimized TS structure is compared to the full QM/MM barrier, providing a measure for the specific catalytic effect of the enzyme.

An alternative approach follows the inherent partitioning of the QM/MM energy. Any energy difference, such as the activation barrier, can be dissected into QM, QM-MM, and pure MM contributions. The QM-MM component is further split into QM-MM electrostatic, QM-MM van der Waals, and QM-MM bonded terms. The MM component may be likewise divided into MM electrostatic, MM van der Waals, and MM bonded terms.

7

Survey of Biomolecular QM/MM Studies

Since the pioneering work of Warshel and Levitt in 1976 [1], QM/MM methods have become increasingly popular in the area of biomolecular modelling. The large majority of applications have been concerned with enzymatic reactions. Obviously, QM/MM approaches are the method of choice for the investigation of chemical transformations, while MM methods serve well in purely structural studies. The focus on enzymes is certainly due to the ubiquity and importance of these biocatalysts, but also to the relatively wide availability of experimental, especially structural and kinetic, data and of specialized force fields (which is, of course, in turn related to the relevance of enzymes). Other biomolecules (nucleic acids, carbohydrates, lipids) have found comparably less attention.

We provide here a survey of biomolecular QM/MM studies in tabular form, considering contributions that have appeared between 2000 and April 2006 (see Tables 1–9). References to earlier work can be found in the reviews cited in Sect. 1. The number of publications has been steadily increasing since the mid-1990s, but especially from 2003 onwards. The list is extensive, but certainly not exhaustive. Apart from inadvertent omissions, a contribution is likely to be missing if none of the pertinent keywords (QM/MM, combined quantum mechanics/molecular mechanics, etc.) is present in the title or abstract. The entries are grouped by the type of biomolecule investigated; the sorting of enzymes follows the IUBMB (International Union of Biochemistry and Molecular Biology) enzyme classification (EC nomenclature [724]). Note that the reaction investigated does not necessarily correspond to the enzyme's main function, which determines its classification. To make the tables less congested, the entries in the columns "biomolecule", "process studied", "QM level", and "MM level" are not repeated in successive rows for the same

Table 1 Oxidoreductases (EC 1), O₂- and electron-transport proteins

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
<i>Flavin-dependent oxidoreductases</i>						
<i>p</i> -Hydroxybenzoate hydroxylase	OH transfer	AM1	CHARMM	Opt	Influence of substituted substrates, flavin cofactors	[479]
		HF, DFT, MP2	GROMOS	Opt	Comparison of free-energy methods	[480]
		AM1		TI(MD)		[301]
				Opt, QM/MM FEP, US(MD)		[321]
Phenol hydroxylase	OH transfer			Opt, TI(MD)	Influence of substrate protonation state	[481]
		AM1	CHARMM	Opt	Influence of substrate deprotonation, substituted substrates	[482]
		AM1	CHARMM	US(MD), rate constants, KIE	VTST with tunnelling	[483]
Acyl-CoA dehydrogenase	H ⁺ , H ⁻ transfer	AM1	CHARMM	Rate constants, KIE	VTST with tunnelling	[484]
Liver alcohol dehydrogenase	H ⁻ transfer	AM1	CHARMM	Opt, reaction paths, rate constants, KIE	VTST with tunnelling	[485]
		AM1, HF	AMBER	Opt, US(MD)	Influence of specific residues, mutations	[472]
Cholesterol oxidase	Electronic excitations FAD oxidation FAD reduction	SCC-DFTB, DFT, HF	CHARMM	Opt, US(MD)	Redox potentials Redox potential from “dual-topology-single-coordinate” FEP	[486]
		CIS	AMBER	Single points		[487]
		SCC-DFTB	CHARMM	FEP(MD) FEP		[330]

Table 1 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
<i>NAD(P)-dependent oxidoreductases</i> Class 3 aldehyde dehydrogenase	H ⁺ transfer	PM3	OPLS-AA	US(MD)		[488]
	Thioacetal formation		CHARMM, OPLS-AA	Opt, US(MD)		[489]
Lactate dehydrogenase	H ⁺ , H ⁻ transfer	AM1	CHARMM	Opt		[490]
				TPS		[381]
		AM1, MP2		Opt	Influence of titratable residues	[491]
		AM1		Opt	Corrections for AM1, adjusted vdW parameters	[492]
Dihydrofolate reductase	Substrate binding, deprotonation	PM3	AMBER	FEP(MD)	Comparison to ab initio/continuum solvent; role of LBHB	[493]
	Inhibitor binding	AM1		FEP(MD)		[494]
			AMBER, Tripos	Opt		[495]
	H ⁻ transfer	PM3	AMBER	Opt	Comparison to semi-empirical all-QM	[496]
				Constrained MD		[497]
				Opt	Comparison to semi-empirical all-QM	[498]
		AM1, PM3		FEP(MD)	Overlapping mechanical embedding	[499]
		AM1	CHARMM	MD	Analysis of substrate polarization	[475]
				US(MD), rate constants, KIE	VTST with tunnelling	[500]
	H ⁺ transfer	PM3		Opt	Influence of mutations	[501]
			AMBER	TI(MD)	QM alchemical perturbation	[502]
		AM1	CHARMM	Opt		[503]

Table 1 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Acetohydroxy acid isomeroreductase	Alkyl migration	AM1	OPLS	US(MD)		[504]
<i>TTQ-dependent oxidoreductases</i>						
Methylamine dehydrogenase	H ⁺ transfer	PM3	CHARMM AMBER	KIE Opt, rate constants, KIE Opt, rate constants, KIE Opt, reaction paths, rate constants, KIE	VTST with tunnelling VTST with tunnelling VTST with tunnelling VTST with tunnelling	[505] [506] [507] [508]
<i>Non-haem-iron oxidoreductases</i>						
Methane monooxygenase	Structure O ₂ binding O ₂ activation, OH transfer	DFT	AMBER OPLS-AA	Opt Opt Opt		[509] [510] [511]
Phe hydroxylase	H abstraction			Opt, KIE	Different substrates	[512]
Ribonucleotide reductase	Hydroxylation	DFT	AMBER	Opt	Comparison of mechanisms	[513]
HIF-1 α asparaginyl hydroxylase	Structure O ₂ activation	DFT CASSCF	AMBER EFP (AMBER)	Opt Opt	Electronic structure	[509] [514]
<i>Haem-dependent oxidoreductases</i>						
Cytochrome c peroxidase	Electronic structure of ferryl intermediate	DFT	CHARMM	Opt		[515]

Table 1 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Cytochrome P450	Spin states, structure of haem	DFT	CHARMM	Opt		[516]
	Substrate binding		AMBER	Opt		[517]
	Characterization of ferryl intermediate		CHARMM	Opt	Comparison between human isoforms	[518]
				Opt	Influence of mutation	[519]
	EPR, Mössbauer parameters of ferryl intermediate			Opt		[520]
	Formation of ferryl intermediate			Opt		[521]
	H abstraction, OH transfer			Opt		[752]
	H abstraction		CHARMM, Opt	Opt		[522]
			OPLS-AA			[523]
			CHARMM	Opt		[753]
			CHARMM	Opt		[415]
	Electronic structure of intermediate, Mössbauer parameters					
Ascorbate peroxidase	Product release			Opt		[524]
	Substrate binding, O ₂ activation, hydroxylation		OPLS-AA	Opt		[525]
	Hydroxylation		CHARMM	Opt	Review	[526]
	Electronic structure of ferryl intermediate		CHARMM	Opt		[515]
Horseradish peroxidase	Electronic structure of ferryl intermediate	DFT	CHARMM	Opt		[527]
Haem oxygenase	Active-site structure	DFT	AMBER	Opt		[528]

Table 1 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Nitrite reductase	H ⁺ transfer	DFT	AMBER	Opt		[529]
NO synthase	Active-site structure			Opt		[530]
	Form of NHA intermediate	HF, DFT	AMBER	Opt		[531]
	NO binding	DFT		Opt	Influence of mutations, haem oxidation state	[532]
<i>Other haem proteins</i>						
Truncated haemoglobin N (trHbN)	NO oxidation to NO ₃ ⁻	DFT	AMBER	Opt		[533]
Various haem proteins	O ₂ binding	DFT	AMBER	Opt		[534]
Haemerythrin	O ₂ binding	DFT	OPLS-AA	Opt		[511, 535]
Myoglobin	Active-site structure with bound CO	DFT	CHARMM	CP-MD		[536]
	Vibrational analysis with bound CO		AMBER	Opt		[537]
	Ligand binding		CHARMM	CP-MD		[538]
	CO binding			Opt, MECP	Non-adiabatic TST	[539]
	CO dissociation		OPLS-AA	Opt	Analysis of protein structural changes	[540]
			CHARMM	Opt		[541]
<i>V-, Cu-dependent oxidoreductases</i>						
V haloperoxidase	Active-site protonation state, formation of hydroperoxo intermediate	DFT	AMBER	CP-MD, metadynamics		[397]
V-dependent chloroperoxidase	Active-site structure	DFT	OPLS-AA	Opt	Protonation state of active site	[542]

Table 1 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Multicopper oxidase	O ₂ activation	DFT	AMBER	Opt	Electronic structure of various protonation and oxidation states	[543]
Galactose oxidase	Oxidation, protonation, dehydrogenation	DFT	CHARMM	CP-MD	Comparison to mimetic	[544]
Dopamine β -mono-oxygenase	Active-site structure	DFT	AMBER	Opt	Comparison of different Cu-oxo species	[545]
<i>Other metalloproteins</i>	H abstraction			Opt		[546]
Rubredoxin	Active-site structure, redox potentials	DFT	AMBER	Opt	Influence of mutations	[547]
Due ferri 1	Active-site structure	DFT	AMBER	CP-MD	Zn analogue of a mimetic Fe protein	[548]
Blue copper proteins	Structure	DFT	AMBER	Opt		[549]
Plastocyanin	EPR parameters			Single points		[550]
<i>Other oxidoreductases</i>	Redox potential	DFT	UFF, AMBER	Opt		[551]
Mammalian glutathione peroxidase	Active-site structure	HF, DFT	AMBER	Opt		[552]
<i>Other proteins</i>						
Thioredoxin	Disulfide reduction	HF, DFT, MP2	UFF	Opt		[553]
Granulocyte colony-stimulating factor	Met oxidation by H ₂ O ₂	DFT	CHARMM	Opt, TI(MD)		[554]

Table 2 Transferases (EC 2)

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Glutathione <i>S</i> -trans-ferase	Nucleophilic epoxide opening	AM1	CHARMM	US(MD)		[555]
Catechol <i>O</i> -methyltrans-ferase	Methyl transfer	AM1	CHARMM	Opt, US(MD) Opt, KIE Opt, US(MD) Opt, MD QM/MM FEP Opt, QM/MM FEP	Corrections to PMF	[478] [556] [557] [558] [317, 318] [559]
Histone lysine methyltrans-ferase SET7/9	Methyl transfer	DFT DFT, MP2	OPLS-AA CHARMM AMBER			
Human purine nucleoside phosphorylase	Vibrations in bound substrate, intermediate	AM1	CHARMM	Opt, MD		[560]
Protein kinase, thymidine phosphorylase	Phosphorylation	HF, DFT	AMBER	Opt	Evaluation of semi-empirical QM	[561]
Pyrimidine nucleoside phosphorylase	Active-site structure	DFT	CHARMM	Opt		[562]
cAMP-dependent protein kinase	Serine phosphorylation	DFT, MP2 DFT	AMBER	Opt Opt	Effect of enzyme phosphorylation	[474] [563]

Table 2 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
HIV-1 reverse transcriptase	H ⁺ transfer, phosphorylation	AM1	CHARMM	Opt, US(MD)		[564]
Hypoxanthine-guanine-xanthine	H ⁺ transfer,	AM1, HF, DFT	OPLS-AA	Opt, US(MD)		[565]
phosphoribosyltransferase	phosphoryl transfer					
Citrate synthase	Enolization	HF, MP2	CHARMM	Opt		[566]
	Fluorescence quantum yield	INDO/S-CIS		MD		[567]
Fluorinase	S _N 2 displacement	DFT	CHARMM	Opt		[568]

Table 3 Hydrolases (EC 3)

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
<i>Peptidases and other amidases</i>						
Aminopeptidase	Peptide hydrolysis	AM1	Tripos	Opt		[569]
Leu aminopeptidase	Peptide hydrolysis	AM1	Tripos	Opt		[570]
Met aminopeptidase	Active-site protonation states	DFT	AMBER	CP-MD		[571]
Carboxypeptidase A	Inhibitor binding	AM1	AMBER, Tripos	Opt		[495]
α -Chymotrypsin	S _N 2 attack on reactive inhibitor	HF	AMBER	Opt		[572]
	Inhibitor binding, aziridinium formation	AM1, DFT		Opt		[573]
	Active-site structure	DFT	EFP	Opt, NMR shifts	Structural and spectroscopic properties of H-bonds	[574]
				Opt, NMR shifts, H/D fractionation	Structural and spectroscopic properties of H-bonds	[575]
Subtilisin	Structure of tetrahedral intermediate	PM3	AMBER	Single point		[576]
α -Lytic protease	Active-site structure	DFT	EFP	Opt, NMR shifts, H/D fractionation	Structural and spectroscopic properties of H-bonds	[575]
Thermolysin	Peptide hydrolysis	AM1	AMBER	Opt		[577]
β -Trypsin	Inhibitor binding	AM1	AMBER, Tripos	Opt		[495]

Table 3 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Trypsin	Peptide hydrolysis	HF, MP2	EFP (OPLS-AA, AMBER)	Opt		[578]
	Acylation		AMBER	QM/MM FEP		[579]
	H ⁺ transfer	HF		Opt, QM/MM FEP	Influence of mutations	[580]
				Opt, QM/MM FEP, NMR parameters		[581]
Elastase	Substrate binding	AM1	CHARMM	Opt	QM/MM-PB/SA; different substrates	[582]
	Deacylation	HF	CHARMM	MD		[583]
	Inhibitor binding	HF, DFT		US(MD)		[584]
	Inhibitor binding, peptide hydrolysis	PM3	AMBER	Opt		[485]
Kumamolisin-As (sedolisin, serine-carboxyl peptidase)		PM3, DFT		Opt		[585]
	Peptide hydrolysis	SCC-DFTB	CHARMM	US(MD)		[586, 587]
HIV-1 protease	Peptide hydrolysis	DFT	AMBER	CP-MD, TI(MD), KIE		[401]
	Inhibitor binding	AM1	AMBER, Tripos	Opt		[495]
Caspases			CHARMM	MD		[588]
	Thioester hydrolysis	DFT	GROMOS	CP-MD		[589]
	Amide hydrolysis	DFT	AMBER	Opt		[590]
Peptide deformylase	N-Peptidyl formamide hydrolysis	DFT	OPLS-AA	Opt	Structure, protonation state of bound inhibitors	[591]

Table 3 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Matrix metalloproteinases	Structure, interaction with bound inhibitors	DFT	OPLS-AA	Opt	Part of a docking approach	[592]
β -Lactamase	Acylation	AM1	CHARMM	Opt		[593]
		PM3	AMBER	Opt		[594]
		AM1	CHARMM	Opt		[595]
Class A β -lactamase	Deacylation			Opt		[596]
	Acylation	AM1, DFT	CHARMM	Opt		[473]
		HF, MP2	AMBER	Opt		[597]
	Deacylation	DFT	CHARMM	Opt		[598]
Class C β -lactamase	Deacylation	AM1, DFT		Opt	Comparison to penicillin-binding protein	[599]
		DFT	OPLS-AA	Opt		
Metallo- β -lactamase Zn- β -lactamase	Substrate binding	DFT	AMBER	Opt		[600]
	Protonation states	SCC-DFTB	CHARMM	Opt	Structural influence of active-site protonation	[601]
Monozinc β -lactamase Monozinc CphA β -lactamase	Deacylation	DFT	AMBER	CP-MD		[602]
	Active-site structure, dynamics with bound substrate	SCC-DFTB	CHARMM	Opt, MD		[603]
Dinuclear zinc metallo- β -lactamase OXA-10 β -lactamase Penicillin-binding proteins	Active-site structure	PM3	AMBER	MD		[604]
	Lys carboxylation	DFT	AMBER	Opt		[605]
	Deacylation	DFT	OPLS-AA	Opt	Comparison to β -lactamase Bound substrate or acyl intermediate	[599]
	Active-site structure	AM1		MD		[606]

Table 3 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Human serum albumin	Lactam aminolysis	PM3	AMBER	Opt		[607]
Fatty acid amide hydrolase	Acylation	PM3, DFT	CHARMM	Opt		[608]
<i>Esterases</i>						
Acetylcholinesterase	Acylation	HF, DFT, MP2	AMBER	Opt		[609]
		HF, MP2	OPLS-AA	Opt		[248]
	Active-site structure	HF, DFT	AMBER, UFF	Opt		[610]
Human butyrylcholinesterase	Ester hydrolysis	HF, DFT, MP2	AMBER	Opt		[611, 612]
Lipase	Acylation	DFT	CHARMM	Opt		[613]
<i>Glycosylases</i>						
Uracil-DNA glycosylase	Oxocarbenium formation, H ₂ O attack	AM1, HF, DFT	CHARMM	Opt		[614]
	Oxocarbenium formation, hydrolysis	AM1		Opt	Mixed implicit/explicit solvent with charge scaling	[81]
	Substrate binding	SCC-DFTB	CHARMM	US(MD)	Alchemical transformation	[615]
8-Oxoguanine DNA glycosylase I	Structure of bound substrate	DFT	CHARMM	CP-MD		[616]
Bacillus 1,3-1,4- β -glucanase						
<i>Phosphatases</i>						
F ₁ -ATPase	Phosphate hydrolysis	HF, MP2	AMBER	Opt		[617]
		DFT		Opt		[618]
Myosin	Phosphate hydrolysis	HF, DFT	CHARMM	Opt		[619]
RNase A	Phosphate hydrolysis	AM1	CHARMM	Opt		[620]

Table 3 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Human H-Ras p21 ^{ras} – GAP-334	GTP hydrolysis	HF, MP2	EFP (AMBER)	Opt		[621]
Human H-Ras p21 ^{ras}	Vibrational spectrum, structure of bound GTP	HF	CHARMM	Opt		[622]
		DFT		CP-MD		[623]
<i>Other hydrolases</i>						
Soluble epoxide hydrolase	Phosphate hydrolysis	DFT	AMBER	CP-MD		[624]
Adenosine deaminase	Nucleophilic attack	HF, DFT	AMBER	Opt		[625, 626]
Cytidine deaminase	Inhibitor (de)hydration	SCC-DFTB	CHARMM	US(MD)		[627, 628]
4-Chlorobenzoyl-	S _N Ar displacement	PM3	CHARMM	US(MD)		[629–631]
CoA dehalogenase	S _N 2 displacement	AM1	CHARMM	Opt		[309]
Haloalkane dehalogenase		PM3		US(MD), reaction rates, KIE	VTST	[449]
	Ester hydrolysis	AM1	OPLS-AA	Opt, US(MD)		[632]
			CHARMM	US(MD), recrossing factor	Reactive-flux calculation	[633]
				Opt	Semi-empirical QM with valence-bond correction	[313]

Table 4 Lyases (EC 4)

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Enolase Methylglyoxal synthase	Enolization, dehydration	HF, DFT	AMBER	QM/MM-FEP	Comparison to triose-phosphate isomerase Alternative mechanisms VTST with tunnelling	[471] [634]
	Phosphate elimination	SCC-DFTB	CHARMM	Opt		
Macrophomate synthase Ornithine decarboxylase	C–C bond formation	PM3	OPLS-AA	FEP(MC)		[635]
	Decarboxylation	AM1	CHARMM	US(MD), rate constants, KIE		[636]
Orotidine 5'-monophosphate decarboxylase	H ⁺ transfer, decarboxylation	DFT	AMBER	CP-MD, fast-growth TI		[369]
	H ⁺ transfer	DFT	OPLS-AA	Opt		[637]
Carbonic anhydrase II	CO ₂ binding, hydration	PM3, DFT	CHARMM	Opt, MD	Collective reaction coordinate	[638]
	Vibrational frequencies	PM3		MD		[639]
	Long-range H ⁺ transfer	SCC-DFTB		Opt		[91]
Diol dehydratase	H ⁺ transfer	PM3		MD	B ₁₂ -dependent	[640]
	H abstraction,	DFT	AMBER	Opt		[641]
	OH transfer					

Table 5 Isomerases (EC 5)

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
4-Oxalocrotonate tautomerase	H ⁺ transfers	HF, DFT	AMBER	Opt, QM/MM FEP	Effect of mutations Different substrates	[642]
				Opt		[643]
				Reaction-path opt, QM/MM FEP		[644]
Mandelate racemase	H ⁺ transfers, epimerization	AM1, DFT PM3	CHARMM AMBER	Opt	Effect of mutations	[754]
				Opt, MD	Comparison of mechanisms for different substrates	[645]
				Opt		[646]
Glutamate racemase Triosephosphate isomerase	H ⁺ transfers H ⁺ transfers, enolization	AM1 DFT AM1, DFT AM1, DFT, MP2 SCC-DFTB	CHARMM CHARMM CHARMM	Opt, US(MD)	Alternative mechanisms	[647]
				Opt, MD		[648]
				Opt		[310]
				Opt	Evaluation of QM methods, size of QM region, size of active region, boundary schemes Comparison to methylglyoxal synthase	[311]
				Opt		[151]
				Opt		[634]
H ⁺ transfer	H ⁺ transfer	DFT AM1	OPLS-AA CHARMM	Opt	VTST; comparison to models; influence of specific vibrational modes VTST; role of tunnelling	[649]
				Rate constants		[650]
				Rate constants, KIE		[651]

Table 5 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Triosephosphate isomerase	H ⁺ transfer	HF	AMBER	Opt, recrossing factor	Reaction-path potential, reactive-flux calculation	[652]
Xylose isomerase	Deamidation	AM1	AMBER	Opt	VTST; evaluation of different boundary schemes	[653]
	1,2 H-shift	PM3, HF, DFT		Opt, KIE		[140]
Cyclophilin A	1,2 H-shift, H ⁺ transfer	PM3	CHARMM	US(MD), KIE	VTST with tunnelling	[654]
	Peptide <i>cis/trans</i> isomerization at Pro	SCC-DFTB	CHARMM	US(MD)		[655]
Chorismate mutase	Claisen rearrangement	DFT	AMBER	Opt	QM/MM implementation in the SIESTA real-space DFT program	[139]
		AM1	CHARMM	Opt		[656]
				Opt, US(MD)		[657]
		HF		Opt		[658]
		DFT	AMBER	Opt		[659]
		AM1	OPLS-AA	Opt, FEP(MC)	Influence of substrate conformation	[444]
		SCC-DFTB	CHARMM	US(MD)	Influence of specific residues on substrate conformations	[660]

Table 5 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Chorismate mutase	Claisen rearrangement	AM1		Opt, MD, KIE		[661]
		AM1, DFT		Opt		[662]
		AM1		FEP(MD)		[445]
		HF		Opt		[663]
		AM1		Opt, KIE, equilibrium isotope effect		[664]
		HF, MP2		Opt	DTSS analysis	[477]
		DFT		Opt		[665]
			AMBER	Fast-growth TI(MD), US(MD)		[755]
		AM1	OPLS-AA	FEP(MC)	Effect of mutation	[666]
		AM1, DFT		Opt, KIE	Dual-level QM approach	[276]
Methylmalonyl-CoA mutase		SCC-DFTB, DFT	AMBER	Opt		[442]
		HF		Opt	Comparison to FMO all-QM treatment	[667]
		AM1, DFT	CHARMM	Opt, US(MD)		[668]
	Claisen, Cope rearrangement					
	Substrate binding	AM1		Opt	Comparison of substrate conformations	[669]
	Substrate conformation	SCC-DFTB		MD		[670]
	Substrate conformation			US(MD)		[671]
	Radical rearrangement	DFT	CHARMM	Opt		[672]

Table 6 Ligases (EC 6)

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Threonyl-tRNA synthetase	Active-site structure	AM1, PM3	CHARMM	Opt		[673]

Table 7 Photo-active proteins

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Bacterial photosynthetic reaction centre	Quinone IR spectra	DFT	CHARMM	CP-MD		[674]
	Structure of bacteriochlorophyll dimer	AM1	AMBER	Opt	Effect of mutations	[675, 676]
Bacteriorhodopsin	Electron transfer	DFT, SCC-DFTB		Opt		[677]
	Electronic excitation, H ⁺ transfer	HF, CASCF, MRMP	AMBER	Opt, excitation energies, IR		[272]
	Electronic excitation	AM1, CIS	CHARMM, OPLS	MD, MC	QM/MM MC in solution, MM MD in enzyme/lipid	[678]
		HF, DFT; CIS, TDHF, TDDFT	AMBER	Opt		[49]
	H ⁺ transfer	SCC-DFTB	CHARMM	Opt	Coordinate driving vs. reaction-path opt	[679]
Photo-isomerization		HF, DFT		Opt	Reaction-path opt	[680]
		QCFF/PI	ENZYMIX	Opt, MD		[681]
				Quantum dynamics	Semi-classical trajectories	[682]
Structure of bound intermediates		SCC-DFTB	CHARMM	Opt		[683]
	Vibrational analysis of different states	HF	AMBER	Opt		[684]
	Structure and dynamics of solvated proton	DFT	GROMOS	CP-MD	Protein in lipid bilayer	[685]

Table 7 continued

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
Bovine rhodopsin	Structure of bound chromophore	SCC-DFTB	CHARMM	MD		[686]
	Structure of bound retinal analogue			MD		[687]
	Active-site structure	DFT	AMBER	Opt	Review	[688]
	Electronic excitation, structure of photoproduct	DFT, TDDFT		Opt		[689]
Rhodopsins	Electronic excitation	SCC-DFTB, CASSCF, CASPT2	CHARMM	MD	Different retinal derivatives	[690]
	NMR parameters	DFT	AMBER	Opt		[691]
	Chromophore structure and dynamics	DFT	AMBER	CP-MD	Comparison of different environments	[692]
	Electronic excitation	HF, CASSCF		Opt, excitation energies		[693]
		HF, SAC-CI		Opt		[694]
		CASSCF, CASPT2		Opt		[695]
Green fluorescent protein	Structure of bound chromophore	Various	Various	Opt	Method comparison	[696]
		DFT	UFF	Opt		[697]
	Chromophore structure, electronic excitation	CASSCF, CASPT2	CHARMM	Opt		[698]
	Electronic excitation	AM1, TDDFT		Opt		[699]
Phot-LOV1	Photo-excitation	HF, DFT	AMBER	Opt	TDDFT on isolated QM part only	[700]
Photo-active yellow protein	Photo-isomerization	PM3, CASSCF	GROMOS	MD		[701]

Table 8 Other proteins, protein–ligand complexes

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
β -Helix antifreeze protein	Interaction with water	AM1, DFT	AMBER	Energy		[725]
Crambin	Structure	SCC-DFTB	TIP3P	MD	All-QM protein in MM solvent, comparison to MM MD	[726]
Turkey ovomucoid third domain	pK_a calculation	HF, MP2	EFP	Opt		[727]
Hen egg white lysozyme–Fab HyHEL-5/Fab D44.1	Structure of antigen–antibody complex	AM1	CVFF	Opt		[728]
Various protein–ligand complexes	Ligand charges for docking	DFT	OPLS-AA	Single points		[729]
Thrombin	Inhibitor binding	HF	CHARMM	ESP on vdW surface		[730]
		AM1	AMBER, Tripos	Opt		[495]

Table 9 Oligonucleotide adducts

Biomolecule	Process studied	QM level	MM level	Calculation type	Comments	Refs.
DNA-cisplatin	structure, NMR parameters	DFT	AMBER	CP-MD		[731]
DNA-Hoechst 33258	Structure	AM1	AMBER	Opt	Effect of substituents	[732]
DNA-netropsin	Structure	HF	CHARMM	Opt		[733]
DNA-Pt complex	Structure	HF, DFT	UFF	Opt		[732]
DNA-Pt ₂ complex	Structure	DFT	AMBER	CP-MD		[734]
DNA-Ru complex	Structure	DFT	AMBER	CP-MD		[735]
DNA-duocarmycin	SN2 attack	DFT	AMBER	CP-MD		[736]
U1A-RNA	Protein-nucleotide association	DFT	OPLS-AA	Opt		[737]

molecule if they remain unchanged. "Comments", however, apply only to the specific row.

Acknowledgements H.M.S. thanks Drs. Johannes Kästner and Tell Tuttle for insightful discussions. This work was supported by the Volkswagenstiftung.

References

1. Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103(2):227–249
2. Warshel A, Karplus M (1972) Calculation of ground and excited state potential surfaces of conjugated molecules I: formulation and parametrization. *J Am Chem Soc* 94(16):5612–5625
3. Singh UC, Kollman PA (1986) A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ exchange reaction and gas phase protonation of polyethers. *J Comput Chem* 7(6):718–30
4. Field MJ, Bash PA, Karplus M (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J Comput Chem* 11(6):700–33
5. Gao J (1996) Methods and applications of combined quantum mechanical and molecular mechanical potentials. In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 7. VCH, New York, pp 119–185
6. Gao J (1996) Hybrid quantum mechanical/molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. *Acc Chem Res* 29(6):298–305
7. Cunningham MA, Bash PA (1997) Computational enzymology. *Biochimie* 79(11):687–689
8. Gao J, Thompson MA (eds) (1998) Combined quantum mechanical and molecular mechanical methods. ACS symposium series, vol 712. American Chemical Society, Washington, DC
9. Merz KM Jr (1998) Quantum mechanical-molecular mechanical coupled potentials. In: Gao J, Thompson MA (eds) Combined quantum mechanical and molecular mechanical methods. ACS symposium series, vol 712. American Chemical Society, Washington, DC, pp 2–15
10. Gao J (1998) Hybrid quantum mechanical/molecular mechanical (QM/MM) methods. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1257–1263
11. Amara P, Field MJ (1998) Combined quantum mechanical and molecular mechanical potentials. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, pp 431–437
12. Ruiz-López MF, Rivail JL (1998) Combined quantum mechanics and molecular mechanics approaches to chemical and biochemical reactivity. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, pp 437–448
13. Merz KM Jr, Stanton RV (1998) Quantum mechanical/molecular mechanical (QM/MM) coupled potentials. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 4. Wiley, Chichester, pp 2330–2343
14. Friesner RA, Beachy MD (1998) Quantum mechanical calculations on biological systems. *Curr Opin Struct Biol* 8(2):257–262

15. Beck B, Clark T (1998) Some biological applications of semiempirical MO theory. *Perspect Drug Disc Des* 9–11:131–159
16. Mordasini T, Thiel W (1998) Combined quantum mechanical and molecular approaches. *Chimia* 52(6):288–291
17. Monard G, Merz KM Jr (1999) Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Acc Chem Res* 32(10):904–911
18. Hillier IH (1999) Chemical reactivity studied by hybrid QM/MM methods. *THEOCHEM* 463(1–2):45–52
19. Amara P, Field MJ (1999) Hybrid potentials for large molecular systems. In: Leszczynski J (ed) *Computational molecular biology. Theoretical and computational chemistry*, vol 8. Elsevier, Amsterdam
20. Bruice TC, Kahn K (2000) Computational enzymology. *Curr Opin Chem Biol* 4:540–544
21. Sherwood P (2000) Hybrid quantum mechanics/molecular mechanics approaches. In: Grotendorst J (ed) *Modern methods and algorithms of quantum chemistry. NIC Series*, vol 1. John von Neumann Institute for Computing, Jülich, pp 257–277
22. Lyne PD, Walsh OA (2001) Computer simulation of biochemical reactions with QM-MM methods. In: Becker OM, Watanabe M (eds) *Computational biochemistry and biophysics*. Dekker, New York, pp 221–236
23. Mulholland AJ (2001) The QM/MM approach to enzymatic reactions. In: Eriksson LA (ed) *Theoretical biochemistry: processes and properties of biological systems. Theoretical and computational chemistry*, vol 9. Elsevier, Amsterdam, pp 597–653
24. Field MJ (2002) Simulating enzyme reactions: Challenges and perspectives. *J Comput Chem* 23(1):48–58
25. Gogonea V (2002) The QM/MM method. An overview. *Internet Electron J Mol Des* 1(4):173–184
26. Castillo R, Oliva M, Martí S, Moliner V, Tuñón I, and Andrés J (2002) Hybrid QM/MM studies on chemical reactivity. *Recent Res Dev Quantum Chem* 3:51–74
27. Monard G, Prat-Resina X, González-Lafont A, Lluch JM (2003) Determination of enzymatic reaction pathways using QM/MM methods. *Int J Quant Chem* 93(3):229–244
28. Ridder L, Mulholland AJ (2003) Modeling biotransformation reactions by combined quantum mechanical/molecular mechanical approaches: from structure to activity. *Curr Top Med Chem* 3(11):1241–1256
29. Náray-Szabó G, Berente I (2003) Computer modelling of enzyme reactions. *THEOCHEM* 666–667:637–644
30. Ryde U (2003) Combined quantum and molecular mechanics calculations on metalloproteins. *Curr Opin Chem Biol* 7(1):136–142
31. Peräkylä M (2003) Ab initio methods in the study of reaction mechanisms – their role and perspectives in medicinal chemistry. In: Carloni P, Alber F (eds) *Quantum medicinal chemistry. Methods and principles in medicinal chemistry*, vol 17. Wiley, Weinheim, pp 157–176
32. Perruccio F, Ridder L, Mulholland AJ (2003) Quantum-mechanical/molecular-mechanical methods in medicinal chemistry. In: Carloni P, Alber F (eds) *Quantum medicinal chemistry. methods and principles in medicinal chemistry*, vol 17. Wiley, Weinheim, pp 177–198
33. Rivail JL (2004) Hybrid quantum mechanical/molecular mechanical methods. In: Bultinck P, De Winter H, Langenaeker W, Tollenaere JP (eds) *Computational medicinal chemistry for drug discovery*. Dekker, New York, pp 119–131

34. Friesner RA (2004) Combined quantum and molecular mechanics (QM/MM). *Drug Discovery Today: Technol* 1(3):253–260
35. Mulholland AJ (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today* 10(20):1393–1402
36. Friesner RA, Guallar V (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu Rev Phys Chem* 56:389–427
37. Lin H, Truhlar DG (2006) QM/MM: What have we learned, where are we, and where do we go from here? *Theor Chem Acc*, Online first, 8 July 2006
38. Maple JR (1998) Force Fields: A general discussion. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1015–1024
39. MacKerell AD Jr (2001) Atomistic models and force fields. In: Becker OM, MacKerell AD Jr, Roux B, Watanabe M (eds) *Computational biochemistry and biophysics*. Dekker, New York, pp 7–38
40. Ponder JW, Case DA (2003) Force fields for protein simulations. In: Daggett V (ed) *Protein simulations. Advances in protein chemistry*, vol 66. Academic Press, San Diego, pp 27–85
41. Liljefors T, Gundertofte K, Norrby P-O, Pettersson I (2004) Molecular mechanics and comparison of force fields. In: Bultinck P, De Winter H, Langenaeker W, Tollenaere JP (eds) *Computational medicinal chemistry for drug discovery*. Dekker, New York, p 1–28
42. MacKerell AD Jr (2004) Empirical force fields for biological macromolecules: Overview and issues. *J Comput Chem* 25(13):1584–1604
43. Maseras F, Morokuma K (1995) IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J Comput Chem* 16(9):1170–1179
44. Humbel S, Sieber S, Morokuma K (1996) The IMOMO method: Integration of different levels of molecular orbital approximations for geometry optimization of large systems: Test for *n*-butane conformation and S_N2 reaction: $Cl^- + RCl$. *J Chem Phys* 105(5):1959–1967
45. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) ONIOM: A multilayered integrated MO + MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and $Pt(P(t-Bu)_3)_2 + H_2$ oxidative addition. *J Phys Chem* 100(50):19357–19363
46. Froese RDJ, Morokuma K (1998) Hybrid methods. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1244–1257
47. Dapprich S, Komáromi I, Byun KS, Morokuma K, Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *THEOCHEM* 461–462:1–21
48. Komáromi I, Muszbek L (2002) Application of the IMOMM (integrated molecular orbital molecular mechanics) method for biopolymers. In: Buzaneva E, Scharff P (eds) *Frontiers of multifunctional nanosystems. NATO science series II: Mathematics, physics and chemistry*, vol 57. Kluwer, Dordrecht, pp 17–28
49. Vreven T, Morokuma K (2003) Investigation of the $S_0 \rightarrow S_1$ excitation in bacteriorhodopsin with the ONIOM(MO:MM) hybrid method. *Theor Chem Acc* 109(3):125–132
50. Vreven T, Byun KS, Komáromi I, Dapprich S, Montgomery JA Jr, Morokuma K, Frisch MJ (2006) Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J Chem Theory Comput* 2(3):815–826

51. Bakowies D, Thiel W (1996) Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J Phys Chem* 100(25):10580–10594
52. Antes I, Thiel W (1998) On the treatment of link atoms in hybrid methods. In: Gao J, Thompson MA (eds) Combined quantum mechanical and molecular mechanical methods. ACS Symposium Series, vol 712. American Chemical Society, Washington, DC, pp 50–65
53. Rick SW, Stuart SJ (2002) Potentials and algorithms for incorporating polarizability in computer simulations. In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 18. Wiley, Hoboken, NJ, pp 89–146
54. Yu H, van Gunsteren WF (2005) Accounting for polarization in molecular simulation. *Comput Phys Commun* 172(2):69–85
55. Stern HA, Rittner F, Berne BJ, Friesner RA (2001) Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J Chem Phys* 115(5):2237–2251
56. Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B* 107(24):5933–5947
57. Grossfield A, Ren P, Ponder JW (2003) Ion solvation thermodynamics from simulation with a polarizable force field. *J Am Chem Soc* 125(50):15671–15682
58. Lamoureux G, Roux B (2003) Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J Chem Phys* 119(6):3025–3039
59. Lamoureux G, MacKerell AD Jr, Roux B (2003) A simple polarizable model of water based on classical Drude oscillators. *J Chem Phys* 119(10):5185–5197
60. Yu H, Hansson T, van Gunsteren WF (2003) Development of a simple, self-consistent polarizable model for liquid water. *J Chem Phys* 118(1):221–234
61. Yu H, van Gunsteren WF (2004) Charge-on-spring polarizable water models revisited: from water clusters to liquid water to ice. *J Chem Phys* 121(19):9549–9564
62. Patel S, Brooks CL III (2004) CHARMM fluctuating charge force field for proteins: I Parameterization and application to bulk organic liquid simulations. *J Comput Chem* 25(1):1–16
63. Patel S, MacKerell AD Jr, Brooks CL III (2004) CHARMM fluctuating charge force field for proteins: II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J Comput Chem* 25(12):1504–1514
64. Vorobyov IV, Anisimov VM, MacKerell AD Jr (2005) Polarizable empirical force field for alkanes based on the classical Drude oscillator model. *J Phys Chem B* 109(40):18988–18999
65. Anisimov VM, Lamoureux G, Vorobyov IV, Huang N, Roux B, MacKerell AD Jr (2005) Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J Chem Theory Comput* 1(1):153–168
66. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21(12):1049–1074
67. Cieplak P, Caldwell J, Kollman PA (2001) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and *N*-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comput Chem* 22(10):1048–1057
68. Wang Z-X, Zhang W, Wu C, Lei H, Cieplak P, Duan Y (2006) Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J Comput Chem* 27(6):781–790

69. Banks JL, Kaminski GA, Zhou R, Mainz DT, Berne BJ, Friesner RA (1999) Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *J Chem Phys* 110:741–754
70. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA (1999) Fluctuating charge, polarizable dipole, and combined models: parameterization from ab initio quantum chemistry. *J Phys Chem B* 103(22):4730–4737
71. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA (2002) Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J Comput Chem* 23(16):1515–1531
72. Kaminski GA, Stern HA, Berne BJ, Friesner RA (2004) Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. *J Phys Chem A* 108(4):621–627
73. Ren P, Ponder JW (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J Comput Chem* 23(16):1497–1506
74. Thompson MA, Schenter GK (1995) Excited states of the bacteriochlorophyll *b* dimer of *Rhodospseudomonas viridis*: a QM/MM study of the photosynthetic reaction center that includes MM polarization. *J Phys Chem* 99(17):6374–86
75. Gao J (1997) Energy components of aqueous solution: insight from hybrid QM/MM simulations using a polarizable solvent model. *J Comput Chem* 18(8):1061–1071
76. Bryce RA, Vincent MV, Malcolm NOJ, Hillier IH, Burton NA (1998) Cooperative effects in the structuring of fluoride water clusters: ab initio hybrid quantum mechanical/molecular mechanical model incorporating polarizable fluctuating charge solvent. *J Chem Phys* 109:3077–3085
77. Jensen L, van Duijnen PT, Snijders JG (2003) A discrete solvent reaction field model within density functional theory. *J Chem Phys* 118(2):514–521
78. Nam K, Gao J, York DM (2005) An efficient linear-scaling Ewald method for long-range electrostatic interactions in combined QM/MM calculations. *J Chem Theory Comput* 1(1):2–13
79. Riccardi D, Schaefer P, Cui Q (2005) pK_a calculations in solution and proteins with QM/MM free energy perturbation simulations: a quantitative test of QM/MM protocols. *J Phys Chem B* 109(37):17715–17733
80. Schaefer P, Riccardi D, Cui Q (2005) Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J Chem Phys* 123(1):014905/1–14
81. Dinner AR, Lopez X, Karplus M (2003) A charge-scaling method to treat solvent in QM/MM simulations. *Theor Chem Acc* 109(3):118–124
82. Gregersen BA, York DM (2005) Variational electrostatic projection (VEP) methods for efficient modeling of the macromolecular electrostatic and solvation environment in activated dynamics simulations. *J Phys Chem B* 109(1):536–556
83. Gregersen BA, York DM (2006) A charge-scaling implementation of the variational electrostatic projection method. *J Comput Chem* 27(1):103–115
84. Berkowitz M, McCammon JA (1982) Molecular dynamics with stochastic boundary conditions. *Chem Phys Lett* 90:215–217
85. Brünger A, Brooks CL III, Karplus M (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem Phys Lett* 105(5):495–500
86. Brooks CL III, Brünger A, Karplus M (1985) Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers* 24:843–865
87. Brooks CL III, Karplus M (1989) Solvent effects on protein motion and protein effects on solvent motion: dynamics of the active site region of lysozyme. *J Mol Biol* 208:159–181

88. King G, Warshel A (1989) A surface constrained all-atom solvent model for effective simulations of polar solutions. *J Chem Phys* 91(6):3647–3661
89. Beglov D, Roux B (1994) Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J Chem Phys* 100:9050–9063
90. Im W, Bernèche S, Roux B (2001) Generalized solvent boundary potential for computer simulations. *J Chem Phys* 114(7):2924–2937
91. König PH, Ghosh N, Hoffmann M, Elstner M, Tajkhorshid E, Frauenheim T, Cui Q (2006) Toward theoretical analysis of long-range proton transfer kinetics in biomolecular pumps. *J Phys Chem A* 110(2):548–563
92. Riccardi D, Schaefer P, Yang Y, Yu H, Ghosh N, Prat-Resina X, König P, Li G, Xu D, Guo H, Elstner M, Cui Q (2006) Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J Phys Chem B* 110(13):6458–6469
93. Murphy RB, Philipp DM, Friesner RA (2000) A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. *J Comput Chem* 21(16):1442–1457
94. Freindorf M, Shao Y, Furlani TR, Kong J (2005) Lennard–Jones parameters for the combined QM/MM method using the B3LYP/6-31+G*/AMBER potential. *J Comput Chem* 26(12):1270–1278
95. Riccardi D, Li G, Cui Q (2004) Importance of van der Waals interactions in QM/MM simulations. *J Phys Chem B* 108(20):6467–6478
96. Eurenus KP, Chatfield DC, Brooks BR, Hodoscek M (1996) Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *Int J Quant Chem* 60(6):1189–1200
97. Koga N, Morokuma K (1990) A simple scheme of estimating substitution or substituent effects in the ab initio MO method based on the shift operator. *Chem Phys Lett* 172(3–4):243–248
98. Matsubara T, Maseras F, Koga N, Morokuma K (1996) Application of the new “integrated MO + MM” (IMOMM) method to the organometallic reaction $\text{Pt}(\text{PR}_3)_2 + \text{H}_2$ ($\text{R} = \text{H}, \text{Me}, t\text{-Bu}, \text{and Ph}$). *J Phys Chem* 100(7):2573–2580
99. Senn HM (2001) Transition-metal-catalysed hydroamination of alkenes: ab initio molecular-dynamics studies. PhD Thesis, Diss ETH no 13972. ETH Zürich, Zürich
100. Eichler U, Kölmel CM, Sauer J (1997) Combining ab initio techniques with analytical potential functions for structure predictions of large systems: method and application to crystalline silica polymorphs. *J Comput Chem* 18(4):463–477
101. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) A hybrid method for solutes in complex solvents: density functional theory combined with empirical force fields. *J Chem Phys* 110(21):10452–10467
102. Lyne PD, Hodoscek M, Karplus M (1999) A hybrid QM–MM potential employing Hartree–Fock or density functional methods in the quantum region. *J Phys Chem A* 103(18):3462–3471
103. de Vries AH, Sherwood P, Collins SJ, Rigby AM, Rigutto M, Kramer GJ (1999) Zeolite structure and reactivity by combined quantum-chemical–classical calculations. *J Phys Chem B* 103(29):6133–6141
104. Field MJ, Albe M, Bret C, Proust-De Martin F, Thomas A (2000) The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J Comput Chem* 21(12):1088–1100
105. Woo TK, Cavallo L, Ziegler T (1998) Implementation of the IMOMM methodology for performing combined QM/MM molecular dynamic simulations and frequency calculations. *Theor Chem Acc* 100(5–6):307–313

106. Swart M (2003) AddRemove: a new link model for use in QM/MM studies. *Int J Quant Chem* 91(2):177–183
107. Das D, Eurenium KP, Billings EM, Sherwood P, Chatfield DC, Hodošček M, Brooks BR (2002) Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J Chem Phys* 117(23):10534–10547
108. Ferré N, Olivucci M (2003) The amide bond: pitfalls and drawbacks of the link atom scheme. *THEOCHEM* 632:71–82
109. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) Frontier bonds in QM/MM methods: a comparison of different approaches. *J Phys Chem A* 104(8):1720–1735
110. Waszkowycz B, Hillier IH, Gensmantel N, Payling DW (1990) A theoretical study of the hydrolysis by phospholipase A₂: the catalytic role of the active site and substrate specificity. *J Chem Soc, Perkin Trans 2* (7):1259–1268
111. Waszkowycz B, Hillier IH, Gensmantel N, Payling DW (1991) A combined quantum mechanical/molecular mechanical model of the potential energy surface of ester hydrolysis by the enzyme phospholipase A₂. *J Chem Soc, Perkin Trans 2* (2):225–231
112. Waszkowycz B, Hillier IH, Gensmantel N, Payling DW (1991) A quantum mechanical/molecular mechanical model of the inhibition of the enzyme phospholipase A₂. *J Chem Soc, Perkin Trans 2* (11):1819–1832
113. Vasilyevy VV (1994) Tetrahedral intermediate formation in the acylation step of acetylcholinesterases. A combined quantum chemical and molecular mechanical model. *THEOCHEM* 110(2):129–41
114. Lin H, Truhlar DG (2005) Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *J Phys Chem A* 109(17):3991–4004
115. König PH, Hoffmann M, Frauenheim T, Cui Q (2005) A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J Phys Chem B* 109(18):9082–9095
116. Sherwood P, de Vries AH, Collins SJ, Greatbanks SP, Burton NA, Vincent MV, Hillier IH (1997) Computer simulation of zeolite structure and reactivity using embedded cluster methods. *Faraday Discuss* 106:79–92
117. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, Billeter S, Terstegen F, Thiel S, Kendrick J, Rogers SC, Casci J, Watson M, King F, Karlsen E, Sjøvoll M, Fahmi A, Schäfer A, Lennartz C (2003) QUASI: a general purpose implementation of the QM/MM approach and its application to problems in catalysis. *THEOCHEM* 632(1–3):1–28
118. Amara P, Field MJ (2003) Evaluation of an ab initio quantum mechanical/molecular mechanical hybrid-potential link-atom method. *Theor Chem Acc* 109(1):43–52
119. CPMD, v. 3.9, IBM Corp., 1990–2001, MPI für Festkörperforschung, Stuttgart, 1997–2004. <http://www.cpmc.org/>
120. Antes I, Thiel W (1999) Adjusted connection atoms for combined quantum mechanical and molecular mechanical methods. *J Phys Chem A* 103:9290–9295
121. Zhang Y, Lee T-S, Yang W (1999) A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J Chem Phys* 110(1):46–54
122. Zhang Y (2005) Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *J Chem Phys* 122(2):024114/1–7
123. Alary F, Poteau R, Heully J-L, Barthelat J-L, Daudey J-P (2000) A new method for modelling spectator chemical groups in ab initio calculations: effective group potentials. *Theor Chem Acc* 104(3–4):174–178

124. Poteau R, Ortega I, Alary F, Solis AR, Barthelat J-C, Daudey J-P (2001) Effective group potentials. 1. Method. *J Phys Chem A* 105(1):198–205
125. Poteau R, Alary F, El Makarim HA, Heully J-L, Barthelat J-C, Daudey J-P (2001) Effective group potentials. 2. Extraction and transferability for chemical groups involved in covalent or donor–acceptor bonds. *J Phys Chem A* 105(1):206–214
126. Bessac F, Alary F, Carissan Y, Heully J-L, Daudey J-P, Poteau R (2003) Effective group potentials: a powerful tool for hybrid QM/MM methods? *THEOCHEM* 632(1–3): 43–59
127. Carissan Y, Bessac F, Alary F, Heully J-L, Poteau R (2006) What can we do with an effective group potential? *Int J Quant Chem* 106(3):727–733
128. DiLabio GA, Hurley MM, Christiansen PA (2002) Simple one-electron quantum capping potentials for use in hybrid QM/MM studies of biological molecules. *J Chem Phys* 116(22):9578–9584
129. DiLabio GA, Wolkow RA, Johnson ER (2005) Efficient silicon surface and cluster modeling using quantum capping potentials. *J Chem Phys* 122(4):044708/1–5
130. Yasuda K, Yamaki D (2004) Simple minimum principle to derive a quantum-mechanical/molecular-mechanical method. *J Chem Phys* 121(9):3964–3972
131. Lilienfeld OA, Tavernelli I, Rothlisberger U, Sebastiani D (2005) Variational optimization of effective atom centered potentials for molecular properties. *J Chem Phys* 122(1):014113/1–6
132. Laio A, VandeVondele J, Rothlisberger U (2002) A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J Chem Phys* 116(16):6941–6947
133. Slaviček P, Martínez TJ (2006) Multicentered valence electron effective potentials: A solution to the link atom problem for ground and excited electronic states. *J Chem Phys* 124(8):084107/1–10
134. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) Quantum mechanical computations on very large molecular systems: the local self-consistent field method. *J Comput Chem* 15(3):269–282
135. Monard G, Loos M, Théry V, Baka K, Rivail J-L (1996) Hybrid classical quantum force field for modeling very large molecules. *Int J Quant Chem* 58(2):153–159
136. Assfeld X, Rivail J-L (1996) Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method. *Chem Phys Lett* 263:100–106
137. Assfeld X, Ferré N, Rivail JL (1998) The local self-consistent field: principles and applications to combined quantum mechanical–molecular mechanical computations on biomacromolecular systems. In: Gao J, Thompson MA (eds) *Combined quantum mechanical and molecular mechanical methods*. ACS symposium series, vol 712. American Chemical Society, Washington, DC, pp 234–249
138. Ferré N, Assfeld X, Rivail J-L (2002) Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *J Comput Chem* 23(6):610–624
139. Hall RJ, Hindle SA, Burton NA, Hillier IH (2000) Aspects of hybrid QM/MM calculations: the treatment of the QM/MM interface region and geometry optimization with an application to chorismate mutase. *J Comput Chem* 21(16):1433–1441
140. Nicoll RM, Hindle SA, MacKenzie G, Hillier IH, Burton NA (2001) Quantum mechanical/molecular mechanical methods and the study of kinetic isotope effects: modelling the covalent junction region and application to the enzyme xylose isomerase. *Theor Chem Acc* 106(1–2):105–112
141. Fornili A, Sironi M, Raimondi M (2003) Determination of extremely localized molecular orbitals and their application to quantum mechanics/molecular mechan-

- ics methods and to the study of intramolecular hydrogen bonding. *THEOCHEM* 632(1–3):157–172
142. Fornili A, Moreau Y, Sironi M, Assfeld X (2006) On the suitability of strictly localized orbitals for hybrid QM/MM calculations. *J Comput Chem* 27(4):515–523
143. Philipp DM, Friesner RA (1999) Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J Comput Chem* 20(14):1468–1494
144. Murphy RB, Philipp DM, Friesner RA (2000) Frozen orbital QM/MM methods for density functional theory. *Chem Phys Lett* 321(1–2):113–120
145. Gao J, Amara P, Alhambra C, Field MJ (1998) A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J Phys Chem A* 102(24):4714–4721
146. Amara P, Field MJ, Alhambra C, Gao J (2000) The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives. *Theor Chem Acc* 104(5):336–343
147. Garcia-Viloca M, Gao J (2004) Generalized hybrid orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method. *Theor Chem Acc* 111(2–6):280–286
148. Pu J, Gao J, Truhlar DG (2004) Generalized hybrid orbital (GHO) method for combining ab initio Hartree–Fock wave functions with molecular mechanics. *J Phys Chem A* 108(4):632–650
149. Pu J, Gao J, Truhlar DG (2004) Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J Phys Chem A* 108(25):5454–5463
150. Pu J, Gao J, Truhlar DG (2005) Generalized hybrid-orbital method for combining density functional theory with molecular mechanicals. *Chem Phys Chem* 6(9):1853–1865
151. Lennartz C, Schäfer A, Terstegen F, Thiel W (2002) Enzymatic reactions of triose-phosphate isomerase: a theoretical calibration study. *J Phys Chem B* 106(7):1758–1767
152. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J Phys Chem B* 105(2):569–585
153. Elstner M, Frauenheim T, Suhai S (2003) An approximate DFT method for QM/MM simulations of biological structures and processes. *THEOCHEM* 632(1–3):29–41
154. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* 58(11):7260–7268
155. Schütz M, Hetzer G, Werner H-J (1999) Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J Chem Phys* 111(13):5691–5705
156. Schütz M, Werner H-J (2000) Local perturbative triples correction (T) with linear cost scaling. *Chem Phys Lett* 318(4–5):370–378
157. Schütz M (2000) Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J Chem Phys* 113(22):9986–10001
158. Schütz M, Werner H-J (2001) Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J Chem Phys* 114(2):661–681
159. Schütz M (2002) Low-order scaling local electron correlation methods. V. Connected triples beyond (T): Linear scaling local CCSDT-1b. *J Chem Phys* 116(20):8772–8785

160. Schütz M, Manby FR (2003) Linear scaling local coupled cluster theory with density fitting. Part I: 4-external integrals. *Phys Chem Chem Phys* 5(16):3349–3358
161. Werner H-J, Manby FR, Knowles PJ (2003) Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J Chem Phys* 118(18):8149–8160
162. Werner H-J, Manby FR (2006) Explicitly correlated second-order perturbation theory using density fitting and local approximations. *J Chem Phys* 124(5):054114/1–12
163. Car R, Parrinello M (1985) Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett* 55:2471–2474
164. Marx D, Hutter J (2000) Ab initio molecular dynamics: Theory and implementation. In: Grotendorst J (ed) *Modern methods and algorithms of quantum chemistry*. NIC Series, vol 1. John von Neumann Institute of Computing, Jülich, pp 301–449, <http://www.fz-juelich.de/nic-series/Volume3/marx.pdf>
165. Woo TK, Blöchl PE, Ziegler T (2000) Towards solvation simulations with a combined ab initio molecular dynamics and molecular mechanics approach. *THEOCHEM* 506(1–3):313–334
166. Blöchl PE (1995) Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *J Chem Phys* 103:7422–7428
167. Biswas PK, Gogonea V (2005) A regularized and renormalized electrostatic coupling Hamiltonian for hybrid quantum-mechanical-molecular-mechanical calculations. *J Chem Phys* 123:164114/1–9
168. Laio A, VandeVondele J, Rothlisberger U (2002) D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J Phys Chem B* 106(29):7300–7307
169. Laio A, Gervasio FL, VandeVondele J, Sulpizi M, Rothlisberger U (2004) A variational definition of electrostatic potential derived charges. *J Phys Chem B* 108(23):7963–7968
170. Laino T, Mohamed F, Laio A, Parrinello M (2005) An efficient real space multigrid QM/MM electrostatic coupling. *J Chem Theory Comput* 1(6):1176–1184
171. MacKerell AD Jr (1998) Protein Force Fields. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, pp 2191–2200
172. MacKerell AD Jr (2005) Empirical force fields for proteins: current status and future directions. In: Spellmeyer DC (ed) *Annual reports in computational chemistry*, vol 1. Elsevier, Amsterdam, pp 91–102
173. Cieplak P (1998) Nucleic acid force fields. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, pp 1922–1930
174. Cheatham TE III (2005) Molecular modeling and atomistic simulation of nucleic acids. In: Spellmeyer DC (ed) *Annual reports in computational chemistry*, vol 1. Elsevier, Amsterdam, pp 75–89
175. Woods RJ (1998) Carbohydrate force fields. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, pp 220–233
176. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, organic molecules. *J Am Chem Soc* 117:5179–5197
177. Kollman P, Caldwell JW, Ross WS, Pearlman DA, Case DA, DeBolt S, Cheatham TE III, Ferguson D, Seibel G (1998) AMBER: a program for simulation of biological and organic molecules. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, pp 11–13

178. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24(16):1999–2012
179. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616
180. MacKerell AD Jr, Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y, Karplus M (1998) CHARMM: The energy function and its parameterization with an overview of the program. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, pp 271–277
181. Foloppe N, MacKerell AD Jr (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macro-molecular target data. *J Comput Chem* 21(2):86–104
182. MacKerell AD Jr, Banavali NK (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J Comput Chem* 21(2):105–120
183. van Gunsteren WF, Daura X, Mark AE (1998) GROMOS force field. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1211–1216
184. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF (1999) The GROMOS biomolecular simulation program package. *J Phys Chem A* 103(19):3596–3607
185. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118(45):11225–11236
186. Jorgensen WL (1998) OPLS force fields. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, pp 1986–1989
187. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105(28):6474–6487
188. Allinger NL, Yuh YH, Lii JH (1989) Molecular mechanics: The MM3 force-field for hydrocarbons. 1. *J Am Chem Soc* 111(23):8551–8566
189. Lii JH, Allinger NL (1989) Molecular mechanics. The MM3 force-field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J Am Chem Soc* 111(23):8566–8575
190. Lii JH, Allinger NL (1989) Molecular mechanics. The MM3 force-field for hydrocarbons. 3. The van der Waals potentials and crystal data for aliphatic and aromatic hydrocarbons. *J Am Chem Soc* 111(23):8576–8582
191. Allinger NL, Li FB, Yan LQ (1990) Molecular mechanics: The MM3 force-field for alkenes. *J Comput Chem* 11(7):848–867
192. Allinger NL, Rahman M, Lii JH (1990) A molecular mechanics force-field (MM3) for alcohols and ethers. *J Am Chem Soc* 112:8293–8307
193. Lii JH, Allinger NL (1991) The MM3 force-field for amides, polypeptides and proteins. *J Comput Chem* 12(2):186–199
194. Allinger NL (1998) Force Fields: MM3. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1028–1033

195. Allinger NL, Chen KS, Lii JH (1996) An improved force field (MM4) for saturated hydrocarbons. *J Comput Chem* 17(5-6):642-668
196. Langley CH, Lii JH, Allinger NL (2001) Molecular mechanics (MM4) calculations on carbonyl compounds. I. Aldehydes. *J Comput Chem* 22(13):1396-1425
197. Langley CH, Lii JH, Allinger NL (2001) Molecular mechanics calculations on carbonyl compounds. II. Open-chain ketones. *J Comput Chem* 22(13):1426-1450
198. Langley CH, Lii JH, Allinger NL (2001) Molecular mechanics calculations on carbonyl compounds. III. Cycloketones. *J Comput Chem* 22(13):1451-1475
199. Langley CH, Lii JH, Allinger NL (2001) Molecular mechanics calculations on carbonyl compounds. IV. Heats of formation. *J Comput Chem* 22(13):1476-1483
200. Langley CH, Lii JH, Allinger N (2003) Molecular mechanics (MM4) calculations on carbonyl compounds. I-IV (Erratum). *J Comput Chem* 24(10):1283-1286
201. Langley CH, Allinger NL (2002) Molecular mechanics (MM4) calculations on amides. *J Phys Chem A* 106(23):5638-5652
202. Allinger NL, Chen KH, Lii JH, Durkin KA (2003) Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds. *J Comput Chem* 24(12):1447-1472
203. Lii JH, Chen KH, Durkin KA, Allinger NL (2003) Alcohols, ethers, carbohydrates, and related compounds. II. The anomeric effect. *J Comput Chem* 24(12):1473-1489
204. Lii JH, Chen KH, Grindley TB, Allinger NL (2003) Alcohols, ethers, carbohydrates, and related compounds. III. The 1,2-dimethoxyethane system. *J Comput Chem* 24(12):1490-1503
205. Lii JH, Chen KH, Allinger NL (2003) Alcohols, ethers, carbohydrates, and related compounds. IV. Carbohydrates. *J Comput Chem* 24(12):1504-1513
206. Halgren TA (1996) Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 17(5-6):490-519
207. Halgren TA (1996) Merck molecular force field. 2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J Comput Chem* 17(5-6):520-552
208. Halgren TA (1996) Merck molecular force field. 3. Molecular geometries and vibrational frequencies for MMFF94. *J Comput Chem* 17(5-6):553-586
209. Halgren TA, Nachbar RB (1996) Merck molecular force field. 4. Conformational energies and geometries for MMFF94. *J Comput Chem* 17(5-6):587-615
210. Halgren TA (1996) Merck molecular force field. 5. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J Comput Chem* 17(5-6):616-641
211. Halgren TA (1998) Force Fields: MMFF94. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1033-1035
212. Halgren TA (1999) MMFF VI. MMFF94s option for energy minimization studies. *J Comput Chem* 20(7):720-729
213. Halgren TA (1999) MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular interaction energies and geometries. *J Comput Chem* 20(7):730-748
214. Rappé AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* 114:10024-10035
215. Casewit CJ, Colwell KS, Rappé AK (1992) Application of a universal force field to organic molecules. *J Am Chem Soc* 114:10035-10046

216. Casewit CJ, Colwell KS, Rappé AK (1992) Application of a universal force field to main group compounds. *J Am Chem Soc* 114:10046–10053
217. Rappé AK, Colwell KS, Casewit CJ (1993) Application of a universal force field to metal complexes. *Inorg Chem* 32:3438–3450
218. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The AMBER biomolecular simulation programs. *J Comput Chem* 26(16):1668–1688
219. Jorgensen WL, Tirado-Rives J (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* 26(16):1689–1700
220. te Velde GT, Bickelhaupt FM, Baerends EJ, Guerra CF, van Gisbergen SJA, Snijders GJ, Ziegler T (2001) Chemistry with ADF. *J Comput Chem* 22(9):931–967
221. Guest MF, Bush IJ, van Dam HJJ, Sherwood P, Thomas JMH, van Lenthe JH, Havenith RWA, Kendrick J (2005) The GAMESS-UK electronic structure package: algorithms, developments and applications. *Mol Phys* 103(6–8):719–747
222. Kendall RA, Apra E, Bernholdt DE, Bylaska EJ, Dupuis M, Fann GI, Harrison RJ, Ju JL, Nichols JA, Nieplocha J, Straatsma TP, Windus TL, Wong AT (2000) High performance computational chemistry: an overview of NWChem a distributed parallel application. *Comput Phys Commun* 128(1–2):260–283
223. Colombo MC, Guidoni L, Laio A, Magistrato A, Maurer P, Piana S, Röhrig U, Spiegel K, Sulpizi M, VandeVondele J, Zumstein M, Röthlisberger U (2002) Hybrid QM/MM Car-Parrinello simulations of catalytic and enzymatic reactions. *Chimia* 56(1–2):13–19 2002.
224. Sebastiani D, Röthlisberger U (2003) Advances in density-functional-based modeling techniques – recent extensions of the Car-Parrinello approach. In: Carloni P, Alber F (eds) *Quantum medicinal chemistry. Methods and principles in medicinal chemistry*, vol 17. Wiley, Weinheim, pp 5–39
225. Woo TK, Margl PM, Blöchl PE, Ziegler T (1997) A combined Car-Parrinello QM/MM implementation for ab initio molecular dynamics simulations of extended systems: application to transition metal catalysis. *J Phys Chem B* 101:7877–7880
226. Woo TK, Margl PM, Deng L, Ziegler T (1998) A combined Car-Parrinello quantum mechanical-molecular mechanical implementation for ab initio molecular dynamics simulations of extended systems. In: Gao J, Thompson MA (eds) *Combined quantum mechanical and molecular mechanical methods. ACS symposium series*, vol 712. American Chemical Society, Washington, DC, pp 128–147
227. Woo TK, Margl PM, Deng L, Cavallo L, Ziegler T (1999) Towards more realistic computational modeling of homogenous catalysis by density functional theory: Combined QM/MM and ab initio molecular dynamics. *Catal Today* 50:479–500
228. Thiel W (2004) MNDO99, V. 6.1. Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany
229. Gale JD, Rohl AL (2003) The general utility lattice program (GULP). *Mol Simul* 29(5):291–341
230. Gale JD (2005) GULP: Capabilities and prospects. *Z Kristallogr* 220(5–6):552–554
231. Warshel A (1991) *Computer modeling of chemical reactions in enzymes and solutions*. Wiley, New York
232. Åqvist J, Warshel A (1993) Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem Rev* 93:2523–2544
233. Shurki A, Warshel A (2003) Structure/function correlations of proteins using MM, QM/MM, and related approaches: Methods, concepts, pitfalls, and current progress. In: Daggett V (ed) *Protein simulations. Advances in protein chemistry*, vol 66. Academic Press, San Diego, pp 249–313

234. Warshel A (2003) Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu Rev Biophys Biomol Struct* 32:425–443
235. Jensen JH, Day PN, Gordon MS, Basch H, Cohen D, Garmer DR, Krauss M, Stevens WJ (1994) Effective fragment method for modeling intermolecular hydrogen-bonding effects on quantum mechanical calculations. In: Smith DA (ed) *Modeling the hydrogen bond*. ACS symposium series, vol 569. American Chemical Society, Washington, DC, pp 139–151
236. Day PN, Jensen JH, Gordon MS, Webb SP, Stevens WJ, Krauss M, Garmer D, Basch H, Cohen D (1996) An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J Chem Phys* 105(5):1968–1986
237. Kairys V, Jensen JH (2000) QM/MM boundaries across covalent bonds: a frozen localized molecular orbital-based approach for the effective fragment potential method. *J Phys Chem A* 104(28):6656–6665
238. Gordon MS, Freitag MA, Bandyopadhyay P, Jensen JH, Kairys V, Stevens WJ (2001) The effective fragment potential method: a QM-based MM approach to modeling environmental effects in chemistry. *J Phys Chem A* 105(2):293–307
239. Grigorenko BL, Nemukhin AV, Topol IA, Burt SK (2002) Modeling of biomolecular systems with the quantum mechanical and molecular mechanical method based on the effective fragment potential technique: proposal of flexible fragments. *J Phys Chem A* 106(44):10663–10672
240. Nemukhin AV, Grigorenko BL, Bochenkova AV, Topol IA, Burt SK (2002) A QM/MM approach with effective fragment potentials applied to the dipeptide-water structures. *THEOCHEM* 581(1–3):167–175
241. Nemukhin AV, Grigorenko BL, Topol IA, Burt SK (2003) Flexible effective fragment QM/MM method: validation through the challenging tests. *J Comput Chem* 24(12):1410–1420
242. Ryde U, Olsen L, Nilsson K (2002) Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J Comput Chem* 23(11):1058–1070
243. Ryde U, Nilsson K (2003) Quantum chemistry can locally improve protein crystal structures. *J Am Chem Soc* 125(47):14232–14233
244. Ryde U, Nilsson K (2003) Quantum refinement—a combination of quantum chemistry and protein crystallography. *THEOCHEM* 632(1–3):259–275
245. Tronrud DE (2004) Introduction to macromolecular refinement. *Acta Crystallogr D: Biol Crystallogr* 60:2156–2168
246. Kleywegt GJ, Henrick K, Dodson EJ, van Aalten DMF (2003) Pound-wise but penny-foolish: How well do micromolecules fare in macromolecular refinement? *Structure* 11(9):1051–1059
247. Klähn M, Braun-Sand S, Rosta E, Warshel A (2005) On possible pitfalls in ab initio quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions. *J Phys Chem B* 109(32):15645–15650
248. Zhang Y, Kua J, McCammon JA (2003) Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies. *J Phys Chem B* 107(18):4459–4463
249. Schlegel HB (1995) Geometry optimization on potential energy surfaces. In: Yarkony DR (ed) *Modern electronic structure theory*. Advanced series in physical chemistry, vol 2/I. World Scientific, Singapore, pp 459–500
250. Schlegel HB (1998) Geometry optimization I. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1136–1142
251. Schlick T (1998) Geometry optimization II. In: von Ragué Schleyer P (ed) *Encyclopedia of computational chemistry*, vol 2. Wiley, Chichester, pp 1142–1157

252. Jensen F (1998) Transition structure optimization techniques. In: von Ragué Schleyer P (ed) Encyclopedia of computational chemistry, vol 5. Wiley, Chichester, pp 3114–2123
253. Schlegel HB (2003) Exploring potential energy surfaces for chemical reactions: an overview of some practical methods. *J Comput Chem* 24(12):1514–1527
254. Farkas Ö, Schlegel HB (1998) Methods for geometry optimization of large molecules. I. An $O(N^2)$ algorithm for solving systems of linear equations for the transformation of coordinates and forces. *J Chem Phys* 109(17):7100–7104
255. Farkas Ö, Schlegel HB (1999) Methods for optimizing large molecules. Part II. Quadratic search. *J Chem Phys* 111(24):10806–10814
256. Farkas Ö, Schlegel HB (2002) Methods for optimizing large molecules. Part III. An improved algorithm for geometry optimization using direct inversion in the iterative subspace (GDIIS). *Phys Chem Chem Phys* 4(1):11–15
257. Farkas Ö, Schlegel HB (2003) Geometry optimization methods for modeling large molecules. *THEOCHEM* 666:31–39
258. Paizs B, Fogarasi G, Pulay P (1998) An efficient direct method for geometry optimization of large molecules in internal coordinates. *J Chem Phys* 109(16):6571–6576
259. Baker J, Kinghorn D, Pulay P (1999) Geometry optimization in delocalized internal coordinates: an efficient quadratically scaling algorithm for large molecules. *J Chem Phys* 110:4986–4991
260. Paizs B, Baker J, Suhai S, Pulay P (2000) Geometry optimization of large biomolecules in redundant internal coordinates. *J Chem Phys* 113(16):6566–6572
261. Németh K, Coulaud O, Monard G, Ángyán JG (2000) Linear scaling algorithm for the coordinate transformation problem of molecular geometry optimization. *J Chem Phys* 113(14):5598–5603
262. Németh K, Coulaud O, Monard G, Ángyán JG (2001) An efficient method for the coordinate transformation problem of massively three-dimensional networks. *J Chem Phys* 114(22):9747–9753
263. Prat-Resina X, Garcia-Viloca M, Monard G, González-Lafont A, Lluch JM, Bofill JM, Anglada JM (2002) The search for stationary points on a quantum mechanical/molecular mechanical potential-energy surface. *Theor Chem Acc* 107(3):147–153
264. Billeter SR, Turner AJ, Thiel W (2000) Linear scaling geometry optimization and transition state search in hybrid delocalized internal coordinates. *Phys Chem Chem Phys* 2:2177–2186
265. Baker J, Kessi A, Delley B (1996) The generation and use of delocalized internal coordinates in geometry optimization. *J Chem Phys* 105:192–212
266. Pulay P, Fogarasi G (1992) Geometry optimization in redundant internal coordinates. *J Chem Phys* 96(4):2856–2860
267. Moliner V, Turner AJ, Williams IH (1997) Transition-state structural refinement with GRACE and CHARMM: realistic modeling of lactate dehydrogenase using a combined quantum/classical method. *Chem Commun* (14):1271–1272
268. Turner AJ, Moliner V, Williams IH (1999) Transition-state structural refinement with GRACE and CHARMM: flexible QM/MM modeling for lactate dehydrogenase. *Phys Chem Chem Phys* 1(6):1323–1331
269. Prat-Resina X, González-Lafont A, Lluch JM (2003) How important is the refinement of transition state structures in enzymatic reactions? *THEOCHEM* 632(1–3):297–307
270. Vreven T, Morokuma K, Farkas Ö, Schlegel HB, Frisch MJ (2003) Geometry optimization with QM/MM, ONIOM, and other combined methods. I: Microiterations and constraints. *J Comput Chem* 24(6):760–769

271. Vreven T, Frisch MJ, Kudin KN, Schlegel HB, Morokuma K (2006) Geometry optimization with QM/MM methods. II: Explicit quadratic coupling. *Mol Phys* 104(5–7): 701–714
272. Hayashi S, Ohmine I (2000) Proton transfer in bacteriorhodopsin: structure, excitation, IR spectra, and potential energy surface analyses by an ab initio QM/MM method. *J Phys Chem B* 104(45):10678–10691
273. Zhang Y, Liu H, Yang W (2000) Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J Chem Phys* 112(8):3483–3492
274. Prat-Resina X, Bofill JM, González-Lafont À, Lluch JM (2004) Geometry optimization and transition state search in enzymes: different options in the microiterative method. *Int J Quant Chem* 98(4):367–377
275. Martí S, Moliner V, Tuñón I, Williams IH (2005) Computing kinetic isotope effects for chorismate mutase with high accuracy: a new DFT/MM strategy. *J Phys Chem B* 109(9):3707–3710
276. Martí S, Moliner V, Tuñón I (2005) Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. *J Chem Theory Comput* 1(5):1008–1016
277. Martí S, Moliner V, Tuñón I (2006) Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems (Erratum). *J Chem Theory Comput* 2(1):216
278. Mills G, Jónsson H (1994) Quantum and thermal effects in H₂ dissociative adsorption: evaluation of free-energy barriers in multidimensional quantum systems. *Phys Rev Lett* 72(7):1124–1127
279. Jónsson H, Mills G, Jacobsen KW (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF (eds) *Classical and quantum dynamics in condensed phase simulations*. World Scientific, Singapore, pp 387–404
280. Henkelman G, Uberuaga BP, Jónsson H (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J Chem Phys* 113(22):9901–9904
281. Henkelman G, Jónsson H (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J Chem Phys* 113(22):9978–9985
282. Henkelman G, Jóhannesson G, Jónsson H (2002) Methods for finding saddle points and minimum energy paths. In: Schwartz SD (ed) *Theoretical methods in condensed phase chemistry*. Progress in theoretical chemistry and physics, vol 5. Kluwer, Dordrecht, pp 269–300
283. Xie L, Liu H, Yang W (2004) Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes. *J Chem Phys* 120(17):8039–8052
284. Ayala PY, Schlegel HB (1997) A combined method for determining reaction paths, minima, and transition state geometries. *J Chem Phys* 107:375–384
285. Liu H, Lu Z, Cisneros GA, Yang W (2004) Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. *J Chem Phys* 121(2):697–706
286. Cisneros GA, Liu H, Lu Z, Yang W (2005) Reaction path determination for quantum mechanical/molecular mechanical modeling of enzyme reactions by combining first order and second order “chain-of-replicas” methods. *J Chem Phys* 122(11):114502/1–7

287. Woodcock HL, Hodošček M, Sherwood P, Lee YS, Schaefer HF III, Brooks BR (2003) Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor Chem Acc* 109(3):140–148
288. Allen MP, Tildesley DJ (1987) Computer simulation of liquids. Oxford University Press, Oxford
289. Beveridge DL, DiCapua FM (1989) Free energy via molecular simulation: application to chemical and biomolecular systems. *Annu Rev Biophys Biophys Chem* 18:431–492
290. Straatsma TP, McCammon JA (1992) Computational alchemy. *Annu Rev Phys Chem* 43:407–435
291. Frenkel D, Smit B (2002) Understanding molecular simulation, 2nd ed. Computational science series, vol 1. Academic Press, San Diego
292. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106:1589–1615
293. Bash PA, Field MJ, Karplus M (1987) Free energy perturbation method for chemical reactions in the condensed phase: a dynamic approach based on a combined quantum and molecular mechanics potential. *J Am Chem Soc* 109:8092–8094
294. Gao J, Xia X (1992) A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* 258(5082):631–635
295. Gao J (1992) Absolute free-energy of solvation from Monte-Carlo simulations using combined quantum and molecular mechanical potentials. *J Phys Chem* 96(2):537–540
296. Stanton RV, Hartsough DS, Merz KM Jr (1993) Calculation of solvation free energies using a density functional/molecular dynamics coupled potential. *J Phys Chem* 97(46):11868–11870
297. Hartsough DS, Merz KM Jr (1995) Potential of mean force calculations on the S_N1 fragmentation of *tert*-butyl chloride. *J Phys Chem* 99(1):384–390
298. Stanton RV, Little LR, Merz KM Jr (1995) An examination of a Hartree–Fock/molecular mechanical coupled potential. *J Phys Chem* 99(48):17344–17348
299. Stanton RV, Hartsough DS, Merz KM Jr (1995) An examination of a density-functional/molecular mechanical coupled potential. *J Comput Chem* 16(1):113–128
300. Hartsough DS, Merz KM Jr (1995) Dynamic force field models: molecular dynamics simulations of human carbonic anhydrase II using a quantum mechanical/molecular mechanical coupled potential. *J Phys Chem* 99(28):11266–11275
301. Senn HM, Thiel S, Thiel W (2005) Enzymatic hydroxylation in *p*-hydroxybenzoate hydroxylase: a case study for QM/MM molecular dynamics. *J Chem Theory Comput* 1(3):494–505
302. Schlegel HB, Millam JM, Iyengar SS, Voth GA, Daniels AD, Scuseria GE, Frisch MJ (2001) Ab initio molecular dynamics: propagating the density matrix with Gaussian orbitals. *J Chem Phys* 114(22):9758–9763
303. Iyengar SS, Schlegel HB, Millam JM, Voth GA, Scuseria GE, Frisch MJ (2001) Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. II. Generalizations based on mass-weighting, idempotency, energy conservation and choice of initial conditions. *J Chem Phys* 115(22):10291–10302
304. Schlegel HB, Iyengar SS, Li XS, Millam JM, Voth GA, Scuseria GE, Frisch MJ (2002) Ab initio molecular dynamics: propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics. *J Chem Phys* 117(19):8694–8704
305. Iyengar SS, Schlegel HB, Voth GA, Millam JM, Scuseria GE, Frisch MJ (2002) Ab initio molecular dynamics: propagating the density matrix with Gaussian orbitals.

- IV. Formal analysis of the deviations from Born–Oppenheimer dynamics. *Isr J Chem* 42(2–3):191–202
306. Iyengar SS, Schlegel HB, Voth GA (2003) Atom-centered density matrix propagation (ADMP): generalizations using Bohmian mechanics. *J Phys Chem A* 107(37):7269–7277
307. Rega N, Iyengar SS, Voth GA, Schlegel HB, Vreven T, Frisch MJ (2004) Hybrid ab-initio/empirical molecular dynamics: combining the ONIOM scheme with the atom-centered density matrix propagation (ADMP) approach. *J Phys Chem B* 108(13):4210–4220
308. Gonzalez-Lafont A, Truong TN, Truhlar DG (1991) Direct dynamics calculations with neglect of diatomic differential overlap molecular orbital theory with specific reaction parameters. *J Phys Chem* 95(12):4618–4627
309. Lau EY, Kahn K, Bash PA, Bruice TC (2000) The importance of reactant positioning in enzyme catalysis: a hybrid quantum mechanics/molecular mechanics study of a haloalkane dehalogenase. *Proc Natl Acad Sci USA* 97(18):9937–9942
310. Cui Q, Karplus M (2001) Triosephosphate isomerase: a theoretical comparison of alternative pathways. *J Am Chem Soc* 123(10):2284–2290
311. Cui Q, Karplus M (2002) Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: verification of methodology and analysis of reaction mechanisms. *J Phys Chem B* 106(7):1768–1798
312. Alhambra C, Corchado JC, Sánchez ML, Gao JL, Truhlar DG (2000) Quantum dynamics of hydride transfer in enzyme catalysis. *J Am Chem Soc* 122:8197–8203
313. Devi-Kesavan LS, Garcia-Viloca M, Gao J (2003) Semiempirical QM/MM potential with simple valence bond (SVB) for enzyme reactions. Application to the nucleophilic addition reaction in haloalkane dehalogenase. *Theor Chem Acc* 109(3):133–139
314. Corchado JC, Coitiño EL, Chuang YY, Fast PL, Truhlar DG (1998) Interpolated variational transition-state theory by mapping. *J Phys Chem A* 102(14):2424–2438
315. Chuang YY, Corchado JC, Truhlar DG (1999) Mapped interpolation scheme for single-point energy corrections in reaction rate calculations and a critical evaluation of dual-level reaction path dynamics methods. *J Phys Chem A* 103(8):1140–1149
316. Ruiz-Pernía JJ, Silla E, Tuñón I, Martí S, Moliner V (2004) Hybrid QM/MM potentials of mean force with interpolated corrections. *J Phys Chem B* 108(24):8427–8433
317. Rod TH, Ryde U (2005) Quantum mechanical free energy barrier for an enzymatic reaction. *Phys Rev Lett* 94(13):138302/1–4
318. Rod TH, Ryde U (2005) Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase. *J Chem Theory Comput* 1(6):1240–1251
319. McQuarrie DA (2000) Statistical mechanics. University Science Books, Sausalito, CA
320. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22(8):1420–1426
321. Kästner J, Senn HM, Thiel S, Otte N, Thiel W (2006) QM/MM free-energy perturbation compared to thermodynamic integration and umbrella sampling: application to an enzymatic reaction. *J Chem Theory Comput* 2(2):452–461
322. Chandrasekhar J, Smith SE, Jorgensen WL (1984) S_N2 Reaction profiles in the gas phase and aqueous solution. *J Am Chem Soc* 106(10):3049–3050
323. Chandrasekhar J, Smith SE, Jorgensen WL (1985) Theoretical examination of the S_N2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J Am Chem Soc* 107(1):154–163

324. Zheng YJ, Merz KM Jr (1992) Mechanism of the human carbonic anhydrase II catalyzed hydration of carbon dioxide. *J Am Chem Soc* 114(26):10498–10507
325. Stanton RV, Peräkylä M, Bakowies D, Kollman PA (1998) Combined ab initio and free energy calculations to study reactions in enzymes and solution: amide hydrolysis in trypsin and aqueous solution. *J Am Chem Soc* 120(14):3448–3457
326. Bentzien J, Florián J, Glennon TM, Warshel A (1998) Quantum mechanical-molecular mechanical approaches for studying chemical reactions in proteins and solution. In: Gao J, Thompson MA (eds) Combined quantum mechanical and molecular mechanical methods. ACS symposium series, vol 712. American Chemical Society, Washington, DC, pp 16–34
327. Bentzien J, Muller RP, Florián J, Warshel A (1998) Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymic reactions: the nucleophilic attack in subtilisin. *J Phys Chem B* 102:2293–2301
328. Štrajbl M, Hong G, Warshel A (2002) Ab initio QM/MM simulation with proper sampling: “first principle” calculations of the free energy of the autodissociation of water in aqueous solution. *J Phys Chem B* 106:13333–13343
329. Rosta E, Klähn M, Warshel A (2006) Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *J Phys Chem B* 110:2934–2941
330. Li G, Zhang X, Cui Q (2003) Free energy perturbation calculations with combined QM/MM potentials complications, simplifications, and applications to redox potential calculations. *J Phys Chem B* 107(33):8643–8653
331. Li G, Cui Q (2003) pK_a Calculations with QM/MM free energy perturbations. *J Phys Chem B* 107(51):14521–14528
332. Yang W, Bitetti-Putzer R, Karplus M (2004) Chaperoned alchemical free energy simulations: a general method for QM, MM, and QM/MM potentials. *J Chem Phys* 120(20):9450–9453
333. Hu H, Yang W (2005) Dual-topology/dual-coordinate free-energy simulation using QM/MM force field. *J Chem Phys* 123(4):041102/1–4
334. Fixman M (1974) Classical statistical mechanics of constraints: a theorem and application to polymers. *Proc Natl Acad Sci USA* 71:3050–3053
335. Carter EA, Ciccotti G, Hynes JT, Kapral R (1989) Constrained reaction coordinate dynamics for the simulation of rare events. *Chem Phys Lett* 156:472–477
336. Sprik M, Ciccotti G (1998) Free energy from constrained molecular dynamics. *J Chem Phys* 109:7737–7744
337. den Otter WK, Briels WJ (1998) The calculation of free-energy differences by constrained molecular-dynamics simulations. *J Chem Phys* 109:4139–4146
338. den Otter WK (2000) Thermodynamic integration of the free energy along a reaction coordinate in Cartesian coordinates. *J Chem Phys* 112:7283–7292
339. den Otter WK, Briels WJ (2000) Free energy from molecular dynamics with multiple constraints. *Mol Phys* 98(12):773–781
340. Darve E, Pohorille A (2001) Calculating free energies using average force. *J Chem Phys* 115:9169–9183
341. Schlitter J, Klähn M (2003) A new concise expression for the free energy of a reaction coordinate. *J Chem Phys* 118(5):2057–2060
342. Schlitter J, Klähn M (2003) The free energy of a reaction coordinate at multiple constraints: a concise formulation. *Mol Phys* 101(23–24):3439–3443
343. Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K (2001) Steered molecular dynamics investigations of protein function. *J Mol Graphics Model* 19(1):13–25
344. Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 11(2):224–230

345. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78:2690–2693
346. Jarzynski C (1997) Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys Rev E* 56:5018–5035
347. Crooks GE (1998) Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J Stat Phys* 90:1481–1487
348. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys Rev E* 60:2721–2726
349. Park S, Schulten K (2004) Calculating potentials of mean force from steered molecular dynamics simulations. *J Chem Phys* 120(13):5946–5961
350. Cuendet MA (2006) Statistical mechanical derivation of Jarzynski's identity for thermostated non-Hamiltonian dynamics. *Phys Rev Lett* 96(12):120602/1–4
351. Hendrix DA, Jarzynski C (2001) A “fast-growth” method of computing free energy differences. *J Chem Phys* 114:5974–5981
352. Jensen M, Park S, Tajkhorshid E, Schulten K (2002) Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc Natl Acad Sci USA* 99:6731–6736
353. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J Chem Phys* 119:3559–3566
354. Hummer G, Szabo A (2001) Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc Natl Acad Sci USA* 98:3658–3661
355. Jarzynski C (2001) How does a system respond when driven away from thermal equilibrium? *Proc Natl Acad Sci USA* 98:3636–3638
356. Liphardt J, Dumont S, Smith SB, Tinoco I Jr, Bustamante C (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* 296(5574):1832–1835
357. Egolf DA (2002) Statistical mechanics: far from equilibrium. *Science* 296(5574):1813–1815
358. Hummer G (2001) Fast-growth thermodynamic integration: error and efficiency analysis. *J Chem Phys* 114:7330–7337
359. Hu H, Yun RH, Hermans J (2002) Reversibility of free energy simulations: slow growth may have a unique advantage. (With a note on use of Ewald summation). *Mol Simul* 28(1–2):67–80
360. Zuckerman DM, Woolf TB (2002) Theory of a systematic computational error in free energy differences. *Phys Rev Lett* 89(18):180602/1–4
361. Zuckerman DM, Woolf TB (2002) Overcoming finite-sampling errors in fast-switching free-energy estimates: extrapolative analysis of a molecular system. *Chem Phys Lett* 351(5–6):445–453
362. Zuckerman DM, Woolf TB (2004) Systematic finite-sampling inaccuracy in free energy differences and other nonlinear quantities. *J Stat Phys* 114(5–6):1303–1323
363. Ytreberg FM, Zuckerman DM (2004) Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. *J Comput Chem* 25(14):1749–1759
364. Gore J, Ritort F, Bustamante C (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc Natl Acad Sci USA* 100(22):12564–12569
365. Fox RF (2003) Using nonequilibrium measurements to determine macromolecule free-energy differences. *Proc Natl Acad Sci USA* 100(22):12537–12538
366. Rodriguez-Gomez D, Darve E, Pohorille A (2004) Assessing the efficiency of free energy calculation methods. *J Chem Phys* 120(8):3563–3578

367. Lua RC, Grosberg AY (2005) Practical applicability of the Jarzynski relation in statistical mechanics: a pedagogical example. *J Phys Chem B* 109(14):6805–6811
368. Lechner W, Oberhofer H, Dellago C, Geissler PL (2006) Equilibrium free energies from fast-switching trajectories with large time steps. *J Chem Phys* 124(4):044113/1–12
369. Raugei S, Cascella M, Carloni P (2004) A proficient enzyme: insights on the mechanism of orotidine monophosphate decarboxylase from computer simulations. *J Am Chem Soc* 126(48):15730–15737
370. Chandler D (1998) Finding transition pathways: throwing ropes over rough mountain passes, in the dark. In: Berne BJ, Ciccotti G, Coker DF (eds) *Classical and quantum dynamics in condensed phase simulations*. World Scientific, Singapore, pp 51–66
371. Dellago C, Bolhuis PG, Csajka FS, Chandler D (1998) Transition path sampling and the calculation of rate constants. *J Chem Phys* 108(5):1964–1977
372. Dellago C, Bolhuis PG, Chandler D (1998) Efficient transition path sampling: application to Lennard–Jones cluster rearrangements. *J Chem Phys* 108(22):9236–9245
373. Bolhuis PG, Dellago C, Chandler D (1998) Sampling ensembles of deterministic transition pathways. *Faraday Discuss* 110:421–436
374. Csajka FS, Chandler D (1998) Transition pathways in a many-body system: application to hydrogen-bond breaking in water. *J Chem Phys* 109(3):1125–1133
375. Dellago C, Bolhuis PG, Chandler D (1999) On the calculation of reaction rate constants in the transition path ensemble. *J Chem Phys* 110(14):6617–6625
376. Bolhuis PG, Dellago C, Geissler PL, Chandler D (2000) Transition path sampling: throwing ropes over mountains in the dark. *J Phys Condens Matter* 12:A147–A152
377. Geissler PL, Dellago C, Chandler D, Hutter J, Parrinello M (2001) Autoionization in liquid water. *Science* 291(5511):2121–2124
378. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* 53:291–318
379. Dellago C, Bolhuis PG, Geissler PL (2002) Transition path sampling. *Adv Chem Phys* 123:1–78
380. Zahn D (2005) Unprejudiced identification of reaction mechanisms from biased transition path sampling. *J Chem Phys* 123(4):044104/1–7
381. Basner JE, Schwartz SD (2005) How enzyme dynamics helps catalyze a reaction in atomic detail: a transition path sampling study. *J Am Chem Soc* 127:13822–13831
382. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99(20):12562–12566
383. Iannuzzi M, Laio A, Parrinello M (2003) Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys Rev Lett* 90(23):238302/1–4
384. Micheletti C, Laio A, Parrinello M (2004) Reconstructing the density of states by history-dependent metadynamics. *Phys Rev Lett* 92(17):170601/1–4
385. Wu Y, Schmitt JD, Car R (2004) Mapping potential energy surfaces. *J Chem Phys* 121(3):1193–1200
386. Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M (2005) Assessing the accuracy of metadynamics. *J Phys Chem B* 109(14):6714–6721
387. Ensing B, Laio A, Parrinello M, Klein ML (2005) A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J Phys Chem B* 109(14):6676–6687

388. Raiteri P, Laio A, Gervasio FL, Micheletti C, Parrinello M (2006) Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J Phys Chem B* 110(8):3533–3539
389. Bussi G, Laio A, Parrinello M (2006) Equilibrium free energies from nonequilibrium metadynamics. *Phys Rev Lett* 96:090601/1–4
390. Ensing B, De Vivo M, Liu ZW, Moore P, Klein ML (2006) Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc Chem Res* 39(2):73–81
391. Huber T, Torda AE, Gunsteren WF (1994) Local elevation: a method for improving the searching properties of molecular-dynamics simulation. *J Comput-Aided Mol Des* 8(6):695–708
392. Grubmüller H (1995) Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys Rev E* 52(3):2893–2906
393. Müller EM, de Meijere A, Grubmüller H (2002) Predicting unimolecular chemical reactions: chemical flooding. *J Chem Phys* 116(3):897–905
394. Wang F, Landau DP (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys Rev Lett* 86(10):2050–2053
395. Wang F, Landau DP (2001) Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram. *Phys Rev E* 64(5):056101/1–16
396. Landau DP, Wang F (2002) Determining the density of states for classical statistical models by a flat-histogram random walk. *Comput Phys Commun* 147(1–2):674–677
397. Rauegi S, Carloni P (2006) Structure and function of vanadium haloperoxidases. *J Phys Chem B* 110(8):3747–3758
398. Rosso L, Tuckerman ME (2002) An adiabatic molecular dynamics method for the calculation of free energy profiles. *Mol Simul* 28:91–112
399. Rosso L, Mináry P, Zhu Z, Tuckerman ME (2002) On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. *J Chem Phys* 116:4389–4402
400. VandeVondele J, Rothlisberger U (2002) Canonical adiabatic free energy sampling (CAFES): a novel method for the exploration of free energy surfaces. *J Phys Chem B* 106:203–208
401. Piana S, Bucher D, Carloni P, Rothlisberger U (2004) Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations. *J Phys Chem B* 108(30):11139–11149
402. Lu Z, Yang W (2004) Reaction path potential for complex systems derived from combined ab initio quantum mechanical and molecular mechanical calculations. *J Chem Phys* 121(1):89–100
403. Miller WH, Handy NC, Adams JE (1980) Reaction path Hamiltonian for polyatomic molecules. *J Chem Phys* 72(1):99–112
404. Nabuurs SB, Spronk CAEM, Krieger E, Vriend G, Hooft RWW (2004) Protein structures: what good is beauty if it cannot be seen? In: Bultinck P, Tollenaere JP, De Winter H, Langenaeker W (eds) (2004) Computational medicinal chemistry for drug discovery. Dekker, New York, pp 387–403
405. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381(6580):272
406. Word JM (2003) Reduce, V. 2.21. Biochemistry Department, Duke University, Durham, NC, <http://www.kinimage.biochem.duke.edu/software/reduce.php>
407. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J Mol Biol* 285(4):1735–1747

408. Garcia-Viloca M, Poulsen TD, Truhlar DG, Gao JL (2004) Sensitivity of molecular dynamics simulations to the choice of the X-ray structure used to model an enzymatic reaction. *Protein Sci* 13(9):2341–2354
409. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8(1):52–56
410. Forrest LR, Honig B (2005) An assessment of the accuracy of methods for predicting hydrogen positions in protein structures. *Proteins Struct Funct Bioinf* 61(2):296–309
411. Jensen JH, Li H, Robertson AD, Molina PA (2005) Prediction and rationalization of protein pK_a values using QM and QM/MM methods. *J Phys Chem A* 109(30):6634–6643
412. Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pK_a values. *Proteins Struct Funct Bioinf* 61(4):704–721
413. Zhang L, Hermans J (1996) Hydrophilicity of cavities in proteins. *Proteins Struct Funct Genet* 24(4):433–438
414. Olsson MHM, Hong GY, Warshel A (2003) Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin. *J Am Chem Soc* 125(17):5025–5039
415. Altun A, Thiel W (2005) Combined quantum mechanical/molecular mechanical study on the pentacoordinated ferric and ferrous cytochrome P450_{cam} complexes. *J Phys Chem B* 109:1268–1280
416. Pauling L (1946) Molecular architecture and biological reactions. *Chem Eng News* 24:1375–1377
417. Villà J, Warshel A (2001) Energetics and dynamics of enzymatic reactions. *J Phys Chem B* 105(33):7887–7907
418. Kollman PA, Kuhn B, Peräkylä M (2002) Computational studies of enzyme-catalyzed reactions: Where are we in predicting mechanisms and in understanding the nature of enzyme catalysis? *J Phys Chem B* 106(7):1537–1542
419. Gao JL, Truhlar DG (2002) Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem* 53:467–505
420. Martí S, Roca M, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2004) Theoretical insights in enzyme catalysis. *Chem Soc Rev* 33(2):98–107
421. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science* 303(5655):186–195
422. Blake CCF, Mair GA, North ACT, Phillips DC, Sarma VR (1967) On the conformation of the hen egg-white lysozyme molecule. *Proc R Soc London B* 167:365–377
423. Blake CCF, Johnson LN, Mair GA, North ACT, Phillips DC, Sarma VR (1967) Crystallographic studies of the activity of hen egg-white lysozyme. *Proc R Soc London B* 167:378–388
424. Ford LO, Machin PA, Phillips DC, Tjian R, Johnson LN (1974) Crystal structure of a lysozyme-tetrasaccharide lactone complex. *J Mol Biol* 88(2):349–360
425. Tapia O, Andrés J, Safont VS (1994) Enzyme catalysis and transition structures in vacuo. Transition structures for the enolization, carboxylation and oxygenation reactions in ribulose-1,5-bisphosphate carboxylase/oxygenase enzyme (Rubisco). *J Chem Soc, Faraday Trans* 90(16):2365–2374
426. Khanjin NA, Snyder JP, Menger FM (1999) Mechanism of chorismate mutase: contribution of conformational restriction to catalysis in the Claisen rearrangement. *J Am Chem Soc* 121(50):11831–11846
427. Vallee BL, Williams RJ (1968) Metalloenzymes: the entatic nature of their active sites. *Proc Natl Acad Sci USA* 59(2):498–505

428. Williams RJ (1971) The entatic state. *Cold Spring Harb Symp Quant Biol* 36:53–62
429. Williams RJP (1985) Metalloenzyme catalysis: the entatic state. *J Mol Catal* 30(1–2): 1–26
430. Williams RJP (1995) Energized (entatic) states of groups and of secondary structures in proteins and metalloproteins. *Eur J Biochem* 234(2):363–381
431. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* 44(2):98–104
432. Storm DR, Koshland DE (1970) A source for the special catalytic power of enzymes: orbital steering. *Proc Natl Acad Sci USA* 66(2):445–452
433. Mesecar AD, Stoddard BL, Koshland DE (1997) Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* 277(5323):202–206
434. Menger FM (1993) Enzyme reactivity from an organic perspective. *Acc Chem Res* 26(4):206–212
435. Kollman PA, Kuhn B, Donini O, Perakyla M, Stanton R, Bakowies D (2001) Elucidating the nature of enzyme catalysis utilizing a new twist on an old methodology: quantum mechanical-free energy calculations on chemical reactions in enzymes and in aqueous solution. *Acc Chem Res* 34(1):72–79
436. Bruice TC, Benkovic SJ (2000) Chemical basis for enzyme catalysis. *Biochemistry* 39:6267–6274
437. Bruice TC (2002) A view at the millennium: the efficiency of enzymatic catalysis. *Acc Chem Res* 35:139–148
438. Hur S, Bruice TC (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc Natl Acad Sci USA* 100(21):12015–12020
439. Hur S, Bruice TC (2003) Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J Am Chem Soc* 125(6):1472–1473
440. Hur S, Bruice TC (2003) Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the *E. coli* mutase: the efficiency of the enzyme catalysis. *J Am Chem Soc* 125(19):5964–5972
441. Hur S, Bruice TC (2003) Just a near attack conformer for catalysis (chorismate to prephenate rearrangements in water, antibody, enzymes, and their mutants). *J Am Chem Soc* 125(35):10540–10542
442. Zhang X, Zhang X, Bruice TC (2005) A definitive mechanism for chorismate mutase. *Biochemistry* 44(31):10443–10448
443. Repasky MP, Guimarães CRW, Chandrasekhar J, Tirado-Rives J, Jorgensen WL (2003) Investigation of solvent effects for the Claisen rearrangement of chorismate to prephenate: mechanistic interpretation via near attack conformations. *J Am Chem Soc* 125(23):6663–6672
444. Guimarães CRW, Repasky MP, Chandrasekhar J, Tirado-Rives J, Jorgensen WL (2003) Contributions of conformational compression and preferential transition state stabilization to the rate enhancement by chorismate mutase. *J Am Chem Soc* 125(23):6892–6899
445. Ranaghan KE, Mulholland AJ (2004) Conformational effects in enzyme catalysis: QM/MM free energy calculation of the “NAC” contribution in chorismate mutase. *Chem Commun* (10):1238–1239
446. Shurki A, Štrajbl M, Villà J, Warshel A (2002) How much do enzymes really gain by restraining their reacting fragments? *J Am Chem Soc* 124:4097–4107
447. Štrajbl M, Shurki A, Kato M, Warshel A (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J Am Chem Soc* 125:10228–10237

448. Lee JK, Houk KN (1997) A proficient enzyme revisited: the predicted mechanism for orotidine monophosphate decarboxylase. *Science* 276(5314):942–945
449. Devi-Kesavan LS, Gao J (2003) Combined QM/MM study of the mechanism and kinetic isotope effect of the nucleophilic substitution reaction in haloalkane dehalogenase. *J Am Chem Soc* 125(6):1532–1540
450. Page MI, Jencks WP (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci USA* 68(8):1678–1683
451. Blow D (2000) So do we understand how enzymes work? *Structure* 8:R77–R81
452. Villà J, Štrajbl M, Glennon TM, Sham YY, Chu ZT, Warshel A (2000) How important are entropic contributions to enzyme catalysis? *Proc Natl Acad Sci USA* 97:11899–11904
453. Cleland WW, Kreevoy MM (1994) Low-barrier hydrogen-bonds and enzymatic catalysis. *Science* 264(5167):1887–1890
454. Frey PA, Whitt SA, Tobin JB (1994) A low-barrier hydrogen-bond in the catalytic triad of serine proteases. *Science* 264(5167):1927–1930
455. Cleland WW, Frey PA, Gerlt JA (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 273(40):25529–25532
456. Pan YP, McAllister MA (1997) Characterization of low-barrier hydrogen bonds. 1. Microsolvation effects. An ab initio and DFT investigation. *J Am Chem Soc* 119(32):7561–7566
457. Pan YP, McAllister MA (1997) Characterization of low-barrier hydrogen bonds. 5. Microsolvation of enol-enolate. An ab initio and DFT investigation. *J Org Chem* 62(23):8171–8176
458. Pan YP, McAllister MA (1998) Characterization of low-barrier hydrogen bonds. 6. Cavity polarity effects on the formic acid formate anion model system. An ab initio and DFT investigation. *J Am Chem Soc* 120(1):166–169
459. Careri G, Fasella P, Gratton E (1979) Enzyme dynamics: the statistical physics approach. *Annu Rev Biophys Bioeng* 8:69–97
460. Karplus M, McCammon JA (1983) Dynamics of proteins: elements and function. *Annu Rev Biochem* 52:263–300
461. Berendsen HJC, Hayward S (2000) Collective protein dynamics in relation to function. *Curr Opin Struct Biol* 10:165–169
462. Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Enzyme dynamics during catalysis. *Science* 295(5559):1520–1523
463. Olsson MHM, Parson WW, Warshel A (2006) Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis. *Chem Rev* 106:1737–1756
464. Hwang JK, Chu ZT, Yadav A, Warshel A (1991) Simulations of quantum-mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *J Phys Chem* 95(22):8445–8448
465. Kohen A, Klinman JP (1999) Hydrogen tunneling in biology. *Chem Biol* 6(7):R191–R198
466. Sutcliffe MJ, Scrutton NS (2000) Enzyme catalysis: over-the-barrier or through-the-barrier? *Trends Biochem Sci* 25(9):405–408
467. Truhlar DG, Gao J, Garcia-Viloca M, Alhambra C, Corchado J, Sanchez ML, Poulsen TD (2004) Ensemble-averaged variational transition state theory with optimized multidimensional tunneling for enzyme kinetics and other condensed-phase reactions. *Int J Quant Chem* 100(6):1136–1152
468. Zhang X, Houk KN (2005) Why enzymes are proficient catalysts: beyond the Pauling paradigm. *Acc Chem Res* 38(5):379–385

469. Bash PA, Field MJ, Davenport RC, Petsko GA, Ringe D, and Karplus M (1991) Computer-simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* 30:5826–5832
470. Mulholland AJ, Richards WG (1997) Acetyl-CoA enolization in citrate synthase: a quantum mechanical/molecular mechanical (QM/MM) study. *Proteins Struct Funct Genet* 27(1):9–25
471. Liu H, Zhang Y, Yang W (2000) How is the active site of enolase organized to catalyze two different reaction steps? *J Am Chem Soc* 122(28):6560–6570
472. Cui Q, Elstner M, Karplus M (2002) A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *J Phys Chem B* 106(10):2721–2740
473. Hermann JC, Hensen C, Ridder L, Mulholland AJ, Höltje HD (2005) Mechanisms of antibiotic resistance: QM/MM modeling of the acylation reaction of a class A β -lactamase with benzylpenicillin. *J Am Chem Soc* 127(12):4454–4465
474. Cheng Y, Zhang Y, McCammon JA (2005) How does the cAMP-dependent protein kinase catalyze the phosphorylation reaction: an ab initio QM/MM study. *J Am Chem Soc* 127(5):1553–1562
475. Garcia-Viloca M, Truhlar DG, Gao J (2003) Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J Mol Biol* 327(2):549–560
476. Sokalski WA (1985) The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory: derivation of simplified models. *J Mol Catal* 30(3):395–410
477. Szeftczyk B, Mulholland AJ, Ranaghan KE, Sokalski WA (2004) Differential transition-state stabilization in enzyme catalysis: quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J Am Chem Soc* 126(49):16148–16159
478. Roca M, Martí S, Andrés J, Moliner V, Tuñón I, Bertrán J, Williams IH (2003) Theoretical modeling of enzyme catalytic power: analysis of “cratic” and electrostatic factors in catechol *O*-methyltransferase. *J Am Chem Soc* 125(25):7726–7737
479. Ridder L, Palfey BA, Vervoort JIM, Rietjens CM (2000) Modeling flavin and substrate substituent effects on the activation barrier and rate of oxygen transfer by *p*-hydroxybenzoate hydroxylase. *FEBS Lett* 478(1–2):197–201
480. Ridder L, Harvey JN, Rietjens IMCM, Vervoort J, Mulholland AJ (2003) Ab initio QM/MM modeling of the hydroxylation step in *p*-hydroxybenzoate hydroxylase. *J Phys Chem B* 107(9):2118–2126
481. Billeter SR, Hanser CFW, Mordasini TZ, Scholten M, Thiel W, van Gunsteren WF (2001) Molecular dynamics study of oxygenation reactions catalysed by the enzyme *p*-hydroxybenzoate hydroxylase. *Phys Chem Chem Phys* 3:688–695
482. Ridder L, Mulholland AJ, Rietjens IMCM, Vervoort J (2000) A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase. *J Am Chem Soc* 122(36):8728–8738
483. Poulsen TD, Garcia-Viloca M, Gao J, Truhlar DG (2003) Free energy surface, reaction paths, and kinetic isotope effect of short-chain acyl-CoA dehydrogenase. *J Phys Chem B* 107(35):9567–9578
484. Alhambra C, Corchado J, Sánchez ML, Garcia-Viloca M, Gao J, Truhlar DG (2001) Canonical variational theory for enzyme kinetics with the protein mean force and multidimensional quantum mechanical tunneling dynamics. Theory and application to liver alcohol dehydrogenase. *J Phys Chem B* 105:11326–11340
485. Tresadern G, Faulder PF, Gleeson MP, Tai Z, MacKenzie G, Burton NA, Hillier IH (2003) Recent advances in quantum mechanical/molecular mechanical calculations

- of enzyme catalysis: hydrogen tunnelling in liver alcohol dehydrogenase and inhibition of elastase by α -ketoheterocycles. *Theor Chem Acc* 109(3):108–117
486. Walker RC, de Souza MM, Mercer IP, Gould IR, Klug DR (2002) Large and fast relaxations inside a protein: calculation and measurement of reorganization energies in alcohol dehydrogenase. *J Phys Chem B* 106(44):11658–11665
487. Formanec MS, Li G, Zhang X, Cui Q (2002) Calculating accurate redox potentials in enzymes with a combined QM/MM free energy perturbation approach. *J Theor Comput Chem* 1(1):53–67
488. Wymore T, Hempel J, Cho SS, MacKerell AD Jr, Nicholas HB Jr, Deerfield DW II (2004) Molecular recognition of aldehydes by aldehyde dehydrogenase and mechanism of nucleophile activation. *Proteins Struct Funct Bioinf* 57(4):758–771
489. Wymore T, Deerfield DW II, Field MJ, Hempel J, Nicholas HB Jr (2003) Initial catalytic events in class 3 aldehyde dehydrogenase: MM and QM/MM simulations. *Chem-Biol Interact* 143–144:75–84
490. Moliner V, Williams IH (2000) Flexible QM/MM modelling embraces alternative mechanisms for lactate dehydrogenase. *Chem Commun* (19):1843–1844
491. Ferrer S, Silla E, Tuñón I, Oliva M, Moliner V, Williams IH (2005) Dependence of enzyme reaction mechanism on protonation state of titratable residues and QM level description: lactate dehydrogenase. *Chem Commun* (47):5873–5875
492. Ferrer S, Ruiz-Pernía JJ, Tuñón I, Moliner V, Garcia-Viloca M, González-Lafont A, Lluch JM (2005) A QM/MM exploration of the potential energy surface of pyruvate to lactate transformation catalyzed by LDH. Improving the accuracy of semiempirical descriptions. *J Chem Theory Comput* 1(4):750–761
493. Cummins PL, Gready JE (2000) QM/MM and SCRF studies of the ionization state of 8-methylpterin substrate bound to dihydrofolate reductase: existence of a low-barrier hydrogen bond. *J Mol Graphics Model* 18(1):42–49
494. Cummins PL, Gready JE (2000) Combined quantum and molecular mechanics (QM/MM) study of the ionization state of 8-methylpterin substrate bound to dihydrofolate reductase. *J Phys Chem B* 104(18):4503–4510
495. Beierlein F, Lanig H, Schürer G, Horn AHC, Clark T (2003) Quantum mechanical/molecular mechanical (QM/MM) docking: an evaluation for known test systems. *Mol Phys* 101(15):2469–2480
496. Titmuss SJ, Cummins PL, Bliznyuk AA, Rendell AP, Gready JE (2000) Comparison of linear-scaling semiempirical methods and combined quantum mechanical/molecular mechanical methods applied to enzyme reactions. *Chem Phys Lett* 320(1–2):169–176
497. Cummins PL, Greatbanks SP, Rendell AP, Gready JE (2002) Computational methods for the study of enzymic reaction mechanisms. 1. Application to the hydride transfer step in the catalysis of dihydrofolate reductase. *J Phys Chem B* 106(38):9934–9944
498. Titmuss SJ, Cummins PL, Rendell AP, Bliznyuk AA, Gready JE (2002) Comparison of linear-scaling semiempirical methods and combined quantum mechanical/molecular mechanical methods for enzymic reactions. II. An energy decomposition analysis. *J Comput Chem* 23(14):1314–1322
499. Cummins PL, Gready JE (2003) Computational methods for the study of enzymic reaction mechanisms. II. An overlapping mechanically embedded method for hybrid semi-empirical-QM/MM calculations. *THEOCHEM* 632(1–3):247–257
500. Garcia-Viloca M, Truhlar DG, Gao J (2003) Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase. *Biochemistry* 42(46):13558–13575

501. Thorpe IF, Brooks CL III (2003) Barriers to hydride transfer in wild type and mutant dihydrofolate reductase from *E. coli*. *J Phys Chem B* 107(50):14042–14051
502. Cummins PL, Gready JE (2005) Computational methods for the study of enzymic reaction mechanisms III: a perturbation plus QM/MM approach for calculating relative free energies of protonation. *J Comput Chem* 26(6):561–568
503. Ferrer S, Silla E, Tuñón I, Martí S, Moliner V (2003) Catalytic mechanism of dihydrofolate reductase enzyme. A combined quantum-mechanical/molecular-mechanical characterization of the N5 protonation step. *J Phys Chem B* 107(50):14036–14041
504. Proust-De Martin F, Dumas R, Field MJ (2000) A hybrid-potential free-energy study of the isomerization step of the acetohydroxy acid isomeroreductase reaction. *J Am Chem Soc* 122(32):7688–7697
505. Alhambra C, Sánchez ML, Corchado J, Gao J, Truhlar DG (2001) Quantum mechanical tunneling in methylamine dehydrogenase. *Chem Phys Lett* 347:512–518
506. Faulder PF, Tresadern G, Chohan KK, Scrutton NS, Sutcliffe MJ, Hillier IH, Burton NA (2001) QM/MM studies show substantial tunneling for the hydrogen-transfer reaction in methylamine dehydrogenase. *J Am Chem Soc* 123(35):8604–8605
507. Tresadern G, Nunez S, Faulder PF, Wang H, Hillier IH, Burton NA (2003) Direct dynamics calculations of reaction rate and kinetic isotope effects in enzyme catalysed reactions. *Faraday Discuss* 122:223–242
508. Tresadern G, Wang H, Faulder PF, Burton NA, Hillier IH (2003) Extreme tunnelling in methylamine dehydrogenase revealed by hybrid QM/MM calculations: potential energy surface profile for methylamine and ethanolamine substrates and kinetic isotope effect values. *Mol Phys* 101(17):2775–2784
509. Torrent M, Vreven T, Musaev DG, Morokuma K, Farkas Ö, Schlegel HB (2002) Effects of the protein environment on the structure and energetics of active sites of metalloenzymes. ONIOM study of methane monooxygenase and ribonucleotide reductase. *J Am Chem Soc* 124(2):192–193
510. Hoffmann M, Khavrutskii IV, Musaev DG, Morokuma K (2004) Protein effects on the O₂ binding to the active site of the methane monooxygenase: ONIOM studies. *Int J Quant Chem* 99(6):972–980
511. Friesner RA, Baik M-H, Ghorman BF, Guallar V, Wirstam M, Murphy RB, Lippard SJ (2003) How iron-containing proteins control dioxygen chemistry: a detailed atomic level description via accurate quantum chemical and mixed quantum mechanics/molecular mechanics calculations. *Coord Chem Rev* 238–239:267–290
512. Ghorman BF, Lippard SJ, Friesner RA (2005) Substrate hydroxylation in methane monooxygenase: quantitative modeling via mixed quantum mechanics/molecular mechanics techniques. *J Am Chem Soc* 127:1025–1037
513. Shiota Y, Yoshizawa K (2004) QM/MM study of the mononuclear non-heme iron active site of phenylalanine hydroxylase. *J Phys Chem B* 108(44):17226–17237
514. Nemukhin AV, Grigorenko BL, Topol IA, Burt SK (2006) Modeling dioxygen binding to the non-heme iron-containing enzymes. *Int J Quant Chem* 106(10):2184–2190
515. Bathelt CM, Mulholland AJ, Harvey JN (2005) QM/MM studies of the electronic structure of the compound I intermediate in cytochrome c peroxidase and ascorbate peroxidase. *Dalton Trans* (21):3470–3476
516. Schöneboom JC, Thiel W (2004) The resting state of P450_{cam}: a QM/MM study. *J Phys Chem B* 108(22):7468–7478
517. Swart M, Groenhof AR, Ehlers AW, Lammertsma K (2005) Substrate binding in the active site of cytochrome P450CAM. *Chem Phys Lett* 403(1–3):35–41
518. Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, Shaik S (2002) The elusive oxidant species of cytochrome P450 enzymes: characterization by combined

- quantum mechanical/molecular mechanical (QM/MM) calculations. *J Am Chem Soc* 124(27):8142–8151
519. Bathelt CM, Zurek J, Mulholland AJ, Harvey JN (2005) Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. *J Am Chem Soc* 127:12900–12908
520. Cohen S, Kumar D, Shaik S (2006) In silico design of a mutant of cytochrome P450 containing selenocysteine. *J Am Chem Soc* 128:2649–2653
521. Schöneboom JC, Neese F, Thiel W (2005) Toward identification of the compound I reactive intermediate in cytochrome P450 chemistry: a QM/MM study of its EPR and Mössbauer parameters. *J Am Chem Soc* 127(16):5840–5853
522. Schöneboom JC, Cohen S, Lin H, Shaik S, Thiel W (2004) Quantum mechanical/molecular mechanical investigation of the mechanism of C–H hydroxylation of camphor by cytochrome P450_{cam}: theory supports a two-state rebound mechanism. *J Am Chem Soc* 126(12):4017–4034
523. Altun A, Guallar V, Friesner RA, Shaik S, Thiel W (2006) The effect of heme environment on the hydrogen abstraction reaction of camphor in P450_{cam} catalysis: a QM/MM study. *J Am Chem Soc* 128:3924–3925
524. Lin H, Schöneboom JC, Cohen S, Shaik S, Thiel W (2004) QM/MM study of the product-enzyme complex in P450_{cam} catalysis. *J Phys Chem B* 108(28):10083–10088
525. Guallar V, Friesner RA (2004) Cytochrome P450CAM enzymatic catalysis cycle: a quantum mechanics/molecular mechanics study. *J Am Chem Soc* 126(27):8501–8508
526. Shaik S, Kumar D, de Visser SP, Altun A, Thiel W (2005) Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem Rev* 105(6):2279–2328
527. Derat E, Cohen S, Shaik S, Altun A, Thiel W (2005) Principal active species of horseradish peroxidase, compound I: a hybrid quantum mechanical/molecular mechanical study. *J Am Chem Soc* 127(39):13611–13621
528. Kamachi T, Yoshizawa K (2005) Water-assisted oxo mechanism for heme metabolism. *J Am Chem Soc* 127(30):10686–10692
529. Martí MA, Crespo A, Bari SE, Doctorovich FA, Estrin DA (2004) QM–MM study of nitrite reduction by nitrite reductase of *Pseudomonas aeruginosa*. *J Phys Chem B* 108(46):18073–18080
530. Källrot N, Nilsson K, Rasmussen T, Ryde U (2005) Theoretical study of structure of catalytic copper site in nitrite reductase. *Int J Quant Chem* 102:520–541
531. Morao I, Tai Z, Hillier IH, Burton NA (2003) What form of N⁸-hydroxy-L-arginine is the intermediate in the mechanism of NO synthase? QM and QM/MM calculations of substrate-active site interactions. *THEOCHEM* 632(1–3):277–285
532. Fernández ML, Martí MA, Crespo A, Estrin DA (2005) Proximal effects in the modulation of nitric oxide synthase reactivity: a QM-MM study. *J Biol Inorg Chem* 10:595–604
533. Crespo A, Martí MA, Kalko SG, Morreale A, Orozco M, Gelpi JL, Luque FJ, Estrin DA (2005) Theoretical study of the truncated hemoglobin HbN: exploring the molecular basis of the NO detoxification mechanism. *J Am Chem Soc* 127(12):4433–4444
534. Martí MA, Crespo A, Capece L, Boechi L, Bikiel DE, Scherlis DA, Estrin DA (2006) Dioxygen affinity in heme proteins investigated by computer simulation. *J Inorg Biochem* 100(4):761–770
535. Wirstam M, Lippard SJ, Friesner RA (2003) Reversible dioxygen binding to hemerythrin. *J Am Chem Soc* 125(13):3980–3987

536. Rovira C, Schulze B, Eichinger M, Evanseck JD, Parrinello M (2001) Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: a QM/MM density functional study. *Biophys J* 81(1):435–445
537. Freindorf M, Shao Y, Brown ST, Kong J, Furlani TR (2006) A combined density functional theory and molecular mechanics (QM/MM) study of FeCO vibrations in carbonmonoxy myoglobin. *Chem Phys Lett* 419(4–6):563–566
538. Rovira C (2003) Applications of Car-Parrinello molecular dynamics in biochemistry – binding of ligands in myoglobin. In: Carloni P, Alber F (eds) *Quantum medicinal chemistry. Methods and principles in medicinal chemistry*, vol 17. Wiley, Weinheim, pp 73–112
539. Harvey JN (2004) Spin-forbidden CO ligand recombination in myoglobin. *Faraday Discuss* 127:165–177
540. Guallar V, Jarzecki AA, Friesner RA, Spiro TG (2006) Modeling of ligation-induced helix/loop displacements in myoglobin: toward an understanding of hemoglobin allostery. *J Am Chem Soc* 128(16):5427–5435
541. Strickland N, Mulholland AJ, Harvey JN (2006) The Fe–CO bond energy in myoglobin: a QM/MM study of the effect of tertiary structure. *Biophys J* 90(4):L27–L29
542. Kravitz JY, Pecoraro VL, Carlson HA (2005) Quantum mechanics/molecular mechanics calculations of the vanadium dependent chloroperoxidase. *J Chem Theory Comput* 1(6):1265–1274
543. Rulíšek L, Solomon EI, Ryde U (2005) A combined quantum and molecular mechanical study of the O₂ reductive cleavage in the catalytic cycle of multicopper oxidases. *Inorg Chem* 44(16):5612–5628
544. Rothlisberger U, Carloni P, Doclo K, Parrinello M (2000) A comparative study of galactose oxidase and active site analogs based on QM/MM Car-Parrinello simulations. *J Biol Inorg Chem* 5(2):236–250
545. Kamachi T, Kihara N, Shiota Y, Yoshizawa K (2005) Computational exploration of the catalytic mechanism of dopamine β -monooxygenase: modeling of its mononuclear copper active sites. *Inorg Chem* 44(12):4226–4236
546. Yoshizawa K, Kihara N, Kamachi T, Shiota Y (2006) Catalytic mechanism of dopamine β -monooxygenase mediated by Cu(III)–Oxo. *Inorg Chem* 45(7):3034–3041
547. Sundararajan M, Hillier IH, Burton NA (2006) Structure and redox properties of the protein, rubredoxin, and its ligand and metal mutants studied by electronic structure calculation. *J Phys Chem A* 110(2):785–790
548. Magistrato A, DeGrado WF, Laio A, Rothlisberger U, VandeVondele J, Klein ML (2003) Characterization of the dizinc analogue of the synthetic diiron protein DF1 using ab initio and hybrid quantum/classical molecular dynamics simulations. *J Phys Chem B* 107(17):4182–4188
549. Ryde U, Olsson MHM (2001) Structure, strain, and reorganization energy of blue copper models in the protein. *Int J Quant Chem* 81(5):335–347
550. Moon S, Patchkovskii S, Salahub DR (2003) QM/MM calculations of EPR hyperfine coupling constants in blue copper proteins. *THEOCHEM* 632:287–295
551. Datta SN, Sudhamsu J, Pandey A (2004) Theoretical determination of the standard reduction potential of plastocyanin in vitro. *J Phys Chem B* 108(23):8007–8016
552. Prabhakar R, Musaei DG, Khavrutskii IV, Morokuma K (2004) Does the active site of mammalian glutathione peroxidase (GPx) contain water molecules? An ONIOM study. *J Phys Chem B* 108(34):12643–12645
553. Bergès J, Rickards G, Rauk A, Houée-Levin C (2006) QM/MM study of electron addition on protein disulfide bonds. *Chem Phys Lett* 421:63–67

554. Chu J-W, Brooks BR, Trout BL (2004) Oxidation of methionine residues in aqueous solutions: free methionine and methionine in granulocyte colony-stimulating factor. *J Am Chem Soc* 126(50):16601–16607
555. Ridder L, Rietjens IMCM, Vervoort J, Mulholland AJ (2002) Quantum mechanical/molecular mechanical free energy simulations of the glutathione *S*-transferase (M1-1) reaction with phenanthrene 9,10-oxide. *J Am Chem Soc* 124(33):9926–9936
556. Ruggiero GD, Williams IH, Roca M, Moliner V, Tuñón I (2004) QM/MM determination of kinetic isotope effects for COMT-catalyzed methyl transfer does not support compression hypothesis. *J Am Chem Soc* 126(28):8634–8635
557. Roca M, Moliner V, Ruiz-Pernía JJ, Silla E, Tuñón I (2006) Activation free energy of catechol *O*-methyltransferase. Corrections to the potential of mean force. *J Phys Chem A* 110(2):503–509
558. Roca M, Andrés J, Moliner V, Tuñón I, Bertrán J (2005) On the nature of the transition state in catechol *O*-methyltransferase. A complementary study based on molecular dynamics and potential energy surface explorations. *J Am Chem Soc* 127(30):10648–10655
559. Hu P, Zhang Y (2006) Catalytic mechanism and product specificity of the histone lysin methyltransferase SET7/9: an ab initio QM/MM-FE study with multiple initial structures. *J Am Chem Soc* 128(4):1272–1278
560. Núñez S, Antoniou D, Schramm VL, Schwartz SD (2004) Promoting vibrations in human purine nucleoside phosphorylase. A molecular dynamics and hybrid quantum mechanical/molecular mechanical study. *J Am Chem Soc* 126(48):15720–15729
561. Sheppard DW, Burton NA, Hillier IH (2000) Ab initio hybrid quantum mechanical/molecular mechanical studies of the mechanisms of the enzymes protein kinase and thymidine phosphorylase. *THEOCHEM* 506:35–44
562. Gao X-F, Huang X-R, Sun C-C (2006) Role of each residue in catalysis in the active site of pyrimidine nucleoside phosphorylase from *Bacillus subtilis*: a hybrid QM/MM study. *J Struct Biol* 154(1):20–26
563. Cheng Y, Zhang Y, McCammon JA (2006) How does activation loop phosphorylation modulate catalytic activity in the cAMP-dependent protein kinase: a theoretical study. *Protein Sci* 15(4):672–683
564. Rungtongmongkol T, Hannongbua S, Mulholland A (2004) Mechanistic study of HIV-1 reverse transcriptase at the active site based on QM/MM method. *J Theor Comput Chem* 3(4):491–500
565. Thomas A, Field MJ (2002) Reaction mechanism of the HGXPRTase from *Plasmodium falciparum*: a hybrid potential quantum mechanical/molecular mechanical study. *J Am Chem Soc* 124(42):12432–12438,
566. Mulholland AJ, Lyne PD, Karplus M (2000) Ab initio QM/MM study of the citrate synthase mechanism. A low-barrier hydrogen bond is not involved. *J Am Chem Soc* 122(3):534–535
567. Kurz LC, Fite B, Jean J, Park J, Erpelding T, Callis P (2005). Photophysics of tryptophan fluorescence: link with the catalytic strategy of the citrate synthase from *Thermoplasma acidophilum*. *Biochemistry* 44(5):1394–1413
568. Senn HM, O'Hagan D, Thiel W (2005) Insight into enzymatic C–F bond formation from QM and QM/MM calculations. *J Am Chem Soc* 127:13643–13655
569. Schürer G, Lanig H, Clark T (2004) *Aeromonas proteolytica* aminopeptidase: an investigation of the mode of action using a quantum mechanical/molecular mechanical approach. *Biochemistry* 43(18):5414–5427
570. Schürer G, Horn AHC, Gedeck P, Clark T (2002) The reaction mechanism of bovine lens leucine aminopeptidase. *J Phys Chem B* 106(34):8815–8830

571. Klein CDP, Schiffmann R, Folkers G, Piana S, Röhrlisberger U (2003) Protonation states of methionine aminopeptidase and their relevance for inhibitor binding and catalytic activity. *J Biol Chem* 278(48):47862–47867
572. Cross JB, Vreven T, Meroueh SO, Mobashery S, Schlegel HB (2005) Computational investigation of irreversible inactivation of the zinc-dependent protease carboxypeptidase A. *J Phys Chem B* 109(10):4761–4769
573. Phoon L, Burton NA (2005) Assessment of a mechanism for reactive inhibition of carboxypeptidase A with QM/MM methods. *J Mol Graphics Model* 24(2):94–101
574. Molina PA, Sikorski RS, Jensen JH (2003) NMR chemical shifts in the low-pH form of α -chymotrypsin. A QM/MM and ONIOM-NMR study. *Theor Chem Acc* 109(3):100–107
575. Molina PA, Jensen JH (2003) A predictive model of strong hydrogen bonding in proteins: the $N^{\delta 1}$ -H- $O^{\delta 1}$ hydrogen bond in low-pH α -chymotrypsin and α -lytic protease. *J Phys Chem B* 107(25):6226–6233
576. Pieraccini S, Sironi M, Colombo G (2006) Modeling enzymatic processes: a molecular simulation analysis of the origins of regioselectivity. *Chem Phys Lett* 418:373–376
577. Antonczak S, Monard G, Ruiz-López M, Rivail JL (2000) Insights in the peptide hydrolysis mechanism by thermolysin: a theoretical QM/MM study. *J Mol Model* 6:527–538
578. Nemukhin AV, Grigorenko BL, Rogov AV, Topol IA, Burt SK (2004) Modeling of serine protease prototype reactions with the flexible effective fragment potential quantum mechanical/molecular mechanical method. *Theor Chem Acc* 111(1):36–48
579. Ishida T, Kato S (2003) Theoretical perspectives on the reaction mechanism of serine proteases: the reaction free energy profiles of the acylation process. *J Am Chem Soc* 125(39):12035–12048
580. Ishida T, Kato S (2004) Role of Asp102 in the catalytic relay system of serine proteases: a theoretical study. *J Am Chem Soc* 126(22):7111–7118
581. Ishida T (2006) Low-barrier hydrogen bond hypothesis in the catalytic triad residue of serine proteases: correlation between structural rearrangement and chemical shifts in the acylation process. *Biochemistry* 45(17):5413–5420
582. Gräter F, Schwarzl SM, Dejaegere A, Fischer S, Smith JC (2005) Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J Phys Chem B* 109:10474–10483
583. Topf M, Várnai P, Richards WG (2002) Ab initio QM/MM dynamics simulation of the tetrahedral intermediate of serine proteases: insights into the active site hydrogen-bonding network. *J Am Chem Soc* 124(49):14780–14788
584. Topf M, Richards WG (2004) Theoretical studies on the deacylation step of serine protease catalysis in the gas phase, in solution, and in elastase. *J Am Chem Soc* 126(44):14631–14641
585. Gleeson MP, Hillier IH, Burton NA (2004) Theoretical analysis of peptidyl α -keto-heterocyclic inhibitors of human neutrophil elastase: insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. *Org Biomol Chem* 2(16):2275–2280
586. Guo H, Wlodawer A, Guo H (2005) A general acid-base mechanism for the stabilization of a tetrahedral adduct in a serine-carboxyl peptidase: a computational study. *J Am Chem Soc* 127:15662–15663
587. Xu Q, Guo H, Wlodawer A, Guo H (2006) The importance of dynamics in substrate-assisted catalysis and specificity. *J Am Chem Soc* 128(18):5994–5995

588. Hensen C, Hermann JC, Nam K, Ma S, Gao J, Höltje H-D (2004) A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J Med Chem* 47(27):6673–6680
589. Sulpizi M, Laio A, VandeVondele J, Cattaneo A, Rothlisberger U, Carloni P (2003) Reaction mechanism of caspases: insights from QM/MM Car–Parrinello simulations. *Proteins Struct Funct Genet* 52(2):212–224
590. Corminboeuf C, Hu P, Tuckerman ME, Zhang Y (2006) Unexpected deacetylation mechanism suggested by a density functional theory QM/MM study of histone-deacetylase-like protein. *J Am Chem Soc* 128(14):4530–4531
591. Madison V, Duca J, Bennett F, Bohanon S, Cooper A, Chu M, Desai J, Girijavallabhan V, Hare R, Hruza A, Hendrata S, Huang Y, Kravec C, Malcolm B, McCormick J, Miesel L, Ramanathan L, Reichert P, Saksena A, Wang J, Weber PC, Zhu H, Fischmann T (2002) Binding affinities and geometries of various metal ligands in peptide deformylase inhibitors. *Biophys Chem* 101–102:239–247
592. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S (2005) A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J Med Chem* 48(17):5437–5447
593. Pitarch J, Pascual-Ahuir J-L, Silla E, Tuñón I (2000) A quantum mechanics/molecular mechanics study of the acylation reaction of TEM1 β -lactamase and penicillanate. *J Chem Soc, Perkin Trans 2* (4):761–767
594. Díaz N, Suárez D, Sordo TL, Merz KM Jr (2001) Acylation of class A β -lactamases by penicillins: a theoretical examination of the role of serine 130 and the β -lactam carboxylate group. *J Phys Chem B* 105(45):11302–11313
595. Hermann JC, Ridder L, Mulholland AJ, Höltje H-D (2003) Identification of Glu166 as the general base in the acylation reaction of class A β -lactamases through QM/MM modeling. *J Am Chem Soc* 125(32):9590–9591
596. Castillo R, Silla E, Tuñón I (2002) Role of protein flexibility in enzymatic catalysis: quantum mechanical-molecular mechanical study of the deacylation reaction in class A β -lactamases. *J Am Chem Soc* 124(8):1809–1816
597. Meroueh SO, Fisher JE, Schlegel HB, Mobashery S (2005) Ab initio QM/MM study of class A β -lactamase acylation: dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70. *J Am Chem Soc* 127:15397–15407
598. Hermann JC, Ridder L, Höltje H-D, Mulholland AJ (2006) Molecular mechanisms of antibiotic resistance: QM/MM modelling of deacylation in a class A β -lactamase. *Org Biomol Chem* 4:206–210
599. Gherman BF, Goldberg SD, Cornish VW, Friesner RA (2004) Mixed quantum mechanical/molecular mechanical (QM/MM) study of the deacylation reaction in a penicillin binding protein (PBP) versus in a class C β -lactamase. *J Am Chem Soc* 126(24):7652–7664
600. Olsen L, Rasmussen T, Hemmingsen L, Ryde U (2004) Binding of benzylpenicillin to metallo- β -lactamase: a QM/MM study. *J Phys Chem B* 108(45):17639–17648
601. Gu W, Zhu J, Liu H (2002) Different protonation states of the *Bacillus cereus* binuclear zinc metallo- β -lactamase active site studied by combined quantum mechanical and molecular mechanical simulations. *J Theor Comput Chem* 1(1):69–80
602. Dal Peraro M, Llarrull LI, Rothlisberger U, Vila AJ, Carloni P (2004) Water-assisted reaction mechanism of monozinc β -lactamases. *J Am Chem Soc* 126(39):12661–12668
603. Xu D, Zhou Y, Xie D, Guo H (2005) Antibiotic binding to monozinc CphA β -lactamase from *Aeromonas hydrophila*: quantum mechanical/molecular mechanical and density functional theory studies. *J Med Chem* 48(21):6679–6689

604. Park H, Brothers EN, Merz KM Jr (2005) Hybrid QM/MM and DFT investigations of the catalytic mechanism and inhibition of the dinuclear zinc metallo- β -lactamase CcrA from *Bacteroides fragilis*. *J Am Chem Soc* 127(12):4232–4241
605. Li J, Cross JB, Vreven T, Meroueh SO, Mobashery S, Schlegel HB (2005) Lysine carboxylation in proteins: OXA-10 β -lactamase. *Proteins Struct Funct Bioinf* 61(2):246–257
606. Oliva M, Dideberg O, Field MJ (2003) Understanding the acylation mechanisms of active-site serine penicillin-recognizing proteins: a molecular dynamics simulation study. *Proteins Struct Funct Bioinf* 53(1):88–100
607. Díaz N, Suárez D, Sordo TL, Merz KM Jr (2001) A theoretical study of the aminolysis reaction of lysine 199 of human serum albumin with benzylpenicillin: consequences for immunochemistry of penicillins. *J Am Chem Soc* 123(31):7574–7583
608. Lodola A, Mor M, Hermann JC, Tarzia G, Piomelli D, Mulholland AJ (2005) QM/MM modelling of oleamide hydrolysis in fatty acid amide hydrolase (FAAH) reveals a new mechanism of nucleophile activation. *Chem Commun* (35):4399–4401
609. Zhang Y, Kua J, McCammon JA (2002) Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an ab initio QM/MM study. *J Am Chem Soc* 124(35):10572–10577
610. Hurley MM, Wright JB, Lushington GH, White WE (2003) Quantum mechanics and mixed quantum mechanics/molecular mechanics simulations of model nerve agents with acetylcholinesterase. *Theor Chem Acc* 109(3):160–168
611. Gao D, Zhan C-G (2005) Modeling evolution of hydrogen bonding and stabilization of transition states in the process of cocaine hydrolysis catalyzed by human butyrylcholinesterase. *Proteins Struct Funct Bioinf* 62:99–110
612. Zhan C-G, Gao D (2005) Catalytic mechanism and energy barriers for butyrylcholinesterase-catalyzed hydrolysis of cocaine. *Biophys J* 89:3863–3872
613. Funke SA, Otte N, Eggert T, Bocola M, Jaeger K-E, Thiel W (2005) Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Eng Des Sel* 18(11):509–514
614. Dinner AR, Blackburn GM, Karplus M (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* 413(6857):752–755
615. Banerjee A, Yang W, Karplus M, Verdine GL (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 434:612–618
616. Biarnés X, Nieto J, Planas A, Rovira C (2006) Substrate distortion in the Michaelis complex of bacillus 1,3–1,4- β -glucanase: insight from first principles molecular dynamics simulations. *J Biol Chem* 281:1432–1441
617. Dittrich M, Hayashi S, Schulten K (2003) On the mechanism of ATP hydrolysis in F_1 -ATPase. *Biophys J* 85(4):2253–2266
618. Dittrich M, Hayashi S, Schulten K (2004) ATP hydrolysis in the β_{TP} and β_{DP} catalytic sites of F_1 -ATPase. *Biophys J* 87(5):2954–2967
619. Li G, Cui Q (2004) Mechanochemical coupling in myosin: a theoretical analysis with molecular dynamics and combined QM/MM reaction path calculations. *J Phys Chem B* 108(10):3342–3357
620. Lopez X, York DM, Dejaegere A, Karplus M (2002) Theoretical studies on the hydrolysis of phosphate diesters in the gas phase, solution, and RNase A. *Int J Quant Chem* 86(1):10–26
621. Topol IA, Cachau RE, Nemukhin AV, Grigorenko BL, Burt SK (2004) Quantum chemical modeling of the GTP hydrolysis by the RAS-GAP protein complex. *Biochim Biophys Acta* 1700(1):125–136

622. Grigorenko BL, Nemukhin AV, Topol IA, Cachau RE, Burt SK (2005) QM/MM modeling the Ras-GAP catalyzed hydrolysis of guanosine triphosphate. *Proteins Struct Funct Bioinf* 60(3):495–503
623. Klähn M, Schlitter J, Gerwert K (2005) Theoretical IR spectroscopy based on QM/MM calculations provides changes in charge distribution, bond lengths, and bond angles of the GTP ligand induced by the Ras-protein. *Biophys J* 88(6):3829–3844
624. De Vivo M, Ensing B, Klein ML (2005) Computational study of phosphatase activity in soluble epoxide hydrolase: high efficiency through a water bridge mediated proton shuttle. *J Am Chem Soc* 127(32):11226–11227
625. Gleeson MP, Burton NA, Hillier IH (2003) Prediction of the potency of inhibitors of adenosine deaminase by QM/MM calculations. *Chem Commun* (17):2180–2181
626. Gleeson MP, Burton NA, Hillier IH (2003) The mechanism of adenosine deaminase catalysis studied by QM/MM calculations: the role of histidine 238 and the activity of the alanine 238 mutant. *Phys Chem Chem Phys* 5(19):4272–4278
627. Xu Q, Guo H (2004) Quantum mechanical/molecular mechanical molecular dynamics simulations of cytidine deaminase: from stabilization of transition state analogues to catalytic mechanisms. *J Phys Chem B* 108(7):2477–2483
628. Guo H, Rao N, Xu Q, Guo H (2005) Origin of tight binding of a near-perfect transition-state analogue by cytidine deaminase: implications for enzyme catalysis. *J Am Chem Soc* 127(9):3191–3197
629. Xu D, Guo H, Gao J, Cui Q (2004) A QM/MM study of a nucleophilic aromatic substitution reaction catalyzed by 4-chlorobenzoyl-CoA dehalogenase. *Chem Commun* (7):892–893
630. Xu D, Wei Y, Wu J, Dunaway-Mariano D, Guo H, Cui Q, Gao J (2004) QM/MM studies of the enzyme-catalyzed dechlorination of 4-chlorobenzoyl-CoA provide insight into reaction energetics. *J Am Chem Soc* 126(42):13649–13658
631. Wu J, Xu D, Lu X, Wang C, Guo H, Dunaway-Mariano D (2006) Contributions of long-range electrostatic interactions to 4-chlorobenzoyl-CoA dehalogenase catalysis: a combined theoretical and experimental study. *Biochemistry* 45:102–112
632. Soriano A, Silla E, Tuñón I, Martí S, Moliner V, Bertrán J (2004) Electrostatic effects in enzyme catalysis: a quantum mechanics/molecular mechanics study of the nucleophilic substitution reaction in haloalkane dehalogenase. *Theor Chem Acc* 112(4):327–334
633. Nam K, Prat-Resina X, Garcia-Viloca M, Devi-Kesavan LS, Gao J (2004) Dynamics of an enzymatic substitution reaction in haloalkane dehalogenase. *J Am Chem Soc* 126(5):1369–1376
634. Zhang X, Harrison DHT, Cui Q (2002) Functional specificities of methylglyoxal synthase and triosephosphate isomerase: a combined QM/MM analysis. *J Am Chem Soc* 124(50):14871–14878
635. Guimarães CRW, Udier-Blagović M, Jorgensen WL (2005) Macrophomate synthase: QM/MM simulations address the Diels-Alder versus Michael-aldol reaction mechanism. *J Am Chem Soc* 127(10):3577–3588
636. Sicinska D, Truhlar DG, Paneth P (2005) Dependence of transition state structure on substrate: the intrinsic C-13 kinetic isotope effect is different for physiological and slow substrates of the ornithine decarboxylase reaction because of different hydrogen bonding structures. *J Am Chem Soc* 127(15):5414–5422
637. Lundberg M, Blomberg MRA, Siegbahn PEM (2004) Developing active site models of ODCase – from large quantum models to a QM/MM approach. In: Lee JK (ed) *Orotidine monophosphate decarboxylase: a mechanistic dialogue. Topics in current chemistry*, vol 238. Springer, Berlin Heidelberg New York, pp 79–112

638. Loferer MJ, Tautermann CS, Loeffler HH, Liedl KR (2003) Influence of backbone conformations of human carbonic anhydrase II on carbon dioxide hydration: hydration pathways and binding of bicarbonate. *J Am Chem Soc* 125(29):8921–8927
639. Garcia-Viloca M, Nam K, Alhambra C, Gao J (2004) Solvent and Protein effects on the vibrational frequency shift and energy relaxation of the azide ligand in carbonic anhydrase. *J Phys Chem B* 108(35):13501–13512
640. Tautermann CS, Loferer MJ, Voegelé AF, Liedl KR (2003) About the kinetic feasibility of the Lipscomb mechanism in human carbonic anhydrase II. *J Phys Chem B* 107(43):12013–12020
641. Kamachi T, Toraya T, Yoshizawa K (2004) Catalytic roles of active-site amino acid residues of coenzyme B₁₂-dependent diol dehydratase: protonation state of histidine and pull effect of glutamate. *J Am Chem Soc* 126(49):16207–16216
642. Cisneros GA, Liu H, Zhang Y, Yang W (2003) Ab initio QM/MM study shows there is no general acid in the reaction catalyzed by 4-oxalocrotonate tautomerase. *J Am Chem Soc* 125(34):10384–10393
643. Cisneros GA, Wang M, Silinski P, Fitzgerald MC, Yang W (2004) The protein backbone makes important contributions to 4-oxalocrotonate tautomerase enzyme catalysis: understanding from theory and experiment. *Biochemistry* 43(22):6885–6892
644. Cisneros GA, Wang M, Silinski P, Fitzgerald MC, Yang W (2006) Theoretical and experimental determination on two substrates turned over by 4-oxalocrotonate tautomerase. *J Phys Chem A* 110(2):700–708
645. Garcia-Viloca M, González-Lafont À, Lluch JM (2001) A QM/MM study of the racemization of vinylglycolate catalyzed by mandelate racemase enzyme. *J Am Chem Soc* 123(4):709–721
646. Prat-Resina X, Garcia-Viloca M, González-Lafont A, Lluch JM (2002) On the modulation of the substrate activity for the racemization catalyzed by mandelate racemase enzyme: a QM/MM study. *Phys Chem Chem Phys* 4(21):5365–5371
647. Prat-Resina X, González-Lafont À, Lluch JM (2005) Reaction mechanism of the mandelate anion racemization catalyzed by mandelate racemase enzyme: a QM/MM molecular dynamics free energy study. *J Phys Chem B* 109:21089–21101
648. Puig E, Garcia-Viloca M, González-Lafont À, Lluch JM (2006) On the ionization state of the substrate in the active site of glutamate racemase. A QM/MM study about the importance of being zwitterionic. *J Phys Chem A* 110:717–725
649. Guallar V, Jacobson M, McDermott A, Friesner RA (2004) Computational modeling of the catalytic reaction in triosephosphate isomerase. *J Mol Biol* 337(1):227–239
650. Cui Q, Karplus M (2002) Promoting modes and demoting modes in enzyme-catalyzed proton transfer reactions: a study of models and realistic systems. *J Phys Chem B* 106(32):7927–7947
651. Cui Q, Karplus M (2002) Quantum mechanics/molecular mechanics studies of triosephosphate isomerase-catalyzed reactions: effect of geometry and tunneling on proton-transfer rate constants. *J Am Chem Soc* 124:3093–3124
652. Wang M, Lu Z, Yang W (2002) Transmission coefficient calculation for proton transfer in triosephosphate isomerase based on the reaction path potential method. *J Chem Phys* 121(1):101–107
653. Konuklar FAS, Aviyente V, Monard G, Lopez RMF (2004) Theoretical approach to the wear and tear mechanism in triosephosphate isomerase: a QM/MM study. *J Phys Chem B* 108(12):3925–3934
654. Garcia-Viloca M, Alhambra C, Truhlar DG, Gao J (2003) Hydride transfer catalyzed by xylose isomerase: mechanism and quantum effects. *J Comput Chem* 24(2):177–190

655. Li G, Cui Q (2003) What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. *J Am Chem Soc* 125(49):15028–15038
656. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2001) Transition structure selectivity in enzyme catalysis: a QM/MM study of chorismate mutase. *Theor Chem Acc* 105(3):207–212
657. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J, Field MJ (2001) A hybrid potential reaction path and free energy study of the chorismate mutase reaction. *J Am Chem Soc* 123(8):1709–1712
658. Lee YS, Worthington SE, Krauss M, Brooks BR (2002) Reaction mechanism of chorismate mutase studied by the combined potentials of quantum mechanics and molecular mechanics. *J Phys Chem B* 106(46):12059–12065
659. Crespo A, Scherlis DA, Martí MA, Ordejón P, Roitberg AE, Estrin DA (2003) A DFT-based QM–MM approach designed for the treatment of large molecular systems: application to chorismate mutase. *J Phys Chem B* 107(49):13728–13736
660. Guo H, Cui Q, Lipscomb WN, Karplus M (2003) Understanding the role of active-site residues in chorismate mutase catalysis from molecular-dynamics simulations. *Angew Chem Int Ed* 42(13):1508–1511
661. Martí S, Moliner V, Tuñón I, Williams IH (2003) QM/MM calculations of kinetic isotope effects in the chorismate mutase active site. *Org Biomol Chem* 1(3):483–487
662. Ranaghan KE, Ridder L, Szeftczyk B, Sokalski WA, Hermann JC, Mulholland AJ (2003) Insights into enzyme catalysis from QM/MM modelling: transition state stabilization in chorismate mutase. *Mol Phys* 101(17):2695–2714
663. Ranaghan KE, Ridder L, Szeftczyk B, Sokalski WA, Hermann JC, Mulholland AJ (2004) Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction. *Org Biomol Chem* 2(7):968–80
664. Ruggiero GD, Guy SJ, Martí S, Moliner V, Williams IH (2004) Vibrational analysis of the chorismate rearrangement: relaxed force constants, isotope effects and activation entropies calculated for reaction in vacuum, water and the active site of chorismate mutase. *J Phys Org Chem* 17(6–7):592–601
665. Claeysens F, Ranaghan KE, Manby FR, Harvey JN, Mulholland AJ (2005) Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chem Commun* (40):5068–5070
666. Guimarães CRW, Udier-Blagović M, Tubert-Brohman I, Jorgensen WL (2005) Effects of Arg90 neutralization on the enzyme-catalyzed rearrangement of chorismate to prephenate. *J Chem Theory Comput* 1(4):617–625
667. Ishida T, Fedorov DG, Kitauro K (2006) All electron quantum chemical calculation of the entire enzyme system confirms a collective catalytic device in the chorismate mutase reaction. *J Phys Chem B* 110:1457–1463
668. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2004) A comparative study of Claisen and Cope rearrangements catalyzed by chorismate mutase. An insight into enzymatic efficiency: transition state stabilization or substrate preorganization? *J Am Chem Soc* 126(1):311–319
669. Martí S, Andrés J, Moliner V, Silla E, Tuñón I, Bertrán J (2000) A QM/MM study of the conformational equilibria in the chorismate mutase active site. The role of the enzymatic deformation energy contribution. *J Phys Chem B* 104(47):11308–11315
670. Guo H, Cui Q, Lipscomb WN, Karplus M (2001) Substrate conformational transitions in the active site of chorismate mutase: their role in the catalytic mechanism. *Proc Natl Acad Sci USA* 98(16):9032–9037

671. Guo H, Lipscomb WN, Cui Q, Karplus M (2005) Molecular dynamics simulations of yeast chorismate mutase: substrate conformational transitions and enzymatic catalysis. In: Puglisi JD (ed) Structure, dynamics and function of biological macromolecules and assemblies. NATO Science Series I: Life and behavioural sciences, vol 364. IOS Press, Amsterdam, pp 63–74
672. Loferer MJ, Webb BM, Grant GH, Liedl KR (2003) Energetic and stereochemical effects of the protein environment on substrate: a theoretical study of methylmalonyl-CoA mutase. *J Am Chem Soc* 125(4):1072–1078
673. Zurek J, Bowman AL, Sokalski WA, Mulholland AJ (2004) MM and QM/MM modeling of threonyl-tRNA synthetase: model testing and simulations. *Struct Chem* 15(5):405–414
674. Nonella M, Mathias G, Eichinger M, Tavan P (2003) Structures and vibrational frequencies of the quinones in *Rb. sphaeroides* derived by a combined density functional/molecular mechanics approach. *J Phys Chem B* 107(1):316–322
675. Hughes JM, Hutter MC, Reimers JR, Hush NS (2001) Modeling the bacterial photosynthetic reaction center. 4. The structural, electrochemical, and hydrogen-bonding properties of 22 mutants of *Rhodobacter sphaeroides*. *J Am Chem Soc* 123(35):8550–8563
676. Reimers JR, Hutter MC, Hughes JM, Hush NS (2000) Nature of the special-pair radical cation in bacterial photosynthesis. *Int J Quant Chem* 80(6):1224–1243
677. Hasegawa J, Ishida M, Nakatsuji H, Lu Z, Liu H, Yang W (2003) Energetics of the electron transfer from bacteriopheophytin to ubiquinone in the photosynthetic reaction center of *Rhodopseudomonas viridis*: theoretical study. *J Phys Chem B* 107(3):838–847
678. Rajamani R, Gao J (2002) Combined QM/MM study of the opsin shift in bacteriorhodopsin. *J Comput Chem* 23(1):96–105
679. Bondar N, Elstner M, Fischer S, Smith JC, Suhai S (2004) Can coordinate driving describe proton transfer coupled to complex protein motions? *Phase Transitions* 77:47–52
680. Bondar AN, Elstner M, Suhai S, Fischer S, Smith JC (2005) Direct proton transfer in a putative L-state intermediate of the bacteriorhodopsin photocycle. *Phase Transitions* 78:5–9
681. Lee Y-S, Krauss M (2004) Dynamics of proton transfer in bacteriorhodopsin. *J Am Chem Soc* 126(7):2225–2230
682. Warshel A, Chu ZT (2001) Nature of the surface crossing process in bacteriorhodopsin: computer simulations of the quantum dynamics of the primary photochemical event. *J Phys Chem B* 105(40):9857–9871
683. Bondar AN, Smith JC, Fischer SK (2005) QM/MM investigation of the hydrogen-bonding interactions in putative K and early-M intermediates of the bacteriorhodopsin photocycle. *Phase Transitions* 78:671–675
684. Hayashi S, Tajkhorshid E, Kandori H, Schulten K (2004) Role of hydrogen-bond network in energy storage of bacteriorhodopsin's light-driven proton pump revealed by ab initio normal-mode analysis. *J Am Chem Soc* 126(34):10516–10517
685. Rousseau R, Kleinschmidt V, Schmitt UW, Marx D (2004) Modeling protonated water networks in bacteriorhodopsin. *Phys Chem Chem Phys* 6(8):1848–1859
686. Sugihara M, Buss V, Entel P, Hafner J, Bondar AN, Elstner M, Frauenheim T (2004) Ab initio, tight-binding and QM/MM calculations of the rhodopsin chromophore in its binding pocket. *Phase Transitions* 77(1–2):31–45
687. Sugihara M, Buss V (2005) Retinal versus 13-demethyl-retinal inside the rhodopsin binding pocket—a QM/MM study. *Phase Transitions* 78:11–15

688. Gascón JA, Sproviero EM, Batista VS (2006) Computational studies of the primary phototransduction event in visual rhodopsin. *Acc Chem Res* 39(3):184–193
689. Gascon JA, Batista VS (2004) QM/MM study of energy storage and molecular rearrangements due to the primary event in vision. *Biophys J* 87(5):2931–2941
690. Sugihara M, Hufen J, Buss V (2006) Origin and consequences of steric strain in the rhodopsin binding pocket. *Biochemistry* 45(3):801–810
691. Gascón JA, Sproviero EM, Batista VS (2005) QM/MM study of the NMR spectroscopy of the retinyl chromophore in visual rhodopsin. *J Chem Theory Comput* 1(4):674–685
692. Röhrig UF, Guidoni L, Rothlisberger U (2005) Solvent and protein effects on the structure and dynamics of the rhodopsin chromophore. *ChemPhysChem* 6(9):1836–1847
693. Hayashi S, Tajkhorshid E, Pebay-Peyroula E, Royant A, Landau EM, Navarro J, Schulten K (2001) Structural determinants of spectral tuning in retinal proteins: bacteriorhodopsin vs. sensory rhodopsin II. *J Phys Chem B* 105(41):10124–10131
694. Fujimoto K, Hasegawa J, Hayashi S, Kato S, Nakatsuji H (2005) Mechanism of color tuning in retinal protein: SAC-CI and QM/MM study. *Chem Phys Lett* 414(1–3):239–242
695. Ferré N, Olivucci M (2003) Probing the rhodopsin cavity with reduced retinal models at the CASPT2//CASSCF/AMBER level of theory. *J Am Chem Soc* 125(23):6868–6869
696. Wanko M, Hoffmann M, Strodel P, Koslowski A, Thiel W, Neese F, Frauenheim T, Elstner M (2005) Calculating absorption shifts for retinal proteins: computational challenges. *J Phys Chem B* 109(8):3606–3615
697. Patnaik SS, Trohalaki S, Pachter R (2004) Molecular modeling of green fluorescent protein: structural effects of chromophore deprotonation. *Biopolymers* 75(6):441–452
698. Sinicropi A, Andruniow T, Ferré N, Basosi R, Olivucci M (2005) Properties of the emitting state of the green fluorescent protein resolved at the CASPT2//CASSCF/CHARMM level. *J Am Chem Soc* 127(33):11534–11535
699. Marques MAL, López X, Varsano D, Castro A, Rubio A (2003) Time-dependent density-functional approach for biological chromophores: the case of the green fluorescent protein. *Phys Rev Lett* 90(25):258101/1–4
700. Dittrich M, Freddolino PL, Schulten K (2005) When light falls in LOV: a quantum mechanical/molecular mechanical study of photoexcitation in Phot-LOV1 of *Chlamydomonas reinhardtii*. *J Phys Chem B* 109(26):13006–13013
701. Groenhof G, Bouxin-Cademartory M, Hess B, de Visser SP, Berendsen HJC, Olivucci M, Mark AE, Robb MA (2004) Photoactivation of the photoactive yellow protein: why photon absorption triggers a trans-to-cis isomerization of the chromophore in the protein. *J Am Chem Soc* 126(13):4228–4233
702. <http://amber.scripps.edu/>
703. <http://www.cemcomco.com/>
704. <http://www.charmm.org/>
705. <http://www.scm.com/>
706. <http://www.cse.clrc.ac.uk/qcg/gamess-uk/>
707. <http://www.gaussian.com/>
708. <http://www.emsl.pnl.gov/docs/nwchem/>
709. <http://www.schroedinger.com/>
710. <http://www.pt.tu-clausthal.de/atp/>
711. <http://www.chemshell.org/>

712. <http://comp.chem.umn.edu/qmmm/>
713. <http://www.molpro.com/>
714. <http://www.thch.uni-bonn.de/tc/orca/>
715. <http://www.turbomole.com/>
716. <http://www.igc.ethz.ch/gromos/>
717. <http://www.ivec.org/GULP/>
718. http://www.cse.clrc.ac.uk/msi/software/DL_POLY/
719. <http://swift.cmbi.ru.nl/gv/whatcheck/>
720. <http://biotech.ebi.ac.uk:8400/>
721. <http://swift.cmbi.ru.nl/whatif/>
722. <http://propka.chem.uiowa.edu/>
723. <http://hekto.med.unc.edu:8080/HERMANS/software/DOWSER/Dowser.htm>
724. <http://www.chem.qmul.ac.uk/iubmb/enzyme/>
725. Yang Z, Zhou Y, Liu K, Cheng Y, Liu R, Chen G, Jia Z (2003) Computational study on the function of water within a β -helix antifreeze protein dimer and in the process of ice-protein binding. *Biophys J* 85(4):2599–2605
726. Liu H, Elstner M, Kaxiras E, Frauenheim T, Hermans J, Yang W (2001) Quantum mechanics simulation of protein dynamics on long timescale. *Proteins Struct Funct Genet* 44(4):484–489
727. Li H, Hains AW, Everts JE, Robertson AD, Jensen JH (2002) The prediction of protein pK_a's using QM/MM: the pK_a of lysine 55 in turkey ovomucoid third domain. *J Phys Chem B* 106(13):3486–3494
728. Magalhães AL, Maigret B, Gomes JANE, Ramos MJ (2000) Hybrid QM/MM study of a fundamental biological interaction surrounded by its protein environment. *Internet J Chem* 3:11
729. Cho AE, Guallar V, Berne BJ, Friesner R (2005) Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J Comput Chem* 26(9):915–931
730. Mlinsek G, Novic M, Hodoscek M, Solmajer T (2001) Prediction of enzyme binding: human thrombin inhibition study by quantum chemical and artificial intelligence methods based on X-ray structures. *J Chem Inf Comput Sci* 41(5):1286–1294
731. Spiegel K, Rothlisberger U, Carloni P (2004) Cisplatin binding to DNA oligomers from hybrid Car-Parrinello/molecular dynamics simulations. *J Phys Chem B* 108(8):2699–2707
732. Jia M, Qu W, Yang Z, Chen G (2005) Theoretical study of the factors that affect the structure and stability of the adduct of a new platinum anticancer drug with a duplex DNA. *Int J Mod Phys B* 19(15–17/2):2939–2949
733. Dolenc J, Borštnik U, Hodošček M, Koller J, Janežič D (2005) An ab initio QM/MM study of the conformational stability of complexes formed by netropsin and DNA. The importance of van der Waals interactions and hydrogen bonding. *THEOCHEM* 718(1–3):77–85
734. Magistrato A, Ruggerone P, Spiegel K, Carloni P, Reedijk J (2006) Binding of novel azole-bridged dinuclear platinum(II) anticancer drugs to DNA: insights from hybrid QM/MM molecular dynamics simulations. *J Phys Chem B* 110(8):3604–3613
735. Gossens C, Tavernelli I, Rothlisberger U (2005) Rational design of organo-ruthenium anticancer compounds. *Chimia* 59(3):81–84
736. Spiegel K, Rothlisberger U, Carloni P (2006) Duocarmycins binding to DNA investigated by molecular simulation. *J Phys Chem B* 110(8):3647–3660
737. Guallar V, Borrelli KW (2005) A binding mechanism in protein–nucleotide interactions: implication for U1A RNA binding. *Proc Natl Acad Sci USA* 102(11):3954–3959

738. Zhang Y (2006) Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor Chem Acc* 116:43–50
739. Elstner M (2006) The SCC-DFTB method and its application to biological systems. *Theor Chem Acc* 116(1):316–325
740. Claeysens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schütz M, Thiel S, Thiel W, Werner H-J (2006) High-accuracy computation of reaction barriers in enzymes. *Angew Chem Int Ed* 45(41):6856–6859
741. Laino T, Mohamed F, Laio A, Parrinello M (2006) An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J Chem Theor Comput* 2:1370–1378
742. Yu N, Hayik SA, Wang B, Liao N, Reynolds CH, Merz KM Jr (2006) Assigning the protonation states of the key aspartates in β -secretase using QM/MM X-ray structure refinement. *J Chem Theor Comput* 2:1057–1069
743. Schöll-Paschinger E, Dellago C (2006) A proof of Jarzynski's nonequilibrium work theorem for dynamical systems that conserve the canonical distribution. *J Chem Phys* 125(5):054105/1–5
744. Xiong H, Crespo A, Marti M, Estrin D, Roitberg AE (2006) Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. *Theor Chem Acc* 116(1):338–346
745. Olsson MHM, Mavri J, Warshel A (2006) Transition state theory can be used in studies of enzyme catalysis: lessons from simulations of tunneling and dynamical effects in lipoyxygenase and other systems. *Phil Trans R Soc B* 361:1417–1432
746. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM (2006) Electrostatic basis for enzyme catalysis. *Chem Rev* 106(8):3210–3235
747. Bruice TC (2006) Computational approaches: reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency. *Chem Rev* 106(8):3170–3187
748. Antoniou D, Basner J, Núñez S, Schwartz SD (2006) Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis. *Chem Rev* 106(8):3170–3187
749. Gao J, Ma S, Major DT, Nam K, Pu J, Truhlar DG (2006) Mechanisms and free energies of enzymatic reactions. *Chem Rev* 106(8):3188–3209
750. Pu J, Gao J, Truhlar DG (2006) Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem Rev* 106(8):3140–3169
751. Schramm VL (ed) (2006) *Chem Rev* 106(8)
752. Zheng J, Wang D, Thiel W, Shaik S (2006) QM/MM study of mechanisms for compound I formation in the catalytic cycle of cytochrome P450cam. *J Am Chem Soc* 128(40):13204–13215
753. Altun A, Shaik S, Thiel W (2006) Systematic QM/MM investigation of factors that affect the cytochrome P450-catalyzed hydrogen abstraction of camphor. *J Comput Chem* 27:1324–1337
754. Tuttle T, Keinan E, Thiel W (2006) Understanding the enzymatic activity of 4-oxalocrotonate tautomerase and its mutant analogues: a computational study. *J Phys Chem B* 110:19685–19695
755. Crespo A, Martí MA, Estrin DA, Roitberg AE (2005) Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J Am Chem Soc* 127(19):6940–6941

Transition Path Sampling Simulations of Biological Systems

Christoph Dellago¹ (✉) · Peter G. Bolhuis²

¹Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria
Christoph.Dellago@univie.ac.at

²van't Hoff Institute for Molecular Sciences, University of Amsterdam,
 Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

1	Introduction	292
2	Transition Path Sampling Techniques	296
2.1	The Path Ensemble	296
2.2	Defining the Stable States	298
2.3	Sampling the Path Ensemble with Monte Carlo	299
2.4	The Initial Path	300
2.5	Computational Issues	301
2.6	Analysis of the Reaction Mechanism	302
2.7	Calculation of Rate Constants	304
3	Applications	306
3.1	Biomolecular Isomerization	306
3.2	Binding and Unbinding of DNA Base Pairs	307
3.3	Protein Folding	308
3.4	Enzyme Catalysis	311
3.5	Lipid Bilayers	312
3.6	Switches in Biochemical Networks	312
4	Outlook	313
	References	314

Abstract Transition path sampling is a computational method for the simulation of rare but important events occurring in complex systems. Based on a statistical mechanics in trajectory space, transition path sampling can be applied to identify mechanisms and determine rate constants. Here we review the basic ideas and algorithms of transition path sampling and discuss some recent applications of this methodology to problems in molecular biology including protein folding, enzyme catalysis, and processes occurring in lipid bilayers.

Keywords Rare events · Computer simulation · Transition path sampling

Abbreviations

DPS Discrete path sampling
 FFS Forward flux sampling
 MC Monte Carlo

MD Molecular dynamics
TIS Transition interface sampling
TPS Transition path sampling
TS Transition state
TSE Transition state ensemble
TST Transition state theory

1

Introduction

Molecular biology strives to describe the structures and processes occurring in the living cell on the molecular level [1]. Molecular processes can involve proteins, nucleic acids, lipid molecules, or other molecules. Protein-specific processes include enzyme catalyzed reactions, protein folding, binding and regulation, signaling pathways, and transport by molecular motors (trafficking, muscles). Nucleic acids are involved in RNA folding, ribozymatic reactions, DNA replication, transcription, chromatin formation, nucleosomes transitions, and translocation. Lipid molecules are of course crucial for understanding membranes, and are required in the description of protein/nucleic acid translocation, the functioning of ion and water channels, membrane ion permeation, and membrane protein association. To understand such processes it is of crucial importance to obtain detailed knowledge of the involved mechanisms on a molecular level.

Computer simulations have been very successful in providing insight into the structure and dynamics of biomolecules. In particular, the advent of molecular dynamics simulation and the development of appropriate interaction potentials, a.k.a. force fields, for biomolecules has had an enormous impact (see, for instance [2]; see also the paper by Schulten et al. in this volume). However, one of the crucial problems one encounters in the simulation of many of the above-mentioned processes is that they are rare and occur on timescales not accessible to molecular dynamics simulations even on today's fastest computers. For instance, times required by proteins to assemble into their native structure typically range from milliseconds to even hours. On the other hand, the timescale of molecular vibrations sets the basic molecular dynamics time step to about one femtosecond. This large separation of timescales results in a huge number of molecular dynamics steps required for the observation of just one folding event. Investigating such rare processes by means of straightforward molecular dynamics simulation is therefore impractical on current computers.

Rare events are related to the existence of high energy barriers or entropic bottlenecks partitioning the configuration space of the system in different metastable basins (for instance the folded and unfolded state of a protein). In equilibrium, the system fluctuates in these long-lived states most of the time

and the barriers separating them are crossed only rarely. While the time intervals between such transitions may be very long, often the barrier-crossing event is swift when it occurs. An early approach to solve this timescale problem dates back to the 1930s when Eyring [3] and Wigner [4] introduced transition state theory (TST). TST is based on the concept of a transition state, a saddle point in the potential energy surface that the system must cross on its way from reactants to products. Under the assumption that trajectories that have crossed the transition state coming from the product region never recross the transition state, TST yields the reaction rate constant.

Algorithms for the calculations of dynamical corrections to TST were later introduced by Keck, Bennett, and Chandler [5–7] providing an, in principle exact, route to the calculation of reaction rate constants. In the Bennett–Chandler approach, rate constants are computed in a two-step procedure. First, the free energy profile along a postulated reaction coordinate is calculated and used to determine the transition state theory approximation of the reaction rate constant. As the transition state region corresponding to the free energy maximum involves unlikely configurations that are rarely sampled in equilibrium, advanced methods such as umbrella sampling [8] or thermodynamic integration [9, 10] are required in such a free energy calculation. The TST rate constants corresponds to the instantaneous flux through a dividing surface defined by the value of the reaction coordinate at the top of the free energy barrier. In a second step, a dynamical correction factor taking into account correlated recrossings of the dividing surface, the so called transmission coefficient, is determined from molecular dynamics trajectories initiated at the dividing surface. These trajectories can also be analyzed to study the details of the transition mechanism. In the context of biomolecular simulations, transition state theory, with and without dynamical corrections, was mainly used to calculate reaction rate constants of enzymatic reactions [11] but also of conformational changes [12] and trans-membrane transport in pores [13], to name just a few examples.

The applicability of TST and its more sophisticated variants such as variational TST [14] or the above-mentioned Bennett–Chandler approach, relies on the knowledge of an appropriate reaction coordinate. A reaction coordinate is a function of the configurational degrees of freedom of the system that should be capable of characterizing the progress of a transition through the dynamical bottleneck region. If the reaction coordinates are chosen in a way that does not capture the essential mechanisms involved in the transition, the methods discussed in the previous section fail. However, identifying the right coordinates can be exceedingly difficult in a complex phase space. The generic reaction coordinates for protein folding, for instance, are still unknown.

In some cases, relevant transition states may be located by searching for saddle points in the potential energy surface [15–17]. The direction of the unstable mode then serves as a reaction coordinate permitting a transition state

theory calculation of the reaction rate constant. From databases of stationary points in the potential energy surface it is then possible to obtain the kinetics of processes involving transitions between many stable states [18, 19]. While this approach can yield valuable insights into the properties of biomolecular systems [20, 21], it is limited in its applicability due to the exponential growth of the number of stationary points with system size. Stochastic transitions between metastable states consisting of almost invariant subsets of conformations, rather than single potential energy minima, are considered in Markovian state models [22, 23].

A similar viewpoint, in which the dynamics is regarded as consisting of jumps between long-lived states, is adopted in the parallel replica method [24, 25], which exploits the availability of highly parallel computer systems. In this method, independent simulations initiated with different initial conditions are performed in parallel on many processors. Each replica is periodically tested for transitions from a stable state to a new one. As soon as a transition is detected on one processor, all simulations are stopped and then restarted in the new state. Provided that transitions are uncorrelated with each other, the correct distribution of escape times can be reconstructed and in the optimum case a speed-up proportional to the number of processors can be achieved. However, the number of needed replicas required to observe a transition event in a given time increases exponentially with the barrier height. In the realm of biomolecular simulation, this approach has been mainly used to determine the folding pathways of small proteins on thousands of computers running the simulation in parallel [26].

Other rare event approaches that have been applied in biomolecular systems, such as the hyperdynamics method [27, 28] or the flooding method [29, 30] (see also the local elevation method [31]), increase the likelihood of visiting rare transition state conformations by modification of the potential energy surfaces. The modified energy landscape destabilizes the long-lived (meta)stable states and thus accelerates transitions between them. Although the artificial potential energy surface alters the dynamics of the system, certain aspects of the time evolution are conserved on a coarse grained level.

Modifications of the potential energy surface favoring the sampling of rare transition state configurations can also be implemented by introducing time-dependent perturbations that drive the system away from those regions of configuration space that have been already visited. This idea is the basis of the metadynamics method that can be used to map out the free energy of the system as a function of a small set of preselected collective variables [32]. Although, due to the time-dependent bias, the dynamics differs from the true dynamics of the system, metadynamics is a powerful tool for the identification of possible transition mechanisms. The method, hinging on an appropriate definition of the collective variables, has been applied to study various processes ranging from chemical reactions in solution [33] to trans-

membrane transport [34]. Such a reduced description is also the basis of the coarse grained molecular dynamics method, in which short trajectories in the full phase space are used to estimate effective forces acting in a low-dimensional space spanned by some preselected coarse variables [35]. As illustrated by simulating conformational changes of alanine dipeptide in water, this method can considerably accelerate the simulation, albeit at the cost of a detailed description of the dynamics.

In many processes involving rare transitions between long-lived stable states, not only the initial but also the final state is known. For instance, in the case of membrane translocation, enzymatic reactions, or isomerizations detailed information is available both for reactants and products. This additional knowledge is exploited in several approaches based on pathways that start and end in particular regions of configuration space. These methods differ in the particular type of dynamics considered. In early work, Pratt [36] suggested using a Monte Carlo approach to sample the probability distribution of stochastic pathways connecting reactants with products. Other methods are based on a minimization of the classical mechanical action [37, 38]. The advantage of these latter methods is that they can be used with very large time-steps, thus speeding up the simulation considerably. However, the dynamical meaning of such long-timestep trajectories remains unclear and results obtained with these methods should be verified with other approaches. Paths connecting reactant with products regions were also considered by Elber and Karplus, who devised a method to determine transition pathways by minimizing an artificial path “energy”, or action [39]. Other methods follow a similar approach but differ in the particular form of action that is minimized [40–42]. Within this class of algorithms the nudged elastic band method [43] permits a particularly efficient localization of the transition state, which can subsequently be used to estimate reaction rate constants from transition state theory. In this method, forces perpendicular and tangential to the pathway “nudge” the trajectory towards a minimum energy pathway passing through the transition state. In a biological context the nudged elastic band method has, for instance, been used to determine pathways for conformational changes in RNA [44]. The same forces used in the nudged elastic band method are also applied in the so-called string method and its extension, the finite temperature string method [45, 46]. Applicable to stochastic dynamics, the latter algorithm is based on a constrained sampling of the equilibrium distribution in a sequence of hyperplanes along a string that connects reactants with products. These hyperplanes are adapted iteratively such that they are tangential to the isoproability surfaces consisting of all points that have the same probability to relax into one stable state rather than the other. As a result of a string method calculation one obtains a transition tube, in which the transition occurs with high probability, and also the corresponding free energy profile and transition rate constant. The applicability of the

finite temperature string method to biomolecular systems has been demonstrated in an analysis of the isomerization of alanine dipeptide in vacuum and explicit solvent [47].

A more widely applicable approach, the transition path sampling method (TPS) [48, 49], consists in first defining the ensemble of all pathways connecting reactants with products and then sampling this ensemble of dynamical pathways with Monte Carlo techniques. The TPS method is valid for a wide range of dynamics ranging from Monte Carlo dynamics to Langevin dynamics and Newtonian dynamics. The method permits the identification of reaction mechanisms and the calculation of reaction rate constants in complex systems with rough free energy barriers. We will focus on TPS and its application to biomolecular problems in the remaining part of this article, which is structured as follows. In Sect. 2 we give a brief overview the basic techniques involved in path sampling, followed by a description of how to calculate the rate constants. In Sect. 3 we review several papers applying TPS to biologically relevant systems. We end with an outlook on future possibilities.

2

Transition Path Sampling Techniques

2.1

The Path Ensemble

We are interested in rare event processes that occur in complex systems between two (meta)stable long-lived states, denoted A and B , separated by an unknown and, in general, rough and high free energy barrier. The system is (meta)stable in states A and B in the sense that if the system is put there it will remain there for a long time, and only rarely undergoes transitions between the stable states A and B . We assume that no long-lived metastable states exist between states A and B . Further, we assume that states A and B can be characterized using a small set of order parameters.

The goal of a TPS simulation is to collect all likely transition pathways between these states A and B . The basic idea behind TPS is a generalization of the standard Monte Carlo importance sampling procedures [50–52] to trajectory space. When we consider, for instance, the ensemble of all trajectories with a given time length that is small compared to the reaction time, most of these will be localized in or near either stable state A or B . Rare transitions between A and B will comprise only a small subset of these short trajectories. For example, if the process of interest occurs roughly once every millisecond, then only one out of a million 1 ns trajectories will exemplify that process. TPS provides an efficient means to sample such sub-ensembles of rare trajectories. Subsequent analysis of the path ensemble enables the elucidation of the

transition mechanism, i.e., the degrees of freedom that capture the physics of the transition and how these change during the transition. Since the pathways collected with TPS are fully dynamical trajectories rather than artificial pathways, such as minimum energy pathways or elastic bands, it is also possible to extract kinetic information from TPS simulations.

In this section we describe very briefly the basic concepts of TPS in the context of biomolecular simulations. For a more detailed description of this methodology we refer the reader to recent review articles [49, 53–56]. To define the path ensemble we discretize the continuous time evolution of the system as an ordered sequence of L “time slices”:

$$\mathcal{X} = \{x_0, x_1, x_2, \dots, x_i, \dots, x_L\} . \quad (1)$$

Consecutive time slices are separated by a small time increment Δt and so x_i denotes the complete system at time $i\Delta t$. For a Hamiltonian system, for instance, $x = \{r, p\}$ includes the positions r and momenta p of all particles. The probability (density) of observing a given path $\mathcal{X}$ depends on the probability of its initial condition x_0 and the specific dynamics of the system. For Markovian processes, the total path probability is:

$$\mathcal{P} [\mathcal{X}] = \rho(x_0) \prod_{i=0}^{L-1} p(x_i \rightarrow x_{i+1}) , \quad (2)$$

where $\rho(x_0)$ is the probability distribution of the initial conditions. For instance, if the system obeys a canonical Boltzmann distribution, $\rho(x_0) = \exp(-\beta H(x_0))/Z$, where H is the Hamiltonian, $\beta = 1/k_B T$ the inverse temperature, and $Z = \int dx \exp[-\beta H(x)]$ the canonical partition function.

The particular form of the transition probability $p(x_i \rightarrow x_{i+1})$ depends on the exact type of deterministic and stochastic dynamics. For deterministic dynamics (for stochastic transition probabilities see [49, 57]), the state of the system x at time $t = i\Delta t$ is completely determined by the state x_0 at time $t = 0$: $x_i = \phi_t(x_0)$. The time-dependent function $\phi_t(x_0)$ is the propagator of the system. For such deterministic dynamics, the short time transition probability can be written in terms of a Dirac delta function: $p(x_i \rightarrow x_{i+1}) = \delta[x_{i+1} - \phi_{\Delta t}(x_i)]$. Accordingly, the path probability for a Newtonian trajectory is:

$$\mathcal{P} [\mathcal{X}] = \rho(x_0) \prod_{i=0}^{L-1} \delta[x_{i+1} - \phi_{\Delta t}(x_i)] . \quad (3)$$

Other examples of deterministic dynamics include the extended Lagrangian dynamics of Car and Parrinello [59] and various dynamics with thermostats such as the Nose–Hoover thermostat [60–62] or the Gaussian isokinetic thermostat [63, 64]. The above definition of path ensembles also applies for these types of dynamics.

To constrain the path ensemble to those rare paths that connect stable state A with state B , we multiply the path probability $P[\mathcal{X}]$ with characteristic functions:

$$\mathcal{P}_{AB}[\mathcal{X}] = h_A(x_0) \mathcal{P}[\mathcal{X}] h_B(x_L) / Z_{AB}, \quad (4)$$

where $h_A(x)$ and $h_B(x)$ are the characteristic functions of region A and B , respectively, $h_{A,B}(x) = 1$ if $x \in A, B$ and zero otherwise, and Z_{AB} is a normalization factor corresponding to a partition function in trajectory space. Thus, a path that does not start in A and end in B has a zero weight whereas a path connecting A and B , may have a non-zero weight depending on the unrestricted path probability $\mathcal{P}[\mathcal{X}]$. The transition path ensemble $\mathcal{P}_{AB}[\mathcal{X}]$ selects only the reactive trajectories from the ensemble of all possible pathways while leaving the relative probabilities of the reactive trajectories among each other unchanged.

2.2

Defining the Stable States

The main advantage of TPS is that it does not require precise knowledge of the reaction coordinate. Instead, a careful characterization of the initial region A and the final region B is important. Often, such characterization involves finding one or more convenient low dimensional order parameters q , which have to meet several criteria. First and most importantly, the order parameters should discriminate A from B , i.e., region A should not overlap with the basin of attraction of B and vice versa¹. In other words, region A spanned by $h_A(x)$ and region B spanned by $h_B(x)$ should be located entirely within the corresponding basins of attraction. If this criterion is not met, the TPS algorithm will eventually collect mostly non-reactive trajectories. Correct definitions of regions A and B can be based on straightforward MD trajectories in both states, but in our experience, identifying discriminating order parameters can involve considerable trial and error experimentation. Fortunately, it is easy to detect problems associated with an inappropriate choice of order parameter. When problems are caused by non-discriminating regions A and B , the order parameter has to be refined until correct sampling can be achieved.

A second criterion for the order parameters is that the regions should be *representative*: A and B have to be large enough to accommodate typical equilibrium fluctuations. This is sometimes difficult to achieve, and therefore this criterion can be relaxed and the stable states only have to be *visited* frequently enough within the path length. If this criterion is not met, important transition pathways might be missing from the transition path ensemble.

¹ The basin of attraction of a specific stable state consists of all configurations from which trajectories relax into that stable state.

2.3

Sampling the Path Ensemble with Monte Carlo

As mentioned above, the goal of a TPS simulation is to collect reactive trajectories according to their weight in the transition path ensemble. This can be achieved with a Monte Carlo importance sampling procedure similar to a molecular simulation [51]. A molecular MC algorithm samples configurational space by generating a new trial configuration from an old one by randomly selecting a particle and moving it by a random distance. This trial configuration is accepted or rejected according to an acceptance probability satisfying the detailed balance condition, usually, the Metropolis rule. This basic step is then iterated. Provided it is ergodic, this algorithm guarantees the sampling of the desired distribution of configurations. Since this procedure generates a random walk that visits configurations according to their statistical weight, it is also called importance sampling. Monte Carlo procedures can sample arbitrary probability distributions including the path ensemble (Eq. 4), in which case the random walk is in the space of trajectories. The basic step of this MC procedure consists of generating a new path, $\mathcal{X}^{(n)}$, from an old one, $\mathcal{X}^{(o)}$. The Metropolis rule for accepting such a new path is given by:

$$P_{\text{acc}}[\mathcal{X}^{(o)} \rightarrow \mathcal{X}^{(n)}] = \min \left[1, \frac{\mathcal{P}_{AB}[\mathcal{X}^{(n)}] P_{\text{gen}}[\mathcal{X}^{(n)} \rightarrow \mathcal{X}^{(o)}]}{\mathcal{P}_{AB}[\mathcal{X}^{(o)}] P_{\text{gen}}[\mathcal{X}^{(o)} \rightarrow \mathcal{X}^{(n)}]} \right], \quad (5)$$

where $P_{\text{gen}}[\mathcal{X}^{(o)} \rightarrow \mathcal{X}^{(n)}]$ is the probability of creating a new path $\mathcal{X}^{(n)}$ from an existing path $\mathcal{X}^{(o)}$. A variety of schemes to generate new trial paths exist [49], but so far the shooting algorithm has proved to be the most efficient one. (See [49] for other path-generating procedures.) The shooting algorithm generates new pathways by (a) randomly selecting a time slice $x_i^{(o)}$ along the old path, (b) perturbing it to yield $x_i^{(n)}$ (this step is necessary for deterministic dynamics), and (c) generating from this time slice two new path segments forward and backward in time according to the propagation rules of the underlying dynamics. Together, these two new segments form the new pathway. Using this path generation scheme and assuming microscopic reversibility holds for the underlying dynamics, the acceptance rule simplifies enormously [49]:

$$P_{\text{acc}}[\mathcal{X}^{(o)} \rightarrow \mathcal{X}^{(n)}] = h_A(x_0^{(n)}) h_B(x_L^{(n)}) \min \left[1, \frac{\rho[x_i^{(n)}]}{\rho[x_i^{(o)}]} \right]. \quad (6)$$

Hence, any new trial path connecting the stable states A and B is accepted with a probability depending only on the shooting points of the old and the new pathway. This simple acceptance rule also suggests the following algorithm. First, select a shooting point $x_i^{(o)}$ at random and modify it to $x_i^{(n)}$. Then, accept this shooting point when a random number selected from a uniform

distribution in the interval $[0, 1]$ is smaller than $\rho[x_i^{(n)}]/\rho[x_i^{(o)}]$. If the shooting point is accepted, grow either one of the path segments. If this segment reaches the appropriate stable region, grow the other segment. The whole path is finally accepted if the boundary condition for that segment is also satisfied. Rejection of any of these steps results in rejecting the entire trial path. This procedure is then repeated to harvest an ensemble of transition paths. The acceptance rule (Eq. 6) is very general and is valid for any procedure that is microscopically reversible. Cases in which it is applicable include Langevin dynamics, Monte Carlo dynamics, and Newtonian dynamics. We refer the reader to [56] for variants of the shooting algorithm specific to the various types of dynamics.

For biomolecular processes the pathways can become very long. From a computational viewpoint it might therefore be useful to minimize the path length to its absolute minimum: the transition time. This is the part of the trajectory from exiting A until entering B , thus leaving out the uninteresting parts of the trajectories that are inside the stable states. This is allowed, as long as the probability of leaving stable state B again, after coming directly from A , is as likely as an independent transition. In other words, once B (or A) has been reached, no recrossings are allowed. It can be shown that the shooting procedure allows for such a variable path length [65]. The shooting algorithm remains the same, but the integration of the equations of motion from the shooting point continues only until one of the stable states is reached. As the path length varies, the probability $1/L$ to pick a random shooting point is not symmetric anymore with respect to the reverse move. The acceptance rule thus now becomes,

$$P_{\text{acc}} \left[\mathcal{X}^{(o)} \rightarrow \mathcal{X}^{(n)} \right] = h_A \left(x_0^{(n)} \right) h_B \left(x_L^{(n)} \right) \min \left[1, \frac{L^{(o)} \rho[x_i^{(n)}]}{L^{(n)} \rho[x_i^{(o)}]} \right], \quad (7)$$

where $L^{(o)}$ and $L^{(n)}$ denote the length of the old and the new path, respectively. This procedure has the advantage that the path length is restricted to the minimum number of time steps needed to cross the barrier and is hence automatically always optimal.

2.4

The Initial Path

Initiating a TPS simulation with the algorithms described above requires an existing pathway connecting A with B . Hence, the generation of an initial transition pathway is an essential step in the application of TPS. The simplest way to obtain an initial trajectory connecting A and B is by running a long molecular dynamics (or stochastic dynamics) simulation. For most applications, however, the rarity of the process of interest renders this straightforward approach impractical. More efficient ways to create an ini-

tial trajectory exist, but, in general, produce an atypical trajectory with a very low weight in the transition path ensemble $P_{AB}[\mathcal{X}]$. Therefore, the first part of a TPS simulation starting from such artificial trajectory consists of equilibrating the pathway towards the more important regions of trajectory space. This situation is analogous to that encountered in a conventional MC simulation of a molecular fluid. The initial configuration is often a regular lattice, which subsequently relaxes to the more probable fluid regions of configuration space. Similarly, a TPS simulation can start from an artificial pathway which does not even need to be a true dynamical trajectory. Then, repeated application of the MC steps described above move the pathways towards more typical regions of pathways and the actual TPS simulation can begin.

Now, how does one create this artificial initial pathway? In some situations, high-temperature pathways can be used to initiate a TPS simulation. Consider, for example, the folding/unfolding of a protein. At physiological temperatures, a protein in its native states unfolds only very rarely on a molecular timescale. At higher temperatures, however, unfolding occurs sufficiently quickly so that it can be simulated with ordinary molecular dynamics simulations. Such a high temperature trajectory acts then as an initial pathway to start a TPS simulation at the temperature of interest. If high temperature transition pathways are qualitatively very different from those at lower temperatures it might be necessary to carry out an annealing procedure, in which the ensemble of pathways is brought to a lower temperature in small steps. In other cases, one may have some, possibly incomplete, notion of a reaction coordinate. Controlling this presumed reaction coordinate one might be able to drive (bias) the system from A to B obtaining a chain of states from which shooting and shifting moves can be initiated.

A more systematic way to create a new transition pathway is to change the ensemble $\mathcal{P}_{AB}[\mathcal{X}]$ gradually from one which includes all trajectories starting in A (without restrictions on the end point) to one which consists only of trajectories starting in A and ending in B . Conversion of one path ensemble into another is computationally demanding and in most cases more efficient ways to generate an initial trajectory exist.

In our experience, no general recipe is available for the generation of an initial trajectory. Rather, specific procedures have to be developed for this purpose for each application of the TPS method to a new problem.

2.5

Computational Issues

The computational effort in carrying out a TPS calculation scales linearly with the number of trajectories harvested. In particular, to obtain N statistically independent trajectories of length t requires the same order of effort as that required to perform N trajectories. In practice, shooting moves are accepted with probabilities between 0.1 and 0.7, with an optimal acceptance

ratio of about 40% [49]. Further, the correlations in that random walk persist typically for only two or three accepted moves. For the stochastic algorithms this number can be much larger. Thus, for instance, 1000 statistically independent 1 ps trajectories are obtained with roughly the same computational resources required for a single straightforward trajectory of length 10–100 ns. If the rare event occurs with a rate constant of $1 \mu\text{s}^{-1}$, a straightforward trajectory will probably not show a single transition, while the 1000 transition path trajectories each will exhibit an independent example of the event.

As explained in the previous section the efficiency can be improved by roughly a factor of two using a flexible path length, in order to reduce the computational effort to a minimum.

Rare events in biomolecular systems often require pathways considerably longer than a few picoseconds. For instance, a folding protein can take several nanoseconds to diffuse over the folding free energy barrier, even when there is no intermediate long-lived (meta)stable state. The deterministic shooting algorithm runs into problems for such “long” diffusive folding trajectories. While a randomly chosen shooting point might seem to lie in the barrier region (i.e., outside of the stable state definitions) it can in fact already be completely committed to one of the stable states. In that case, the acceptance ratio will be extremely low. Only when shooting from points around the true transition state region, can we expect a reasonable acceptance. Unfortunately there is no way to recognize these points a priori. To alleviate this problem one can employ the stochastic shooting algorithm [66], allowing shooting in one direction, either forward or backwards. To still be able to use deterministic MD in combination with the stochastic algorithm requires the introduction of a small amount of stochasticity in the trajectories, for instance by application of the Andersen thermostat. The Andersen coupling constant can be made sufficiently small so that there is no noticeable difference from completely deterministic dynamics [51]. In this case, the acceptance ratio will improve dramatically, in fact to about 50%. The downside is that the correlation between paths increases. Several tools have been developed to keep track of this correlation [73].

2.6

Analysis of the Reaction Mechanism

As mentioned before, the main advantage of TPS is that it does not require a priori knowledge of the reaction coordinate. Instead one can analyze the path ensemble to obtain transition state ensembles and test proposed reaction coordinates by computing committor distributions.

In a typical complex biomolecular case the order parameters used to define the stable states are bad reaction coordinates. Driving the system from *A* to *B* along such wrong coordinates neglects important barriers that exist in orthogonal directions and leads to large hysteresis effects. A good reaction

coordinate should capture the essence of the dynamics and should allow the prediction of where a trajectory initiated from a given configuration will most likely head. The commitment probability p_B or committor enables precisely that, as it is defined as the probability to relax into stable state B . The committor is a statistical measure for how committed a given configuration is to the product state. A value of $p_B = 0$ indicates no commitment at all while a value of $p_B = 1$ indicates that the configuration is fully committed to B . Fractional values like 0.5 denote that the configuration is equally committed to A and B . The committor (a.k.a p -fold, or splitting probability [67–70]) is a direct statistical indicator of the progress of the reaction. In this sense it is an ideal reaction coordinate. Thus, if a good reaction coordinate is available, the committor can be parametrized in terms of this coordinate [46]. Parameterizing the committor in terms of a small number of atomic variables to gain insight into the reaction mechanism, however, is a highly non-trivial task.

A practical way to compute the committor for an arbitrary configuration r is to start N independent trajectories and compute the fraction of paths p_B that reach stable state B after time t . For deterministic dynamics an average fraction is obtained by drawing the initial momenta for r from a Maxwell–Boltzmann distribution. For stochastic dynamics, the average extends also over noise histories. Assuming that the N trajectories are statistically independent, the standard deviation in p_B calculated from N trajectories is $\sigma = \sqrt{p_B(1 - p_B)/N}$. This expression can be used to terminate a committor calculation after a certain desired accuracy has been reached. As $p_B(1 - p_B)$ has its maximum at $p_B = 1/2$ the largest number of trajectories is needed for configurations with a committor $p_B = 0.5$. Note that the committor depends on the time-length of the trajectories used. This time should be larger than the molecular timescale needed to commit to one of the stable states. A time-independent form of the committor can be defined by counting all trajectories originating from a given configuration r that reach state B before they reach state A .

A configuration r with a committor value $p_B = p_A = 0.5$ marks a transition state. Both stable states A and B are equally accessible from that configuration. Transition states defined in this statistical way often do not coincide with particular features of the potential energy surface, because in general entropic contributions play an important role in determining the location of statistically defined transition states in configuration space. The high dimensional surface with committor value 0.5 divides or separates the basins of attraction of the stable states. Hence, this dividing surface is also called the separatrix [68].

In principle, it is possible to find this separatrix by calculating the committor values for all possible configurations, but this is rather impractical. Moreover, a $p_B = 0.5$ configuration does not hold information on what the probability of the TS itself is [70, 71]. The path ensemble, however, provides a weighted ensemble of transition paths, all by necessity crossing the sep-

aratrix. Locating configurations with $p_A = p_B$ on the ensemble of transition pathways results in the transition state ensemble (TSE). All configurations on transition pathways that have a committor of $p_B = 0.5$ are members of the TSE. Each transition path contributes at least one configuration, but possibly more, to the TSE. Subsequent analysis of the TSE can yield important information about the reaction mechanism.

One can also use the committors to test a proposed parameter q as a reaction coordinate. When q is a good reaction coordinate it should be able to capture the separatrix, and provide a description of the TSE. Calculating distributions of committors for constraint ensembles can be used to test the correctness of the proposed reaction coordinate. This diagnostic tool is not restricted to TPS but can be applied to any case requiring a test of whether or not the proposed order parameter describes the transition, or is only partially correlated (if at all) with the reaction coordinate.

The first step in this procedure is the calculation of the free energy profile as a function of a variable q . Most likely this profile will show a barrier with a maximum at $q = q^*$. Next, configurations with $q = q^*$ are generated, for instance by imposing constraints in a molecular dynamics simulation. Subsequently, commitment probabilities are computed for a representative number of configurations in this ensemble. If each of these configurations is truly in the transition state region, all committors will be around $p_B = 0.5$. In this case, q is a good reaction coordinate. If, in contrast, they are far from one half, say 0 or 1, the configurations are clearly not part of the separatrix, but already committed to A or B. In that case, q is not a good reaction coordinate. Analysis of the distributions of committors in this way allows location of the correct reaction coordinates, at least in the neighborhood of the separatrix. The committor distribution can thus be used to estimate how far a postulated reaction coordinate is removed from the correct reaction coordinate.

2.7

Calculation of Rate Constants

The pathways collected with TPS are true dynamical trajectories and hence can capture the kinetics of the reaction. If the barrier between A and B is high enough to cause a separation of timescales, the phenomenological rate constant k_{AB} for the unimolecular (first order) reaction $A \leftrightarrow B$ can be related to a microscopic time correlation function $C(t)$ of the characteristic functions h_A and h_B that is linear for times $\tau_{\text{mol}} < t \ll \tau_{rxn}$ (τ_{mol} is the molecular timescale and τ_{rxn} the reaction time):

$$C(t) = \frac{\langle h_A(x_0) h_B(x_{t/\Delta t}) \rangle}{\langle h_A \rangle} \approx k_{AB} t. \quad (8)$$

Here, the brackets denote an ensemble average over the initial condition x_0 . In words, $C(t)$ is the conditional probability that if the system is in A at time 0 it

will be in B at time t . To evaluate the time correlation function $C(t)$ in the TPS framework we can relate it to an order parameter $\lambda(x)$ that distinguishes between the reactant region A and product region B and for which the range of values λ takes when the system is in A ($\lambda_{\min}^A < \lambda < \lambda_{\max}^A$) does not overlap with the range of values λ takes in B ($\lambda_{\min}^B < \lambda < \lambda_{\max}^B$). The normalized probability density for finding a particular value of the parameter λ at the endpoints of pathways starting in A is given by:

$$P_A(\lambda', L) = \langle \delta(\lambda' - \lambda(x_L)) \rangle_A, \quad (9)$$

where the average is taken over all pathways starting from A . The function $C(t)$ and hence the rate constant follows from the integrated probability to reach B :

$$C(t) = \int_{\lambda_{\min}^B}^{\lambda_{\max}^B} d\lambda P_A(\lambda, L). \quad (10)$$

This probability is extremely low for high barriers, and the calculation has to be done using biasing techniques such as umbrella sampling (US), in which $C(t)$ is obtained by matching the results of TPS simulations carried out for different λ -windows (see [49] for a detailed description of this procedure).

Transition Interface Sampling. The efficiency of the rate constant procedure can be improved using a slightly different path sampling approach named transition interface sampling (TIS) [65]. The TIS method has in common with the umbrella sampling method that it uses an order parameter to bridge the barrier region between A and B . However, instead of windows, TIS maps the configuration space onto a foliation of *interfaces* characterized by $\lambda(x)$. The TIS method comprises the computation of the flux to leave a stable initial state and, subsequently, the conditional probability for trajectories leaving the initial state to reach the final state. The forward rate constant k_{AB} then follows from:

$$k_{AB} = f_A P(\lambda_B | \lambda_A), \quad (11)$$

where the first factor is the flux f_A to leave region A (defined by $\lambda(x) < \lambda_A$) and the second factor is the conditional probability $P(\lambda_B | \lambda_A)$ to reach region B (defined by order parameter $\lambda(x) > \lambda_B$) once the surface λ_A defining region A is passed. The flux factor can be measured by starting an MD simulation in stable state A . The interface λ_A will be crossed often, resulting in a statistically accurate value for f_A . In contrast, the value of $P(\lambda_B | \lambda_A)$ is very low for a high barrier and cannot be measured directly. The statistics can again be considerably improved by employing a biased path sampling scheme. By sampling paths with the constraint that they come from region A , cross an interface λ_i , and then either go to B or return to A , we can measure the probability

$P(\lambda|\lambda_A)$ to reach values of λ larger than λ_i . These probabilities are binned in a histogram, and after having performed several simulations for different interfaces λ_i , these histograms are joined into a master curve, from which the value of $P(\lambda_B|\lambda_A)$ can be extracted. A more detailed description of TIS can be found in [65]. TIS allows the variable path length description, thereby limiting the required time steps to a minimum. The TIS rate equation is based on an effective positive flux formalism and is less sensitive to recrossings. The shifting moves used in TPS to enhance statistics are unnecessary in the TIS algorithm. Also, multidimensional or discrete order parameters can easily be implemented in TIS. Up to now TIS has only been applied to relatively small biomolecular systems [72, 73]. The forward flux sampling method (FFS) put forward by ten Wolde and collaborators [74] also makes use of the same formalism but differs in the particular way of generating trajectories for the calculation of the probabilities. For a comparison of TIS and FFS see [80].

Partial Path TIS. Some recently developed methods make use of the diffusive nature of paths and the corresponding loss of correlation. For instance, the partial path TIS technique (PPTIS) takes advantage of this loss of correlation by describing the rare event as a Markovian hopping process between the interfaces, with corresponding hopping probabilities. If the loss of correlation between three consecutive interfaces is justified, this method exactly gives the kinetics [81]. The milestone approach of Elber and collaborators is based on similar ideas [82].

Discrete Path Sampling. An entirely different approach for the calculation of rate constants is followed in the discrete path sampling scheme (DPS) of Wales [18]. Here, pathways between long-lived stable states are viewed as consisting of many transitions between local energy minima. Overall rate constants are then computed from the rate constants for these transitions between minima.

3

Applications

3.1

Biomolecular Isomerization

Conformational changes of biomolecules play a crucial role in biological function, for instance in protein folding, but also in signal transduction, and the mechanics of molecular motors. The complexity of such processes and the timescales involved render a straightforward approach impractical. Moreover, the dynamical variables for biomolecular isomerization processes are not known a priori and are most likely multidimensional. In principle,

these variables can be revealed using TPS methodology. In one of the earliest attempts to apply TPS on a biomolecular system [83] we tried to find the reaction coordinates involved in isomerization of the simplest of peptides: alanine dipeptide. The common peptide dihedral angles ϕ and ψ are satisfactory order parameters for describing the stable states. However, this does not imply that they also provide an accurate description for the dynamics of the transition. We established, by analysis of the transition path ensembles for isomerization of alanine dipeptide in both water and vacuum, that other variables in addition to ϕ and ψ are important. In these simulations we employed the AMBER 94 force field, with TIP3P water as a solvent. For alanine dipeptide isomerization in vacuum (from C_{7eq} to C_{ax}) we found by thorough committor distribution analysis that there is at least a third angle involved in the dynamical transition. We identified this angle to be the O – C – N – C $_{\alpha}$ torsion angle. When this angle was constrained, the committor distribution turned out to be peaked at around $p_B = 0.5$. More recent work by Maragliano et al. [84] using the finite temperature string method suggests that including the C $_{\alpha}$ – C – N – H torsion angle, besides the above three, is required to make the committor peaked.

For alanine dipeptide isomerization in TIP3P water (from C_{eq} to α_R) it was found that the solvent degrees of freedom play a role and should be incorporated in a dynamical treatment of the process [83]. Ma and Dinner performed an exhausting search of many possible reaction coordinates and concluded that a complex coordinate involving the electrostatic torque on a particular hydrogen atom arising from the solvent is the best candidate for a description of the reaction coordinate for alanine dipeptide isomerization [85]. In particular, solvent fluctuations in the second solvation shell contributed to this torque. It is therefore not surprising that previous attempts to find these contributions by investigating solvent flux correlations did not succeed [86].

A general conclusion following from the above work is that the influence of the solvent on the reaction coordinate can be subtle. Furthermore, commitment times might be long on a molecular scale. Such issues will render the study of isomerization difficult with deterministic TPS, in particular for larger molecules. A solution to this problem could be to abandon a deterministic approach and describe the transition as a stochastic process, as is done in PPTIS [81], milestoneing [82], and the finite temperature string method [84].

3.2

Binding and Unbinding of DNA Base Pairs

Hagan et al. [87] investigated the room temperature binding and unbinding pathways of a base pair in a CGC DNA oligomer. The aim was to understand the microscopic details of DNA hybridization and melting to validate existing coarse-grained models of DNA. Employing TPS with the CHARMM force

field and TIP3P water they found several possible pathways for the unbinding process. The first path involved hydrogen bond breaking via twisting of the base pairs followed by unstacking of the C-base. In the second path the C-base simultaneously breaks the stacking interaction and hydrogen bonds with the G-base. The authors found, by combining free energy calculations with path sampling, that four parameters/coordinates are important for the unbinding transition. Two monitor the interaction energy between the flipping base and its nearest neighbors in the native structure. The other two describe the distance between the centers of mass of the base and its neighbors. The first pair of coordinates describes the first part of the transition well, where the twisting changes the interaction energies significantly. The second pair of parameters describes the final region well, where the bases are far apart.

The transition states were located in a flat region in the free energy landscape as a function of the distances. The TSE is not determined by the breaking of DNA–DNA hydrogen bonds or subsequent formation of DNA–water hydrogen bonds. It is also not the penetration or disappearance of water between stacking bases, as would be the case for hydrophobic interaction. Rather the unbinding transition requires the base to diffuse away so far that re-formation of hydrogen bonds is unlikely. The unbound state is hence stabilized by entropic forces. However, this entropic stabilization is not as strong as in the case of protein unfolding.

3.3

Protein Folding

Folding lies at the heart of the stability and functioning of each protein in the living cell. The theoretical understanding of protein folding, although much improved over the past decades, is still far from complete. What exactly determines how a protein folds into a single native state? What is the ensemble of reactive pathways? Can we predict the final structure? These are questions of crucial impact for molecular biology. In this section, we summarize briefly the current understanding of two-state folding, which most single domain proteins obey [88–95], and describe some recent TPS work applied to this problem [96].

Two-state proteins have only two stable states, native and unfolded, and therefore show exponential relaxation behavior. The folding process is conceptually explained in terms of energy funnel landscapes. The unfolded state has abundant conformational entropy but a high potential energy as well. The hydrophobic effect drives the collapse of the protein into a molten globule, a compact solvated state, where the protein still can explore many conformations. Before the transition to the native state can occur via formation of all native contacts in combination with water expulsion, the protein first has to find a transition state. This is a state topologically similar to the native state in which a nucleus of tertiary contacts is formed. Some proteins form

the secondary structure elements first, before they form the tertiary structure (the so-called diffusion collision mechanism). Others can only stabilize their secondary structures after the nucleation event, (the so-called nucleation-condensation mechanism).

Simulation has contributed much to the understanding of folding but it is notoriously hard to simulate folding events using accurate atomistic force fields, first of all because of the long time scales involved. A protein folds in microseconds or more, practically prohibiting the use of straightforward implementation of simulating techniques. Second, the reaction coordinate is very complex. This prevents the use of biased sampling (see Sect. 1).

Many simulation methods have been developed since the 1980s to circumvent this problem. One way to speed up the simulations is by coarse graining, e.g., using implicit solvents [75,76], Go-like models [77–79], and lattice models [92]. An alternative approach is path sampling, for which the initial and final states have to be known. Several groups have used two-ended path sampling methods to study folding events [97,98]. These studies revealed details about the specific folding transitions but, because of the large time step between the time slices, do not represent the kinetics accurately. Here, we will focus only on applications employing transition path sampling or TPS related methods.

TPS of hydrophobic collapse. The first step in protein folding is the hydrophobic collapse. Ten Wolde and Chandler [99] studied the collapse of a hydrophobic homopolymer in solvent. The polymer was described with a bead-string model, while the solvent was modeled on an Ising-like lattice, capable of showing a liquid–gas phase transition. Using path sampling they observed a significant cooperative drying effect around the polymer due to hydrophobic interactions. One of the main conclusions following from this work was that the collapse pathways did not follow the free energy landscape saddle point because the solvent coordinate changed on a much faster timescale than the polymer was able to diffuse.

TPS of GB1 beta hairpin. The 16 residue C-terminal hairpin of protein G-B1 is a model system to investigate β sheet formation. Straightforward MD shows that at 300 K there is a large barrier to unfolding, preventing the system to unfold or fold spontaneously in an accessible simulation timescale [100–103]. Previous work established that the folding occurs via an intermediate compact state in which the hydrophobic core is formed but no backbone hydrogen bonds yet. The formation of these hydrogen bonds towards the folded state seems to be the rate-determining step in the folding process. Application of TPS techniques to this transition, using the CHARMM22 force field with an explicit TIP3P solvent at room temperature with an aggregate simulation time of several μ s, elucidated the precise kinetic pathways and obtained the folding rates [72,73].

The general picture of this process is that folding starts from a state in which the hydrophobic core holds the two strands together but no hydrogen bonds are present. The hairpin then has to find the transition state conformation in which specific backbone hydrogen bonds are bridged by water. They form a lubricating strip of water between the strands that help to form the final hydrogen bonds. The conformational entropy of the backbone is reduced in this transition state, as is the entropy of the water. When this transition state has been reached the folding can complete by expulsion of this water.

TIS enabled the calculation of the rate constant, which turned out to be in agreement with experiments ($0.17 \mu\text{s}^{-1}$). The TIS results indicate that there is a folding free energy barrier of more than $10 k_B T$. However, the computed free energy barrier is only approximately $3\text{--}4 k_B T$. The large discrepancy between the rate constant and FE results is caused by an overlap between the folded and unfolded states, thus apparently lowering the barriers in the projection on the used representations. It follows that free energy landscapes cannot always be trusted to give correct barrier heights. The low barrier also shows that TST-based methods [104, 105] cannot be used to calculate rates, as they will result in extremely low transmission coefficients.

DPS of GB1 beta hairpin. Evans and Wales investigated the same GB1 hairpin with a discrete path sampling method [106]. Inspired by TPS, the DPS method samples kinetically important pathways from a previously assembled database of metastable states and transition rates between these states obtained via harmonic TST. The authors employed the CHARMM19 force-field with the EEF1 implicit solvent. The resulting folding rate was predicted only an order of magnitude lower than the experimental value. However, there were some discrepancies with previous work, in particular on the stable states. Whether the discrepancies were due to differences in force fields, undersampling of configuration space or to difficulties in comparing the structures is yet to be determined.

TPS of Trp-cage. Juraszek and Bolhuis [107] studied the Trp-cage mini-protein in explicit solvent with TPS. They employed the OPLSAA force field, with SPC water, and obtained the path ensemble in an aggregate simulation time of several μs . They observed very different folding pathways; one folding according to the diffusion-collision mechanism and one following the nucleation-condensation mechanism. The average path length was about 3 ns, indicating that the transition was extremely diffuse. A committor analysis showed that the water dynamics was not part of the reaction coordinate, something that was already suggested for other small proteins by Pande et al. [108]. This study shows that TPS is indeed capable of finding entire different pathways in a complex system.

3.4

Enzyme Catalysis

Enzymatic reactions are extremely important in molecular biology. The difficulty in studying them by computer simulation lies, besides in the quantum nature of the reaction, in describing the reaction coordinate. In a few studies so far, TPS has enabled researchers to test proposed reaction coordinates, which in most cases led to new insights.

TPS of DNA polymerase β . Radhankrishnan and Schlick [109–111] performed path sampling simulations to investigate the mechanism and fidelity of the DNA repair process by polymerase. To that end they compared the kinetic pathways of the incorporation of a correct nucleotide versus that of an incorrect base pair. Because this process consists of many sequential reactions they split the problem into a series of transitions, a reaction network. Employing QM/MM (Gamess-UK with B3LYP in combination with CHARMM) simulation for the trajectory generation they obtained path ensembles for each of the transitions, requiring an aggregate amount of about 100 ns of QM/MM simulation. The conclusion is that the mismatched base pair in the polymerase indeed leads to a higher barrier, and hence enables a selection criterion. (The method used in this work to estimate the free energy barrier is in our view slightly biased. See, e.g., [112] for an correct treatment).

TPS of Lactate Dehydrogenase. Basner and Schwartz [113] have applied TPS to the enzymatic reaction of lactate dehydrogenase (LDH). LDH catalyzes the interconversion of the hydrogen acid acetate and the keto-acid pyruvate with the coenzyme NADH (nicotinamide adenine dinucleotide). The reaction involves a proton transfer between the histidine of the active site and the oxygen of the lactate, in addition to a hydride transfer between the NADH coenzyme. Previous work revealed that the residues close to the active site have influence on the coenzyme binding, substrate binding and orientation, and on the reaction event. The order of the proton and hydride transfer was still controversial.

Using the QM/MM methodology (AM1 with CHARMM27) Basner and Schwartz performed TPS of the proton transfer and hydride transfer. Due to the demanding calculation, the aggregate simulation time was only a few ps. The authors argue that the mechanism for the LDH-catalyzed reaction from pyruvate to lactate involves both sequential and concerted proton and hydride transfers. They base this conclusion on the fact that both possibilities are found in the transition path ensemble. Large scale residue motions are paramount to the success of the reaction. The important residues behind the hydride donor, the coenzyme NADH, move toward the active site, while the residues located behind the substrate move away from it. Examination of unreactive trajectories helped in revealing this mechanism. Fluctuations

and compression of the residues in the binding pocket turned out to be important. An enzyme-wide collective motion was observed as a contributing factor of the enzyme rate acceleration. Whereas previous studies always had to assume the reaction coordinate in advance, this work is the first to successfully employ TPS for an enzymatic reaction, for which this assumption is not required.

3.5

Lipid Bilayers

Living cells and their intracellular organelles are protected from the environment by membranes: lipid bilayers. These bilayers serve as compartment boundaries, provide mechanical strength, host a wide variety of functional proteins, and act as a capacitor to maintain ionic gradients between the inside and outside of the cell. The computational study of lipid bilayers has taken an enormous flight [114–117]. Many studies have been carried out using atomistic force fields, but to incorporate the bilayers' flexibility, the system sizes need to be large. Most simulation approaches therefore use simplified coarse grained potentials.

Flip-flop of lipids in membranes. Marti and Csajka [118, 119] investigated the so-called flip-flop transition where a lipid molecule with its head-group on one side of the membrane migrates to the other side. Flip-flop has been observed in naturally occurring as well in synthetic membranes. The difficulty of studying such flip-flops lies in the fact that they are very rare. The hydrophilic head-group must go through the hydrophobic interior of the membrane, something that is very unfavorable. The transition therefore happens only on the order of seconds in a simulation with a reasonable system size.

Marti and Csajka employed a simplified lipid model, consisting of a bead spring model with LJ interaction and soft core repulsions, for hydrophobic interactions. They obtained an ensemble of 200 pathways of flip-flop by TPS at different temperatures. The average transition path was found to take less than a picosecond. This separation of timescales is caused by the height of the barrier, which was shown to be of the order of 10 kJ/mol at ambient temperature. The transition state ensemble was also obtained and showed very clearly that a flipping lipid molecule adopts a “rolled up” configuration in the middle of the membrane (i.e., parallel to the membrane interface). Also the amount of disorder/fluctuation was found to be higher in the transition state.

3.6

Switches in Biochemical Networks

Biochemical switches are networks of chemical reactions with more than one steady state. The switches between those states can occur due to fluctuations

in the system. Such networks are ubiquitous in living cells, e.g., the lac repressor in *E. Coli* and cell cycle regulation. Simulating such networks has relied heavily on stochastic simulation methods. However, the stability of the steady states in these networks make straightforward simulations exceedingly long. Transition path sampling would be able alleviate this problem, if it were not for the problem that most biochemical networks are out of equilibrium and hence lack detailed balance. Allen et al. introduced in [74] a novel method, based on TIS, that allows one to evaluate the kinetics of such biochemical switches. Named forward flux sampling (FFS), the method generates the trajectories in a ratchet-like manner. It first requires the definition of a series of interfaces in the same way as in TIS. Starting from state A , trajectories are generated using stochastic dynamics, and the probability of reaching the first interface is computed. The crossing points of the first interface are recorded and used for the generation of the new trajectories. This procedure is iterated until the last interface and the final state is reached. The difference from TIS is that only forward trajectories are generated, hence allowing for time-irreversible dynamics.

Allen et al. applied this method to a model genetic switch and found, although the switch is symmetric on exchanging the final and initial states, that the path ensemble ($A \rightarrow B$) differs from ($B \rightarrow A$). This implies that the distribution of transition paths does not follow the steady state phase space density.

4

Outlook

In this review article we have given an overview of the TPS methods in relation to biologically inspired problems. In biology most process are complex, and mechanistic molecular insight is sometimes hard to achieve. Molecular simulation techniques have advanced to the point that complex biological systems can be modeled in atomistic detail. This advancement complements the experimental trend towards techniques resolving these systems at the molecular level. Characteristic of these biologically important systems are the large differences in timescales that are relevant for functional activity. TPS can bridge this difference in timescales.

In the past few years several groups have applied TPS to widely different biologically relevant processes. While the first applications were on rather modest systems, in past few years it has become clear that even complex processes such as protein folding can be treated with TPS. As available computer power increases continuously, we can expect TPS studies on even more complex systems in the near future. In particular, the field of enzymatic reactions might show a growth in TPS applications, as these reactions are notoriously hard to study with straightforward MD (QM/MM). We also might expect applica-

tion to other problems such as translocation, DNA condensation, transport through ion channels, protein–DNA complexes, protein binding, and other conformational changes. The work on biological networks emphasizes that the TPS methods are not limited to molecular dynamics. It is even conceivable that they can be used for outstanding problems in population biology and evolution.

From the methodology point of view there is room for many improvements. The problem of intermediate metastable states has not been solved satisfactorily. The straightforward solution to this problem is to cut the overall reaction into several subprocess, as was for instance done in [109]. However, this approach requires a lot of knowledge about the process. The use of Markovian state models or equilibrium kinetic networks might help in this respect [23]. Also, the difficulty of sampling multiple pathways still exists. The work on protein folding showed that it is sometimes possible to find qualitatively different pathways. Nevertheless it is very well possible that in other cases finding different mechanisms requires alternative approaches such as parallel tempering [49]. Besides the increased computational efficiency offered by path sampling, much can be gained by the use of more accurate coarse-grained force fields. Improvement in this area most certainly will enable the treatment of larger and more complex biological systems.

To summarize, we have shown that TPS has the potential to become a standard simulation tool in theoretical molecular biology studies. As rare events play an important role in the life sciences it is very well possible that TPS and its related techniques can make the connection from the single molecule level all the way up to system biology, population biology, and maybe even evolutionary biology.

References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) *Molecular biology of the cell*, 4th edn. Garland Science, New York
2. Karplus M (ed) (2002) Special issue on molecular dynamics simulations of biomolecules. *Acc Chem Res* 35:321
3. Eyring H (1935) *J Chem Phys* 3:107
4. Wigner E (1938) *Trans Faraday Soc* 34:29
5. Keck JC (1962) *Discuss Faraday Soc* 33:173
6. Bennett CH (1977) In: Christofferson R (ed) *Algorithms for chemical computations*, ACS Symposium series no. 46. Am Chem Soc, Washington DC
7. Chandler D (1978) *J Chem Phys* 68:2959
8. Torrie GM, Valleau JP (1977) *J Comput Phys* 23:187
9. Kirkwood J (1935) *J Chem Phys* 3:300
10. Carter EA, Ciccotti G, Hynes JT, Kapral R (1989) *Chem Phys Lett* 156:472
11. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) *Science* 303:186
12. Lazaridis T, Paulaitis ME (1994) *J Am Chem Soc* 116:1546
13. Roux B, Karplus M (1991) *J Phys Chem* 95:4856

14. Truhlar DG, Garrett BC (1984) *Ann Rev Phys Chem* 35:159
15. Cerjan CJ, Miller WH (1981) *J Chem Phys* 75:2800
16. Doye JPK, Wales DJ (1997) *Z Phys D* 40:194
17. Munro LJ, Wales DJ (1999) *Phys Rev B* 59:3969
18. Wales DJ (2002) *Mol Phys* 100:3285
19. Wales DJ (1005) *Phys Biol* 2:S86
20. Evans DA, Wales DJ (2003) *J Chem Phys* 119:9947
21. Evans DA, Wales DJ (2004) *J Chem Phys* 121:1080
22. Deuffhard P, Huisinga W, Fischer A, Schütte Ch (2000) *Lin Alg Appl* 315:39
23. Elmer SP, Park S, Pande VS (2005) *J Chem Phys* 123:114902
24. Voter AF (1998) *Phys Rev B* 57:R13985
25. Shirts MR, Pande VS (2001) *Phys Rev Lett* 86:4983
26. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ, Zagrovic B (2003) *Biopolymers* 68:91
27. Voter AF (1997) *Phys Rev Lett* 78:3907
28. Hamelberg D, McCammon JA (2005) *J Am Chem Soc* 127:13778
29. Grubmüller H (1995) *Phys Rev E* 52:2893
30. Schulze BG, Grubmüller H, Evanseck JD (2000) *J Am Chem Soc* 122:8700
31. Huber T, Torda AE, van Gunsteren WF (1994) *J Comput-Aided Mol Design* 8:695
32. Laio A, Parrinello M (2002) *Proc Natl Acad Sci USA* 99:12562
33. Ensing B, de Vivo M, Liu Z, Moore P, Klein ML (2006) *Acc Chem Res* 39:73
34. Ceccarelli M, Danelon C, Laio A, Parrinello M (2004) *Biophys J* 87:58
35. Hummer G, Kevrekidis IG (2003) *J Chem Phys* 118:10762
36. Pratt LR (1986) *J Chem Phys* 85:5045
37. Elber R, Ghosh A, Cardenas A, Stern H (2004) *Adv Chem Phys* 126:123 (2004)
38. Passerone D, Parrinello M (2001) *Phys Rev Lett* 87:108302
39. Elber R, Karplus M (1987) *Chem Phys Lett* 139:375
40. Sevick EM, Bell AT, Theodorou DN (1993) *J Chem Phys* 98:3196
41. Voter AF (1997) *Phys Rev Lett* 78:3908
42. Gillilan RE, Wilson KR (1996) *J Chem Phys* 105:9299
43. Henkelman G, Johansson G, Jonsson H (2000) In: Schwartz SD (ed) *Progress on theoretical chemistry and physics*. Kluwer Academic, Dordrecht, The Netherlands
44. Mathews DH, Case DA (2006) *J Mol Biol* 357:1683
45. E W, Ren W, Vanden-Eijnden E (2002) *Phys Rev B* 66:052301
46. E W, Ren W, Vanden-Eijnden E (2005) *J Phys Chem* 109:6688
47. Ren W, Vanden-Eijnden E, Maragakis P, E W (2005) *J Chem Phys* 123:134109
48. Dellago C, Bolhuis PG, Csajka FS, Chandler D (1998) *J Chem Phys* 108:1964
49. Dellago C, Bolhuis PG, Geissler PL (2002) *Adv Chem Phys* 123:1
50. Chandler D (1987) *Introduction to modern statistical mechanics*. Oxford University Press, New York
51. Frenkel D, Smit B (2002) *Understanding molecular simulation*, 2nd edn. Academic, San Diego, CA
52. Binder K, Heermann D (2002) *Monte Carlo simulation in statistical physics*. Springer, Berlin Heidelberg New York
53. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) *Ann Rev Phys Chem* 53:291
54. Dellago C, Chandler D (2002) Bridging the time scale gap with transition path sampling. In: Nielaba P, Mareschal M, Ciccotti G (eds) *Molecular simulation for the next decade*. Springer, Berlin Heidelberg New York, p 321
55. Dellago C (2005) Transition Path Sampling. In: Yip S (ed) *Handbook of materials modeling*. Springer, Berlin Heidelberg New York, p 1585

56. Dellago C (2007) Transition path sampling and the calculation of free energies. In: Chipot C, Pohorille A (eds) Free energy calculations: Theory and applications in chemistry and biology. Springer series in chemical physics, vol 86. Springer, Berlin Heidelberg New York (in press)
57. Dellago C, Bolhuis PG, Chandler D (1998) *J Chem Phys* 108:9263
58. Bolhuis PG, Dellago C, Chandler D (1998) *Faraday Discuss* 110:421
59. Car R, Parrinello M (1985) *Phys Rev Lett* 55:2471
60. Nosé S (1984) *J Chem Phys* 81:511
61. Nosé S (1984) *Mol Phys* 52:255
62. Hoover WG (1985) *Phys Rev A* 31:1695
63. Hoover WG, Ladd AJC, Moran B (1982) *Phys Rev Lett* 48:1818
64. Evans DJ (1983) *J Chem Phys* 78:3297
65. van Erp TS, Moroni D, Bolhuis PG (2003) *J Chem Phys* 118:7762
66. Bolhuis PG (2003) *J Phys Cond Matter* 15:S113
67. Onsager L (1938) *Phys Rev* 54:554
68. Klosek MM, Matkowsky BJ, Schuss Z (1991) *Ber Bunsen Phys Chem* 95:331
69. Pande V, Grosberg AY, Tanaka T, Shakhnovich EI (1998) *J Chem Phys* 108:334
70. Hummer G (2004) *J Chem Phys* 120:516
71. Best RB, Hummer G (2005) *Proc Natl Acad Sci USA* 102:6732
72. Bolhuis PG (2003) *Proc Natl Acad Sci USA* 100:12129
73. Bolhuis PG (2005) *Biophys J* 88:50
74. Allen R, Warren PB, ten Wolde P (2005) *Phys Rev Lett* 94:018104
75. Qiu D, Shenkin PS, Hollinger FP, Still EC (1997) *J Phys Chem A* 101:3005
76. Lazaridis T, Karplus M (1999) *Prot Struct Func Gen* 35:133
77. Taketomi H, Ueda Y, Go N (1975) *Int J Pept Protein Res* 7:445
78. Clementi C, Garcia AE, Onuchic JN (2003) *J Mol Biol* 326:933
79. Liwo A, Khalili M, Scheraga HA (2005) *Proc Natl Acad Sci USA* 102:2362
80. van Erp TS (2006) *Condens Mat*, abstract cond-mat/0605260
81. Moroni D, Bolhuis PG, van Erp TS (2004) *J Chem Phys* 120:4055
82. Faradjian AK, Elber R (2004) *J Chem Phys* 120:10880
83. Bolhuis PG, Dellago C, Chandler D (2000) *Proc Natl Acad Sci USA* 97:5877
84. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) *J Chem Phys* 125:024106
85. Ma A, Dinner AR (2005) *J Phys Chem B* 109:6769
86. McCormick TA, Chandler D (2003) *J Phys Chem B* 107:2796
87. Hagan ME, Dinner AR, Chandler D, Chakraborty AK (2003) *Proc Natl Acad Sci USA* 100:13922
88. Dobson CM (2004) *Semin Cell Dev Biol* 15:3
89. Fersht A (1999) *Structure and mechanism in protein science*. Freeman, New York
90. Cheung MS, Chavez LL, Onuchic JN (2004) *Polymer* 45:547
91. Kubelka J, Hofrichter J, Eaton WA (2004) *Curr Opin Struc Biol* 14:76
92. Mirny L, Shakhnovich E (2001) *Annu Rev Bioph Biom* 30:361
93. Shea JE, Brooks CL (2001) *Annu Rev Phys Chem* 52:499
94. Gnanakaran S et al. (2003) *Curr Opin Struc Biol* 13:168
95. Snow CD, Sorin EJ, Rhee YM, Pande V (2005) *Annu Rev Biophys Biomol Struct* 34:43
96. Bolhuis PG (2006) Sampling kinetic protein folding pathways using all atom models. In: Landau DP, Lewis SP, Schüttler H-B (eds) *Computer simulations in condensed matter: from materials to chemical biology*. Springer proceedings in physics. Springer, Berlin Heidelberg New York (in press)
97. Eastman P, Gronbech-Jensen N, Doniach S (2001) *J Chem Phys* 114:3823

98. Elber R, Ghosh A, Cardenas A, Stern H (2003) *Adv Chem Phys* 126:93
99. ten Wolde PR, Chandler D (2002) *Proc Natl Acad Sci USA* 99:6539
100. Pande VS, Rokhsar DS (1999) *Proc Natl Acad Sci USA* 96:9062
101. Roccatano D, Amadei A, Di Nola A, Berendsen HJC (1999) *Protein Sci* 8:2130
102. Garcia AE, Sanbonmatsu KY (2001) *Proteins* 42:345
103. Zhou RH, Berne BJ, Germain R (2001) *Proc Natl Acad Sci USA* 98:14931
104. Chandler D (1978) *J Chem Phys* 68:2959
105. Bennett CH (1977) In: Christofferson R (ed) *Algorithms for chemical computations: ACS Symposium series No 46*. Am Chem Soc, Washington DC
106. Evans DA, Wales DJ (2004) *J Chem Phys* 121:1080
107. Juraszek J, Bolhuis PG (2006) *Proc Natl Acad Sci USA* (in press)
108. Rhee YM et al. (2004) *Proc Natl Acad Sci USA* 101:6456
109. Radhakrishnan R, Schlick T (2004) *Proc Natl Acad Sci USA* 101:5970
110. Radhakrishnan R, Yang LJ, Arora K, Schlick T (2004) *Biophys J* 86:34A
111. Radhakrishnan R, Schlick T (2005) *J Am Chem Soc* 127:13245
112. Moroni D, van Erp TS, Bolhuis PG (2005) *Phys Rev E* 71:056709
113. Basner JE, Schwartz SD (2005) *J Am Chem Soc* 127:13822
114. Goetz R, Lipowsky R (1998) *J Chem Phys* 108:7397
115. Smondyrev AM, Berkowitz ML (1999) *J Comput Chem* 20:531
116. Marrink SJ, Jähnig F, Berendsen HJC (1996) *Biophys J* 71:632
117. Bemporad D, Essex JW, Luttmann C (2004) *J Phys Chem B* 108:4875
118. Marti J, Csajka FS (2004) *Phys Rev E* 69:061918
119. Marti J (2004) *J Phys: Condens Mat* 16:5669

PcrA Helicase, a Molecular Motor Studied from the Electronic to the Functional Level

Markus Ditttrich · Jin Yu · Klaus Schulten (✉)

Beckman Institute, University of Illinois at Urbana-Champaign, Urbana,
 Illinois, IL 61801, USA
kschulte@ks.uiuc.edu

1	General Introduction	320
2	PcrA Helicase, a Prototype Molecular Motor	322
3	Zooming in on ATP Hydrolysis in the Catalytic Site	323
4	Molecular Dynamics Simulations of PcrA Function	332
5	Langevin Simulations of PcrA Function	339
6	Summary	344
	References	346

Abstract Molecular motors are adenosine tri-phosphate (ATP) hydrolysis-driven, cellular proteins responsible for a wide variety of different tasks, such as transport, energy metabolism, and DNA processing. Their operation cycle spans a wide range of length and time scales, from the localized and fast chemical reaction in the catalytic site(s) to the large scale and much slower conformational motions involved in the motors' physiological function. From a computational point of view, this means that currently there exists no single approach capable of capturing the whole spectrum of events during molecular motor function. In the present review, we show for PcrA helicase, a molecular motor involved in the unwinding of double-stranded DNA, how a combination of computational approaches can be used to examine PcrA's function in its entirety as well as in detail. Combined quantum mechanical/molecular mechanical simulations are used to study the catalytic ATP hydrolysis event and its coupling to protein conformational changes. Molecular dynamics simulations then provide a means of studying overall PcrA function on a nanosecond time scale. Finally, to reach physiologically relevant time scales, i.e., milliseconds, stochastic simulations are employed. We show that by combining the three stated approaches one can obtain insight into PcrA helicase function.

Keywords ATP hydrolysis · DNA helicase · Molecular dynamics simulations · Molecular motor · QM/MM · PcrA · Stochastic modeling

Abbreviations

ADP adenosine di-phosphate
 ATP adenosine tri-phosphate
 dsDNA double-stranded DNA

DNA	deoxyribonucleic acid
MD	molecular dynamics
MM	molecular mechanics
nt	nucleotide
P _i	phosphate
QM	quantum mechanics
QM/MM	quantum mechanical/molecular mechanical
RESP	restrained electrostatic potential
RNA	ribonucleic acid
SMD	steered molecular dynamics
ssDNA	single-stranded DNA
vdW	van der Waals

1

General Introduction

Molecular motors are proteins essential for cellular transport, metabolism, and signaling. They are powered by the energy stored in nucleotide triphosphates, mostly adenosine tri-phosphate (ATP). At the heart of molecular motor function is the conversion of the chemical energy stored in ATP into a mechanical force which drives the physiological function of the system. The chemo-mechanical coupling underlying the conversion of ATP's energy is still poorly understood. Indeed, even though several molecular motor proteins are known at a structural level, e.g., myosin, kinesin, F₁-ATPase, or helicase, and even though for each of them a vast amount of biochemical data is available, decades of study have not revealed how the chemo-mechanical coupling is being achieved in microscopic detail by any of these systems.

However, the availability of structural information allows computational studies to be conducted and a significant number have been reported over the past decade for such diverse systems as myosin [1–5], kinesin [3, 6], F₁-ATPase [7–12], and helicase [13–17]. For F₁-ATPase, e.g., molecular dynamics (MD) simulations provided insight into the inner workings of the rotation of the central stalk inside the hexameric head-unit [8]; combined quantum mechanical/molecular mechanical (QM/MM) simulations of the ATP hydrolysis reaction in the catalytic sites [10–12] on the other hand revealed a mechanism for efficient ATP catalysis and were able to identify several residues involved in the coupling of the chemical reaction to larger scale conformational changes of the protein. On the basis of these available structures and biochemical data, Oster and coworkers [7] have developed stochastic and kinetic models for F₁-ATPase function.

Molecular motor proteins present a number of formidable challenges to the computational researcher. First, their operation involves time scales covering several orders of magnitude, from the fast chemical reaction step which occurs over the course of a few femtoseconds all the way to the mechani-

cal motions responsible for the systems' physiological function taking place on a microsecond up to a millisecond time scale. Second, the events during motor function span a wide range of spatial scales, from the localized and electronic level changes accompanying the bond-breaking event in the catalytic binding site(s) during ATP catalysis ($\sim 1 \text{ \AA}$) to the large scale structural changes involved in physiological function ($10 \text{ \AA}$ to 10 nm).

From a methodological point of view this means that there is presently no single computational approach which is able to model the whole spatial and temporal scale of events underlying molecular motor function. At least three levels of computational methodology are necessary to capture the relevant degrees of freedom. A QM or QM/MM method is required to investigate the electronic events underlying the ATP hydrolysis reaction and its interaction with the protein environment. Typically, however, only localized structural changes and/or relatively short time scales can be accessed using QM/MM simulations and MD is needed to study the properties of the full protein in a solvent environment on a multi nanosecond time scale at the expense of neglecting electronic level processes such bond-breaking. In conjunction with methods such as steered MD (SMD) [48, 49] this allows one to study the larger scale motions underlying motor function, e.g., conformational changes coupled to binding of reactant ATP or unbinding of hydrolysis products ADP and P_i as well as protein movements relayed to the motors' binding partners such as actin in case of myosin. However, using even the most powerful computational resources currently available, the time scales accessible to all-atom MD simulations are still orders of magnitude away from the true physiological ones. Hence, in order to reach the latter regime one has to resort to an approach which focuses on a small number of essential degrees of freedom. One might, for example, model the motor function via the center of mass motion of the relevant protein domains on a low-dimensional potential energy surface governed by a Langevin-type stochastic equation.

The challenge from a modeling point of view, then, is to connect these approaches to obtain a coherent picture of the overall process. Ideally, the insight gained on each level is used to design and improve the methodology of the other levels, possibly in an iterative fashion. As an example, the essential degrees of freedom and the potentials governing the stochastic equations modeling the long time behavior of the system could be extracted from appropriately designed MD simulations. Conversely, insight gained from the stochastic modeling might then catalyze an improved setup of the MD simulations.

Here, we summarize progress on how such a comprehensive computational approach can be used to gain insight into the function of one particular molecular motor, the DNA helicase PcrA. This motor protein will be introduced in the next section. We will then lay out the computational approaches used and finally summarize and interpret recent computational findings.

2

PcrA Helicase, a Prototype Molecular Motor

DNA helicases are important participants in DNA metabolism, which involves processes such as replication, transcription, and repair. They catalyze the separation of double-stranded DNA (dsDNA) into its single-stranded DNA (ssDNA) components. Helicases are typically found to exist in a monomeric [18, 19], dimeric [20], or hexameric form [21] and unwind DNA either in a 3' → 5' or 5' → 3' direction.

PcrA is a monomeric 3' → 5' helicase, and due to its relatively small size (80 kDa) constitutes an ideal candidate for computational study. Indeed, PcrA is one of the smallest molecular motors known to date. Several atomic resolution X-ray structures of PcrA have been reported [18, 19]. The present discussion will focus on two PcrA structures from *Bacillus stearothermophilus* that were determined by Velankar et al. [19] in a substrate (with bound ATP analogue AMP-PNP) and product (without ATP/ADP bound) state at 3.3 Å and 2.9 Å resolution, respectively. Figure 1 shows a molecular view of PcrA which consists of four major domains called 1A, 1B, 2A, and 2B. Also visible is a strand of ssDNA threading through the protein at the interface formed by the A and B domains and which is connected to a short piece of dsDNA bound to the 2B domain. ATP binds to the catalytic binding pocket located at the interface between the 1A and 2A domains.

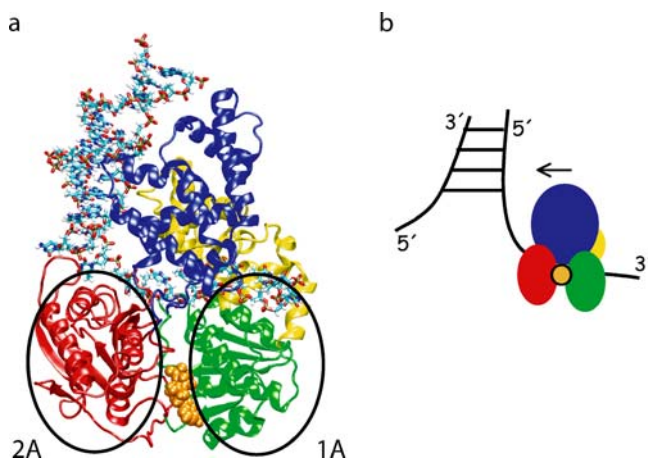


Fig. 1 Molecular view of PcrA helicase. **a** Depicted is the protein in cartoon representation. Also shown is a strand of ssDNA in licorice representation threading through the center of PcrA at the interface of the 1A (green) and 2A domains (red) and which is attached to a short piece of dsDNA adjacent to the 2B domain (blue); also shown is the 1B domain (yellow) and a single molecule of ATP (orange) bound to the catalytic site at the interface between the 1A and 2A domains (depicted in vdW representation). **b** Shown is a schematic view of PcrA translocating along ssDNA toward the dsDNA fork; domains 1A, 2A, 1B, and 2B are shown in green, red, yellow, and blue, respectively

In the following, we will focus on the mechanism of PcrA movement along ssDNA. On the basis of their crystal structure data and on the intuitive idea of alternating affinities, Velankar et al. [19] proposed an “inchworm” model for the translocation of PcrA along ssDNA that involves concerted movements of ssDNA bases across the interface with PcrA. This model does not, however, address how ATP hydrolysis is coupled to domain motion, what residues are most crucial during this process, how the proposed alternating affinities can be justified microscopically and finally, how uni-directionality is achieved. These are the questions that will be discussed in the following sections. We will start by looking at the ATP hydrolysis reaction in the catalytic site and then work our way up toward longer time and larger length scales.

3

Zooming in on ATP Hydrolysis in the Catalytic Site

The catalysis of ATP lies at the heart of PcrA function and takes place in a single catalytic site at the interface between the 1A and 2A domain (Fig. 1). Several important questions arise when considering the role of ATP catalysis during the overall translocation process of PcrA along ssDNA. First and from a purely enzymatic point of view, one would like to know how efficient ATP catalysis is being achieved by the protein. Second, one would like to investigate how the actual catalytic step is coupled to larger scale protein conformational changes eventually leading to PcrA translocation along ssDNA.

An investigation of both questions requires a proper treatment of the electronic degrees of freedom during the catalytic reaction in the biological context provided by the protein environment. QM/MM methods present a powerful and versatile way of addressing these issues, since they allow one to conduct high level electronic structure calculations of a particular region of interest in the presence of a classically treated protein environment. The following paragraph provides a brief outline of the principles behind QM/MM calculations, before we describe some of the insights into PcrA function that have emerged using this approach.

QM/MM Calculations

The underlying theme of QM/MM simulations [22–29] is a quantum mechanical treatment of a certain molecular core region that is undergoing electronic level changes in the presence of a (large) biomolecular environment given by, for example, protein, solvent, or DNA/RNA. The environment itself is assumed to be electronically inert and is, therefore, treated classically, typically with a molecular mechanics method. The main motivation for this separation into a quantum mechanically and a classically treated region is the fact that quantum mechanical calculations are computationally very expensive

and can only be applied to relatively small systems of at most a few hundred atoms. Hence, QM/MM simulations provide a compromise between the need to treat certain parts of the system at an electronic level and the requirement for the presence of the sizeable host biomolecule or biomolecular complex.

Depending on the particular QM/MM method chosen, it is sometimes necessary to limit the number of classically treated MM atoms. In the case of PcrA which contains more than 110 000 atoms in its solvated state (Sect. 4), the conducted QM/MM calculations described below were restricted to a reduced sub-system containing only 20 000 atoms that was centered around the nucleotide ATP bound to the catalytic site. This is depicted schematically in Fig. 2, which shows the full QM/MM system indicating the quantum mechanically (QM) and classically treated regions (MM), respectively.

The remainder of this section will provide a brief overview of QM/MM simulations in general and the implementation used in the studies summarized in the present manuscript in particular. In all of the following it is assumed that the Born–Oppenheimer approximation is valid, i.e., electronic and nuclear degrees of freedom can be separated. In this case, the QM/MM Hamiltonian describing the quantum mechanically modeled core region and the classically treated protein environment can formally be expressed as

$$H = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM/MM}} , \quad (1)$$

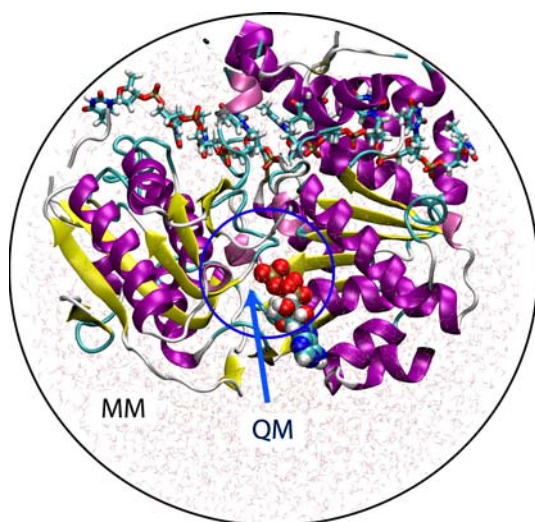


Fig. 2 QM/MM simulation system for PcrA. Shown is the complete model which is centered around the catalytic binding site located at the interface between the 1A and 2A domains and contains about 20 000 atoms. The protein is shown in cartoon representation, a piece of bound ssDNA is depicted in licorice representation, and bound ATP in vdW representation. The classically (MM) and quantum mechanically (QM) treated regions are indicated by *circles* and labeled accordingly

where H_{QM} can be identified with the Hamiltonian of the time-independent Schrödinger equation

$$H_{\text{QM}} |\psi(\{\mathbf{r}\})\rangle = E |\psi(\{\mathbf{r}\})\rangle \quad (2)$$

and the classical contribution H_{MM} is given by Eq. 14 below.

The main challenge for achieving a proper QM/MM description lies in the treatment of the interface between the quantum and classical regions, described by $H_{\text{QM/MM}}$. QM/MM studies of chemical reactions in solution allow for a straightforward separation of both regions, e.g., the solvent being treated using a classical description and the chemically active solute being modeled by a QM method. QM/MM simulations of biomolecular systems like, for example, enzymes, on the other hand, often lead to interfaces separating the quantum and classical regions that cut through covalent bonds, thereby, complicating the description significantly.

Broadly, one can distinguish three types of QM/MM coupling schemes. These are listed below in the order of increasing complexity:

- *Mechanical Embedding*. This represents the crudest level of description. The QM and MM regions are not connected by any covalent bonds and interact only via steric constraints, typically implemented in terms of a van der Waals (vdW) type interaction:

$$H_{\text{QM/MM}} = \sum_{a \in \text{MM}} \sum_{i \in \text{QM}} H_{\text{vdW}}^{ai} . \quad (3)$$

The two sums extend over all nuclei in the classically (MM) and quantum mechanically (QM) treated region, respectively. The choice of vdW parameters is not straightforward. Often, the parameters provided by the classical forcefield are employed, but QM/MM studies with specifically parameterized vdW interactions have also been reported.

- *Electrostatic Embedding*. At this level of QM/MM coupling, in addition to steric constraints between the QM and MM regions, one also takes into account their electrostatic interaction, leading to a QM/MM Hamiltonian of the form:

$$H_{\text{QM/MM}} = \sum_{a \in \text{MM}} \sum_{i \in \text{QM}} (H_{\text{vdW}}^{ai} + H_{\text{ele}}^{ai}) . \quad (4)$$

Here, H_{ele} describes the electrostatic interaction between the QM and MM regions. As will be discussed in more detail below, this term can be evaluated in several different ways.

- *Covalent Embedding*. This most sophisticated level of QM/MM scheme involves, in addition to steric and electrostatic interactions, covalent bonds between QM and MM atoms. Typically, this is the description necessary for simulations of biomolecular systems, and was used in our QM/MM approach.

The next section is devoted to a more in depth exposition of the particular covalent embedding scheme used in our calculations.

RESP Charge-Based Covalent Embedding

A number of different covalent embedding schemes have been proposed since the early QM/MM studies of Warshel [30]. Here, rather than trying to be comprehensive, we focus on the particular method employed in the QM/MM implementation on which the reported results are based and refer to the literature for additional information. See also the review on QM/MM methods by Senn and Thiel in this volume.

When describing quantum mechanical events inside biomolecules one frequently encounters covalent bonds across the QM/MM boundary connecting classically and quantum mechanically treated atoms. This is problematic for several reasons. First and foremost, “cutting” these bonds creates an open valence in the QM region. Several schemes have been devised to deal with this problem. The probably most straightforward method is the so-called link atom approach [23] that is also used in our QM/MM method. Here, the dangling bond is simply terminated by a “dummy” hydrogen atom. More sophisticated schemes have been proposed, for example, the local self-consistent field method by Théry et al. [31] in which a frozen hybrid orbital with a pre-determined density is used to satisfy the valence at the boundary. However, even though the introduction of an additional (hydrogen) atom in the link atom approach can in principle cause artificial behavior at the boundary, it is currently not clear that alternative, more sophisticated methods provide a much improved description.

A second problematic issue arising from a QM/MM boundary that dissects covalent bonds is the difficulty of achieving a proper separation of the QM and MM interaction terms in order to avoid the double counting of certain interactions. This will be discussed in more detail below.

The QM/MM interface Hamiltonian of the covalent embedding scheme describing the quantum-classical interaction can be written as

$$H_{\text{QM/MM}} = \sum_{a \in \text{MM}} \sum_{i \in \text{QM}} (H_{\text{vdW}}^{ai} + H_{\text{ele}}^{ai}) + H_{\text{covalent}}^{\text{QM/MM}}. \quad (5)$$

The $H_{\text{covalent}}^{\text{QM/MM}}$ contribution maintains the proper conformation across the covalent quantum-classical boundary and is provided by the molecular mechanics force field. Quantum mechanically treated atoms in the boundary region experience all classical bonded force contributions that involve at least one classical atom.

In most conventional QM/MM approaches the electronic contribution to the electrostatic interaction H_{ele}^{ai} in Eq. 5 is given via the one electron

operator

$$H_{\text{ele,e}}^{ai} = \frac{q_a}{|\mathbf{R}_a - \mathbf{r}_i|}, \quad (6)$$

where i indexes the electron coordinates and q_a is the partial charge on the classical atom a at position $\mathbf{R}_a$. Unfortunately, as indicated above, this approach leads to problems at the covalent QM/MM boundary where electrostatic interactions between quantum and classical atoms are already implicitly accounted for via the classical bond, angle, and dihedral energies. Subtracting these interactions from the expression Eq. 6 is difficult since the electronic charge is de-localized over the whole QM segment and, therefore, can not be easily partitioned among atoms. In other words, there is no quantum mechanical charge operator that would allow the assignment of partial atomic charges. To address this shortcoming, our QM/MM approach introduces an effective charge operator $\hat{\xi}_p$ for each atomic site in the QM segment, that is evaluated via the restrained electrostatic potential (RESP) charge method [32]. This leads to the following expression for the electrostatic potential

$$\begin{aligned} H_{\text{ele}} &= H_{\text{ele,e}} + H_{\text{ele,n}} \\ &= \gamma \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{\hat{\xi}_p q_a}{|\mathbf{R}_a - \mathbf{R}_p|} + \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{Z_p q_a}{|\mathbf{R}_a - \mathbf{R}_p|}, \end{aligned} \quad (7)$$

where $H_{\text{ele,e}}$ and $H_{\text{ele,n}}$ are the electronic and nuclear contribution, respectively. Here q_a and Z_p are the partial charges on the classical atoms and nuclear charges of the quantum mechanically modeled atoms, respectively. The κ_{ap} mediate the proper exclusion of the classical non-bonded interactions at the QM/MM boundary, and γ is a scaling factor to ensure the proper total charge of the QM/MM system.

Using the Hamiltonian Eq. 1 together with the QM/MM interaction energy Eq. 5, one obtains the following expression for the total energy E

$$\begin{aligned} E &= \langle \Psi | H | \Psi \rangle \\ &= E^{\text{QM}}(\mathbf{P}) + \gamma \sum_{a \in \text{MM}} \sum_{p \in \text{QM}} \kappa_{ap} \frac{\hat{\xi}_p(\mathbf{P}) q_a}{|\mathbf{R}_a - \mathbf{R}_p|} + E^{\text{MM}}. \end{aligned} \quad (8)$$

Here, $|\Psi\rangle$ is the electronic wave function and $\mathbf{P}$ the corresponding electronic density matrix. The term E^{QM} is the expression for the gas phase energy of the QM part, and E^{MM} includes all remaining non-electronic energy contributions. In the second term, $\hat{\xi}_p(\mathbf{P})$ is the effective charge for atoms in the QM segment, which is evaluated via the RESP method by minimizing the follow-

ing expression

$$I = \sum_{\alpha} \omega_{\alpha} \left[\sum_{p \in \text{QM}} \frac{\xi_p(\mathbf{P})}{|\mathbf{R}_{\alpha} - \mathbf{R}_p|} - V_{\alpha}(\mathbf{P}) \right]^2 + 2\lambda_e \left[\sum_{p \in \text{QM}} \xi_p(\mathbf{P}) - N_e \right] + \sum_{p \in \text{QM}} g_p (\xi_p(\mathbf{P}) + Z_p)^2. \quad (9)$$

The first term on the right-hand side attempts to find effective charges $\xi_p(\mathbf{P})$ that best reproduce the exact electronic electrostatic potential $V_{\alpha}(\mathbf{P})$ subject to the constraints expressed in the second and third term. Here, α sums over the points of a grid located at positions $\mathbf{R}_{\alpha}$ surrounding the QM segment at which the electrostatic potential is evaluated. The ω_{α} are weight factors whose value depends on the particular choice for the computational grid. The second term enforces charge conservation and constrains the overall effective charge to be equal to the total number of electrons N_e via a Lagrange multiplier λ_e . Both $V_{\alpha}(\mathbf{P})$ and N_e have to take into account the presence of the dummy hydrogen atoms. Finally, the third term represents a harmonic penalty function that can be adjusted by the parameter g_p to prevent the appearance of unreasonably large effective charges [32].

The minimization of the expression given by Eq. 9 can be cast into a matrix equation for $\xi_p(\mathbf{P})$ which can then be used to formulate a new Fock operator

$$\mathbf{F}^{\text{QM/MM}} = \mathbf{F} + \mathbf{F}^{\text{solv}}, \quad (10)$$

where $\mathbf{F}$ is the gas-phase Fock operator and $\mathbf{F}^{\text{solv}}$ the so-called solvated Fock operator mediating the electrostatic interaction between the QM and MM segment of the QM/MM system [33]. The self-consistent solution of the Roothaan-Hall equations with the Fock operator given by Eq. 10 is the mathematical basis of the QM/MM approach used in the studies described in the remainder of this section.

Finally, we would like to point out that due to the computational cost involved, the energetics derived from ab initio QM/MM calculations typically are enthalpies rather than free energies. The latter are required, to, for example, derive true reaction free energies or equilibrium constants. Hence, for each particular system it is necessary to take into account possible entropic effects when arguing based on calculated enthalpies. There exist, however, semi-empirical QM/MM approaches that are capable of computing true free energy surfaces, at the possible expense of reduced accuracy and transferability [34, 35].

QM/MM Study of the Chemo-Mechanical Coupling in PcrA Helicase

The QM/MM approach just described was employed to investigate the ATP hydrolysis reaction in the catalytic site of PcrA helicase and its coupling to

protein conformational changes, i.e., the chemo-mechanical coupling [36]. The reactant state conformation of the quantum mechanically treated region is shown in the upper left-hand side panel of Fig. 3 which also depicts important distances between atoms.

Figure 3 reveals that all binding pocket residues immediately adjacent to the tri-phosphate part of ATP are either charged or polar, with several water molecules surrounding the γ -phosphate group. A magnesium ion is located between two oxygen atoms of the β and γ -phosphate groups and is hydrogen-bonded to additional water molecules and binding pocket residues. Most of the residues forming the catalytic binding sites are located in the 1A domain of PcrA with the exception of two arginine residues, R287 and R610, which are contributed by the neighboring 2A domain (c.f. bottom panel of Fig. 3).

Using QM/MM calculations, the ATP hydrolysis reaction pathway can be calculated [36] and the conformations of the reactant, transition, and product state are shown in the upper panel of Fig. 3 with the corresponding energies depicted in Fig. 4.

The center column of Fig. 3 shows the transition state conformation during ATP hydrolysis. It features a hydronium ion that evolved via proton transfer from the nucleophilic water, W1, toward W2. During the conversion from the

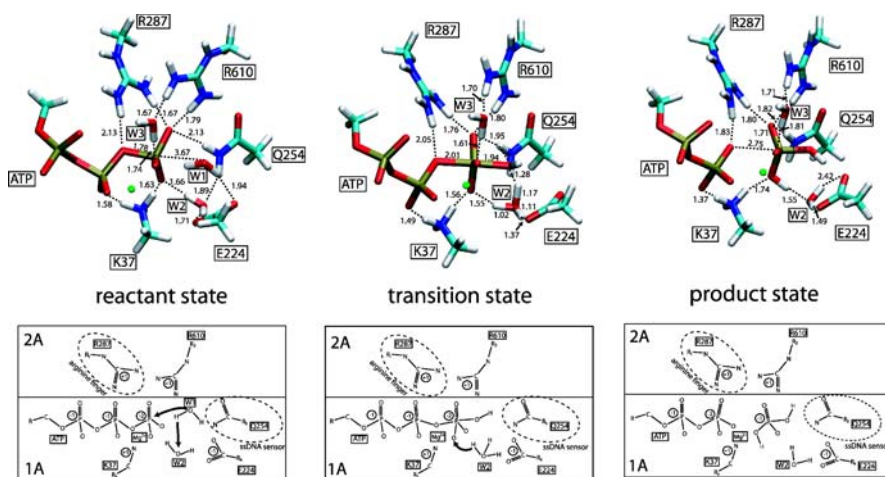


Fig. 3 Conformation of the quantum mechanically treated core region in the catalytic site of PcrA. Shown in the *top panel* are molecular views of the structures of the reactant (*top left*), transition (*top middle*), and final (*top right*) state along the ATP hydrolysis reaction pathway. Important distances between atoms are indicated and given in units of Å. The *bottom panel* depicts schematic views of the catalytic site in each of the three states and shows which of the residues belong to domain 1A or 2A. The *bold arrows* sketch the proton relay mechanism discussed in the text. The *color coding* is as follows: *white* - hydrogen; *light blue* - carbon; *dark blue* - nitrogen; *red* - oxygen; *gold* - phosphorus; *green* - magnesium ion (adapted from [36])

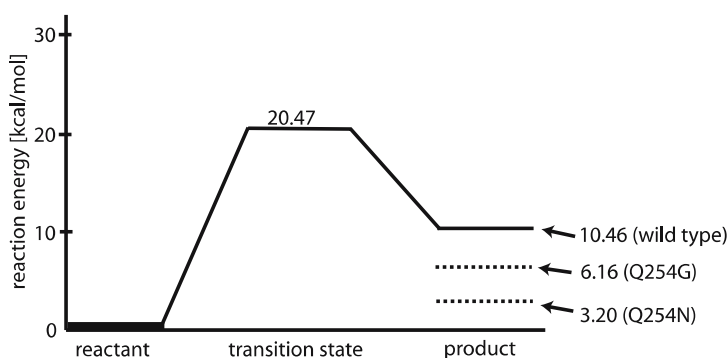


Fig. 4 Energetics of ATP hydrolysis in PcrA. Shown are the energies in kcal/mol of the reactant, transition, and product state during ATP hydrolysis in the catalytic sites of PcrA. For the product state, the energies for the wild type and two mutants, Q254N and Q254G, are provided (adapted from [36])

transition state toward the product conformation, the hydronium ion transfers a second proton to the γ -phosphate group to yield products ADP and P_i . This mechanism of proton transfer from the nucleophilic water toward the γ -phosphate group of ATP is sketched by the bold arrows in the bottom panel of Fig. 3 and was termed a proton relay mechanism [10], since it involves two water molecules, W1 and W2. Until recently, it was thought that direct proton transfer from the nucleophilic water, W1, toward one of the γ -phosphate oxygen atoms of ATP is the physiologically relevant pathway. However, QM/MM studies have shown that the proton relay mechanism is energetically more favorable by up to 20 kcal/mol [36]. Such a mechanism had also been proposed based on theoretical studies of other ATP-driven molecular motors, such as F_1 -ATPase [10, 12] and myosin [1], and had been proposed also based on structural and biochemical evidence for ras [37] and F_1 -ATPase [38]. All combined, this evidence suggests that such a proton relay mechanism is responsible for efficient hydrolysis in a wider class of ATP-driven molecular motors. Since the proton relay mechanism relies on the proper arrangement of water molecules in the binding pocket, which is in turn facilitated by the protein environment, it represents a genuinely enzymatic pathway.

Figure 4 gives the energies of important states along the ATP hydrolysis pathway in PcrA. Several features are noteworthy: First, the QM/MM calculations [36] reveal the presence of a modest transition state barrier of ~ 20 kcal/mol in good agreement with experimentally measured rates ($\sim 10^2$ s $^{-1}$). This moderate barrier is due to the proton relay mechanism as shown by comparison with calculations of direct proton transfer pathways. In addition, there are several binding pocket residues that contribute to a lowering of the transition state barrier. This is shown in Fig. 5 which depicts the change in electrostatic interaction between ATP and the protein environment upon going from the reactant to the transition state. Clearly, K37,

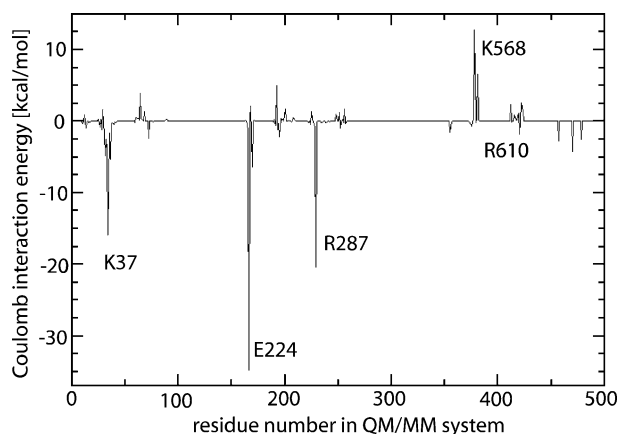


Fig. 5 Electrostatic interactions between catalytic core and protein environment. Shown are the changes in electrostatic interaction between the nucleotide and the protein environment upon going from the reactant to the transition state (adapted from [36])

E224, and R287 stabilize the transition state, whereas K568 leads to a slight de-stabilization. Somewhat surprisingly, even though R287 provides a large contribution, its immediate structural neighbor, R610, does not.

The second important feature to notice in Fig. 4 is the endothermicity of the reaction energy profile in the wild type system, implying that the catalytic site in the conformation studied has the ability to tightly bind ATP, but not yet to hydrolyze it. To investigate this in more detail and to identify means by which the protein could achieve a lowering of the product state energy, *in silico* mutation studies can be used [36]. Here, selected residues are mutated in the model system and the reaction energy profile is recalculated either fully or for selected points along the reaction pathway. In the present case, such an analysis shows, for example, that the mutations Q254N and Q254G exert a significant influence on the product state energy. This, in turn, allows one to hypothesize that motion of Q254 with respect to P_γ leads to the lowering of the product state energy to enable efficient ATP hydrolysis. Q254 links to a protein pocket known to be involved in ssDNA base flipping [19] during translocation via the inchworm mechanism. This suggests a direct connection between translocation along ssDNA and the actual catalytic step in the ATP binding pocket of PcrA via Q254.

Finally, based on the ATP hydrolysis reaction energy profile in PcrA shown in Fig. 4, the *in silico* mutation studies, and the findings from QM/MM simulations of ATP hydrolysis in F_1 -ATPase [10, 12], the overall ATP catalysis reaction energy profile can be assumed to have an equilibrium constant of $K \sim 1$, corresponding to an iso-energetic reaction energy profile [36]. This has important implications, since it means that the actual chemical bond-breaking step in the catalytic sites of PcrA does not release any net free energy and is

hence not coupled to force generation. Similar to F_1 -ATPase [7, 12, 39], it is likely that the binding of reactant ATP to the initially open catalytic cleft at the interface between the 1A and 2A domains leads to a closing motion which provides the actual power stroke propagating PcrA along ssDNA. This picture fits nicely with the above observation that insertion of R287 from domain 2A into the catalytic site located in the 1A domain is required for efficient ATP hydrolysis to take place. Such a coupling of large-scale protein conformation and the chemical reaction assures that ATP hydrolysis does not occur prematurely, which could destroy the directed motion of PcrA along ssDNA [36].

However, due to the static nature of the employed QM/MM method and the limited spatial extent of the QM/MM system, one is limited in one's ability to test the domain motion hypothesis and the coupling of Q254 to ssDNA base flipping on realistic time and length scales. This is where all-atom MD simulations and coarser-grained descriptions come into play. These methods will be the topic of the following sections.

4

Molecular Dynamics Simulations of PcrA Function

Continuing our study of molecular motor function in PcrA, we turn to MD simulations to investigate PcrA's larger scale conformational motions leading to translocation along ssDNA and to study the coupling of ATP hydrolysis to unidirectional translocation. Figure 6 shows a molecular view of an all-atom simulation system of PcrA, containing the protein, DNA, TIP3P water molecules, and a number of ions adding up to a total of more than 110 000 atoms. Before discussing some of the findings for PcrA function obtained via MD simulations we will give a brief exposition of MD simulation methods.

MD Simulations—A Brief Introduction

Classical simulations are based on the solution of the differential equation posed by Newton's second law for a system of N interacting particles

$$\ddot{\mathbf{x}}_i = \frac{1}{m_i} \mathbf{F}_i \quad (i = 1, \dots, N). \quad (11)$$

Here $\ddot{\mathbf{x}}_i$ is the second time derivative of the position of particle i , m_i its mass, and $\mathbf{F}_i$ the total force acting on it. $\mathbf{F}_i$ depends on the positions of all other particles. The calculation of the forces $\mathbf{F}_i$ is the step requiring the most computational effort.

A number of algorithms have been developed to numerically integrate Eq. 11 and, thereby, obtain the particles' trajectories $\mathbf{x}_i(t)$. Many integration schemes use the *Leap frog algorithm* for updating coordinates and velocities

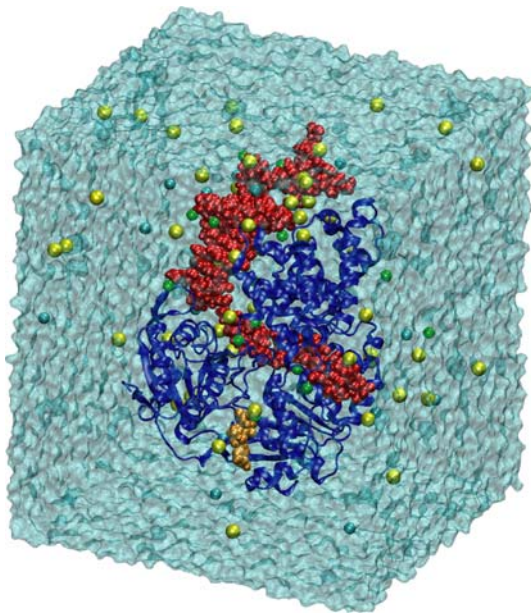


Fig. 6 PcrA all-atom simulation system. Shown is the protein in a *blue* cartoon representation with bound dsDNA and ssDNA shown in *red* vdW representation. The nucleotide ATP bound to the catalytic site at the interface between the 1A and 2A domains is shown as *gold*-colored vdW spheres. The protein-DNA complex is immersed in a large box of TIP3P water molecules containing Cl^- (*cyan*), Na^+ (*yellow*), and Mg^{2+} (*green*) ions (adapted from [51])

of particle i :

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t + \Delta t/2) \quad (12)$$

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i(t - \Delta t/2) + \Delta t \frac{1}{m_i} \mathbf{F}_i, \quad (13)$$

with Δt being the value of the discrete time-step. In biomolecular simulations, the value of Δt has to be a fraction of the fastest vibrational period in the system and, therefore, is in the 1–2 fs range.

All classical simulations are based on a classical approximation to the Hamilton operator in the Schrödinger Eq. 2. With this classical Hamiltonian one can then proceed to calculate energies, forces, and higher derivatives and use them to, for example, compute the minimum energy conformation or simulate the dynamical behavior via solving Eq. 11.

In the following we will focus on the classical Hamiltonians that are commonly used for simulations of biomolecular systems. They are referred to as *force fields* and are known under names such as CHARMM [40], GRO-MACS [41], and AMBER [42]. They all have in common a relatively simple mathematical structure to allow for a computationally efficient force evalu-

ation. Furthermore, these force fields are empirical, i.e., they contain fitting parameters, that are either determined by comparison to experiments, via quantum mechanical calculations, or both.

A typical force field is described by the following Hamiltonian

$$H_{\text{MM}} = H_{\text{bnd}} + H_{\text{ang}} + H_{\text{dih}} + H_{\text{vdW}} + H_{\text{ele}} . \quad (14)$$

Here, the first three terms represent the bonded contributions, namely the bond

$$H_{\text{bnd}} = \sum_a^{\text{bonds}} \frac{1}{2} k_a^{\text{bnd}} (x_a - x_{0a})^2 , \quad (15)$$

and angle terms

$$H_{\text{ang}} = \sum_a^{\text{angles}} \frac{1}{2} k_a^{\text{ang}} (\theta_a - \theta_{0a})^2 , \quad (16)$$

that are both simple harmonic functions of the bond lengths x_a and angles θ_a , respectively. The parameters x_{0a} and θ_{0a} denote the corresponding equilibrium values and the $k_a^{\text{bnd/ang}}$ are the interaction strengths.

The contributions of the dihedral angles ϕ_a are periodic, typically of small magnitude, and are given by

$$H_{\text{dih}} = \sum_a^{\text{dihed}} k_a^{\text{dih}} (1 + \cos(n\phi_a - \phi_0)) . \quad (17)$$

Again, ϕ_0 is the equilibrium value of the dihedral angle, k_a^{dih} the interaction strength, and n determines the periodicity. The last two terms in Eq. 14 are the non-bonded contributions to the force field. The short range part is represented by a Lennard-Jones potential

$$H_{\text{vdW}} = \sum_{a \neq b}^N 4\epsilon_{ab} \left[\left(\frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left(\frac{\sigma_{ab}}{r_{ab}} \right)^6 \right] , \quad (18)$$

where r_{ab} is the distance between particles a and b , and the quantities ϵ_{ab} and σ_{ab} correspond to the depth of the Lennard-Jones potential and to the inter-particle distance where the potential is equal to zero, respectively.

Finally, the last term represents the Coulomb interaction between pairs of atoms with partial charges q_a and q_b

$$H_{\text{ele}} = \sum_{a \neq b}^N \frac{q_a q_b}{r_{ab}} . \quad (19)$$

Because of the $1/r$ dependence, the Coulomb interaction is long-range and, therefore, a straightforward calculation requires a computational effort of

order $O(N^2)$ in the number of particles N . Computationally, it is, therefore, the most expensive part in evaluating Eq. 14 and several numerical approximations, for example the Particle Mesh Ewald method, have been devised to reduce the computational cost of this term to $O(N \log N)$.

Using efficient parallelization schemes, such as that employed by the molecular dynamics program NAMD [43], it is presently possible to calculate the time evolution of a biomolecular system using Newton's equation of motion combined with an expression for the force field similar to the one given in Eq. 14 for a million atoms over a time of tens of ns [44]. MD calculations were traditionally performed in the micro-canonical (NVE) ensemble; currently, most MD simulations are conducted in the canonical (NVT) and grand-canonical (NPT) ensemble, i.e., under conditions of constant volume/pressure. The temperature of the system can be kept constant by coupling the system to an external heat bath at the desired temperature [45]; the pressure can be controlled by using, for example, the Nose-Hoover Langevin piston method [46, 47]. A rather comprehensive introduction into concepts and algorithms underlying classical MD simulations can be found in [43].

Molecular Dynamics Simulations of PcrA

Having said this, it becomes apparent that the main difficulty in applying MD methods to the study of molecular motors is the vast gap between the time-scales available to equilibrium MD simulations (~ 100 ns) and the ones relevant physiologically (μ s to ms). Since the translocation of PcrA by a single nucleotide takes about 20 ms, the accompanying protein motions, among them, for example, the closing of the catalytic cleft between domains 1A and 2A or the motion of bound ssDNA bases, take place on a similar length time-scale. Hence, investigating PcrA function with equilibrium MD simulations is hampered by this gap in time-scale and different strategies are needed.

One possible approach to elucidate the coupling of ATP binding and hydrolysis to changes in PcrA conformation consists of the analysis of equilibrium MD trajectories of PcrA in the ATP and ADP+P_i bound states. Figure 7 shows the per residue RMSD averaged over several ns for both states, clearly illustrating that at least on a ns time scale, there is no significant difference between the two, rendering this approach futile.

The time-scale gap can be partially overcome by applying an external biasing potential, enforcing a certain change in conformation, via, for example steered molecular dynamics simulations (SMD) [48, 49] or umbrella sampling [50]. The application of these approaches to molecular motors, however, suffers from several limitations. First, it requires a certain insight into what the physiologically relevant motions are, in order to bias the simulations properly. Second, biasing can be difficult for all but the simplest motions. Finally, enforcing ms time-scale protein motions on a ns computational time-scale can lead to artifacts or, in the worst case, destroy the protein.

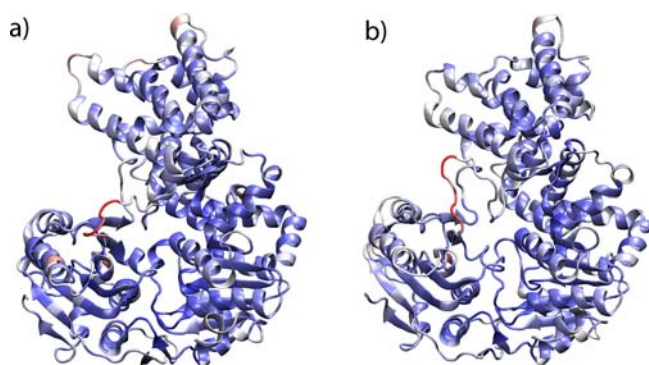


Fig. 7 Trajectory averaged per-residue RMSD in PcrA. Shown are the RMSD values based on ~ 3.5 ns of MD simulations for the ATP (**a**) and ADP/ P_i (**b**) bound PcrA systems. The color scale ranges from blue to red and indicates increasing RMSD values. The most mobile part of the structure is a flexible loop connecting the 2A and 2B domains colored in red

Nevertheless, despite all these drawbacks, significant insight into the system can be gained if the approach is applied properly and can be used, for example, to calculate potentials of mean force or force-extension curves.

In the case of PcrA, such a biasing method can be employed to investigate the mechanism behind the directionality of ssDNA motion in the $3' \rightarrow 5'$ direction. As proposed in the inchworm translocation model [19] derived from structural data, the 1A and 2A domains each have different affinities for ssDNA, and both alternate their affinities between the substrate (with ATP bound, *s*) and product (without ATP/ADP bound, *p*) state. The model, however, does not provide any microscopic evidence to support such a mechanism. To test this idea further, SMD simulations have been used by attaching ten harmonic springs (force constants of $2 \text{ kcal/mol } \text{\AA}^2$) to ten phosphorous atoms of the ssDNA bound to PcrA [51]. The ssDNA was then pulled one half nucleotide (nt) distance forward and backward across the protein-ssDNA interface. The direction of pulling in the *s* and *p* states was chosen according to the anticipated domain motion in each state corresponding to separation and closure of domains 1A and 2A, respectively. The required pulling forces were monitored and are reproduced in Fig. 8. These results show that in the *s* state, the average force needed to move nucleotides 15–18 to the right (corresponding to forward motion of 2A to the left) is smaller than that needed to move nucleotides 18–21 to the left (corresponding to backward motion of 1A to the right). Similarly, in the *p* state, the average force needed to move nucleotides 15–18 to the left (corresponding to backward motion of 2A to the right) is larger than the force to move nucleotides 17–20 to the right (corresponding to forward motion of 1A to the left). This directly supports the idea of alternating affinities between the protein domains and ssDNA in different PcrA states and can explain the directionality of PcrA translocation.

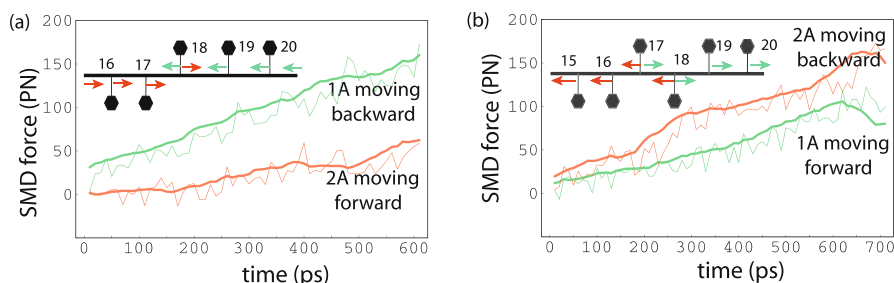


Fig. 8 Comparisons of SMD forces arising during ssDNA pulling simulations in PcrA. Shown are the results with ATP bound (*s* state) (**a**) and without ATP bound (*p* state) (**b**). The *green/red* curve represents the average force needed to move relevant nucleotides in a direction opposite to the corresponding movement of domain 1A/2A (indicated by *green/red* arrows). Numbers are used to label nucleotides on ssDNA. The *thin* curves were measured directly from simulations, while the *thick* curves were smoothed over every ten data points. The results show that in **a**, the average force needed to shift the relevant nucleotides, corresponding to the domain movement of 2A, is smaller than the average force needed to shift the relevant nucleotides corresponding to the domain movement of 1A; the opposite is true in **b** (adapted from [51])

In order to further investigate the microscopic mechanism underlying the alternating affinities between the 1A and 2A domains and the bound ssDNA, all-atom ns time-scale MD simulations were used to derive an effective potential, $U_{i\sigma}(x_i)$, governing individual domain motion of PcrA. Here, σ is a state index describing either *s* or *p*; $i = 1, 2$ refers to domain 1A and 2A, respectively, with x_i being the position of each domain along the bound ssDNA. The effective potential $U_{i\sigma}(x_i)$ is then used to formulate a coarse-grained Langevin dynamics description of PcrA that will be described in the next section. This approach [51] allows one to describe the ATP hydrolysis coupled uni-directional ssDNA translocation of PcrA on the physiologically relevant ms time-scale.

Since the translocation of helicase along ssDNA arises through the sequential binding and unbinding of ssDNA units (nucleotides) to the PcrA domain surfaces, one possible strategy for determining $U_{i\sigma}$ is to calculate the binding free energies of individual ssDNA units with the protein, and then use an appropriate combination of these energies to obtain $U_{i\sigma}$. The calculation of the binding free energy from MD simulations is not straightforward, but can be accomplished by a semi-empirical method [52]. Here, the binding energy E_b is approximated as a weighted sum of electrostatic and vdW interactions between the protein and each individual nucleotide as extracted from equilibrium MD simulations. In the present case, only relative but not absolute energies are needed. Hence, the solvent (water and ion) contribution to the energy can be neglected by assuming a homogeneous solvent-nucleotide interaction along the bound ssDNA region (consisting of poly-thymine in the

simulation model) [51]. From the discrete set of nucleotide-protein interaction energies derived from MD, a continuous site-energy function $E_b(x)$ for each state (s and p) was constructed by polynomial interpolation [51]. Figure 9 shows the continuous site energies $E_b(x)$ as well as the resulting potential energy functions $U_{i\sigma}$ in both the s and p states.

Now we give an example of how to derive $U_{i\sigma}$ from the binding energies $E_b(x)$ [51]. Starting from the s state, as domain 2A moves forward (to the left) by a distance Δx ($\Delta x \leq 0.5$, in unit of a single nt distance), nucleotides 15 to 18 move backward (to the right) by Δx , while the other nucleotides remain stationary. This scenario can be inferred from the available X-ray crystal structures. Hence, the potential U_{2s} , for example, can be expressed as

$$U_{2s}(\Delta x) = \sum_{j=15}^{18} [E_b(x_j + \Delta x) - E_b(x_j)] \quad \Delta x \in [0, 0.5]. \quad (20)$$

The other $U_{i\sigma}$ can be computed in a similar fashion. In the derivation, we assume a symmetric form of $U_{i\sigma}$ with a barrier at the center at $\Delta x = 0.5$, where $A_{i\sigma} = U_{i\sigma}(0.5)$ is defined as the barrier height. According to Eq. 20, $A_{i\sigma}$ in-

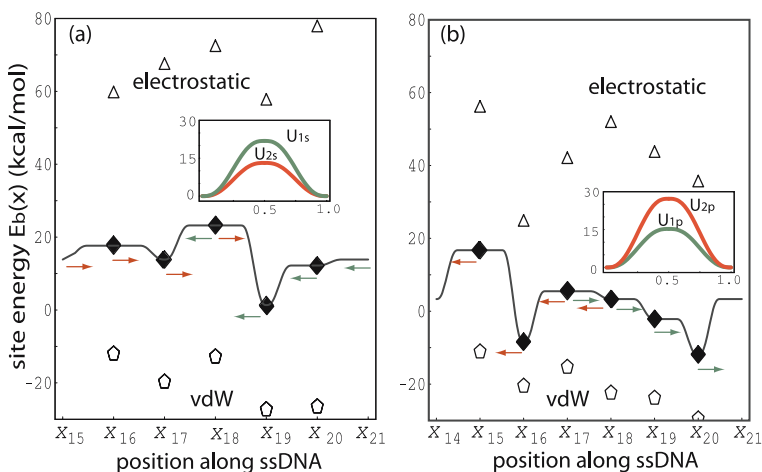


Fig. 9 Site energies of ssDNA nucleotides and individual domain potentials for PcrA. Shown are the result for the substrate (a) and the product state (b). Filled dark diamonds represent the relative binding free energies of nucleotide i at position x_i , i.e., the weighted sum of electrostatic and vdW energies between protein and individual nucleotides. The separate electrostatic and vdW contributions are indicated through non-filled triangles and non-filled pentagons, respectively. A smooth site energy function $E_b(x)$ is drawn through the filled dark diamonds using a third-order polynomial interpolation. The inset shows the potential $U_{i\sigma}(\Delta x)$ as derived from $E_b(x)$ that is experienced by domain 2A (red solid curve) and 1A (green solid curve) as the domains move along ssDNA; the length scale is in units of one nt distance (6.5 Å) (adapted from [51])

cludes a sum of site energy terms $E_b(x_j + 0.5) - E_b(x_j)$. Following [51] one can write this as

$$A_{i\sigma} = \Sigma_{i\sigma} + \Delta, \quad (21)$$

where $\Sigma_{i\sigma}$ are sums of the differences of binding energies $E_b(x_j)$. Δ is a single tunable parameter that accounts for the effect of additional friction due to ssDNA-protein interactions and is assumed to be independent of i and σ . The value of Δ is adjusted so that the translocation speed of PcrA is consistent with that measured from experiment, i.e., about one nt distance every 20 ms.

It is important to point out that the difference between amplitudes, $|A_{1\sigma} - A_{2\sigma}|$, is independent of Δ . The calculations show that $A_{1s} - A_{2s} \sim 9$ kcal/mol in the s state and $A_{1p} - A_{2p} \sim -12$ kcal/mol in the p state [51]. This result, i.e., $A_{1s} > A_{2s}$ and $A_{1p} < A_{2p}$, derived solely on the basis of ns time-scale MD simulations, confirms the idea of alternating affinities between the domains and ssDNA in different states, first proposed by Velankar et al. [19]: in the ATP-bound (s) state domain 2A moves more easily than 1A while in the empty binding site case (state p) the opposite is true. As a result, when ATP binds and the domains 1A and 2A are being pulled together, only 1A moves. When the domains separate again, domain 2A moves. This scenario of alternating 1A and 2A mobilities along DNA results in a unidirectional translocation as one can readily realize by grasping with ones hands (representing domains 1A and 2A) a rope (representing ssDNA) and alternately making a tight and a loose fist with the left and right hand while pulling the rope with the tight fist (c.f. Fig. 6b in [19]). The derived potential energy functions $U_{i\sigma}$ (Fig. 9) will be utilized in a Langevin dynamics description presented below, demonstrating indeed uni-directional translocation along ssDNA.

5 Langevin Simulations of PcrA Function

From experimental studies it is known that PcrA translocates along ssDNA uni-directionally at a speed of about 50 nt/s, presumably consuming one ATP per single nucleotide step which, therefore, takes about 20 ms [53]. To describe PcrA function on this time-scale, a computational modeling approach different from QM/MM or MD is required for two major reasons. First, due to the explicit treatment of electronic and/or atomistic details, the time-scales accessible to QM/MM and MD simulations do not reach the long time-scales needed to model physiological PcrA function. Second, on larger length- and longer time-scales, the essential degrees of freedom governing the dynamics of the system are likely quite different and, therefore, largely inaccessible from simulations of smaller scale models. Therefore, in order to properly investigate the relevant mechanism on this functional level one needs to adopt a different approach.

One possible description of the PcrA ssDNA stepping motor replaces the two translocating domains 1A and 2A by two featureless beads [51], which move along ssDNA during each ATP hydrolysis cycle (Fig. 1). This simple view is based on the two basic conformations of the PcrA-DNA complex found in the available X-ray crystal structures [19].

A straightforward approach to characterize the bead movement along ssDNA is by means of a Langevin equation describing Brownian motion in a potential. Let us first suppose that the interaction between the bead and ssDNA is purely associative, i.e., the attachment of the bead to the DNA strand does not interfere with its movement along ssDNA. Under these conditions, the bead motion can be described by a 1-D Langevin equation,

$$m\ddot{x} = -\gamma\dot{x} + \tilde{f}(t). \quad (22)$$

Here γ is the friction coefficient. The term $\tilde{f}(t)$ represents white noise and obeys the fluctuation-dissipation theorem [54]

$$\langle \tilde{f}(t) \rangle = 0; \quad \langle \tilde{f}(t)\tilde{f}(t') \rangle = 2\gamma k_B T \delta(t - t'). \quad (23)$$

Since a particle the size of a protein in solution has a very low mass to viscosity ratio (the Reynolds number) one can work in the strong friction limit in which the acceleration term $m\ddot{x}$ in Eq. 22 vanishes.

Next, we take into account the interactions between the protein and ssDNA which prevent the beads from sliding freely along ssDNA. The interactions can be modeled by additional potentials in Eq. 22 which govern the movement of individual beads. These additional potentials are just the $U_{i\sigma}$ in Eq. 20 that were derived from MD simulation and originate from the imbalance in binding energy of individual ssDNA nucleotides to PcrA [51]. Since the two beads, corresponding to domain 1A and 2A, move in a concerted fashion, it may be advantageous to introduce an extra interaction term acting between the two domains facilitating the movement of domains. This interaction potential, when non-vanishing, is modeled by a harmonic spring

$$V_{\sigma'} = \frac{1}{2}k(x_1 - x_2 - l_{\sigma'})^2. \quad (24)$$

Here, σ' (s or p) is a state index varying independently from the σ in $U_{i\sigma}$. The force constant of the spring, k , adopts a value of $1k_B T/\text{\AA}^2$, a value empirically estimated from MD simulations [51]. $l_{\sigma'}$ is the equilibrium length of the spring, with $l_p > l_s$, since the two domains 1A and 2A are separated farther in the p state than in the s state. For the sake of simplicity, we use $l_p = l_0$ and $l_s = l_0/3$ with l_0 corresponding to the nucleotide separation in ssDNA ($\sim 6.5 \text{\AA}$).

Putting all of this together, the Langevin equation in the strong friction limit reads

$$\gamma\dot{x}_i = -\frac{\partial W_{\sigma\sigma'}(x_1, x_2)}{\partial x_i} + \tilde{f}(t), \quad (25)$$

where $W_{\sigma\sigma'}(x_1, x_2) = U_{1\sigma}(x_1) + U_{2\sigma}(x_2) + V_{\sigma'}(x_1, x_2)$, with $x_i, i = 1, 2$, being the position of the beads tracking domains 1A and 2A along ssDNA. Equation 25 can be solved numerically assuming discrete time steps of size Δt ,

$$x_i(t + \Delta t) = x_i(t) - \frac{1}{\gamma} \frac{\partial W_{\sigma\sigma'}(x_1, x_2)}{\partial x_i} \Delta t + \sqrt{2D \Delta t} Z. \quad (26)$$

Here, Z is a standard normal random variable (with mean 0 and variance 1) and D is the diffusion coefficient, according to the fluctuation-dissipation theorem related to the friction coefficient through $D = k_B T / \gamma$. D is chosen to assume a value of $10^4 \text{ \AA}^2 / \mu\text{s}$, typical for a 3 nm-radius protein in solution [55]. It is important to keep in mind, however, that D can vary significantly in different physiological environments.

Note that Eq. 25 holds in a particular state σ , denoted by s or p when $V_{\sigma'}$ is absent. Transitions between states are either triggered once some geometrical criterion is satisfied, or are simulated through a Poisson process with a specific rate constant. In the latter scenario one generates uniformly distributed random numbers Y (with $Y \in [0, 1]$) and transitions are accepted if $Y \leq \omega \Delta t$ ($\omega \Delta t \ll 1$), where ω is the rate constant for the transition and Δt is the discrete time step in Eq. 26. The rate constants are estimated [51] based on the rate limiting steps assumed in each scenario and the translocation speed observed experimentally [53].

In the following, we will assume two limiting scenarios [51], the weak and strong coupling scenario, expecting that the most realistic model will lie somewhere between the two. We envision that the translocation of PcrA along ssDNA comes about through an inchworm motion involving separate, but coupled translocations of its 1A and 2A domains (described by beads) that are governed by three factors: (1) There exist geometrical constraints that prohibit the domains to pass each other as well as to separate too far. (2) Binding of ATP favors a narrower separation between domains 1A and 2A while unbinding of ADP and P_i favors a wider separation between the domains. (3) Depending on the state (s or p) of PcrA, the domains experience different effective potentials such that in the ATP bound state (s), 2A can move easily (low energy barrier A_{2s}) and 1A can hardly move (high energy barrier A_{1s}) and in the absence of ATP/ADP (p), 1A can move easily (low energy barrier A_{1p}) and 2A can hardly move (high energy barrier A_{2p}). Below we investigate PcrA translocation in the two scenarios, the weak and strong coupling scenario, and demonstrate how (1)–(3) can endow PcrA with unidirectional motion in the $3' \rightarrow 5'$ direction.

Weak Coupling Scenario

Here, domain 2A and 1A move with vanishing mutual interaction potential, i.e. $V_{\sigma'}(x_1, x_2) = 0$. However, the two domains are still coupled by geometrical constraints. We furthermore assume that the rate limiting steps are indi-

vidual domain movements, i.e., thermally activated barrier-crossing events governed by potentials $U_{i\sigma}$. The barrier-crossing time can be estimated by the mean first-passage time [55, 56], which in turn is mostly determined by $A_{i\sigma}$ given in Eq. 21. In order to obtain a translocation speed of about $6.5 \text{ \AA}/20 \text{ ms}$ a value for the parameter Δ of -2.5 kcal/mol was used. The corresponding results of Langevin dynamics simulations are shown in Fig. 10a. The left panel depicts the stochastic trajectories describing the motion of 1A and 2A along ssDNA over a period of 10 hydrolysis cycles. The right-side panel shows

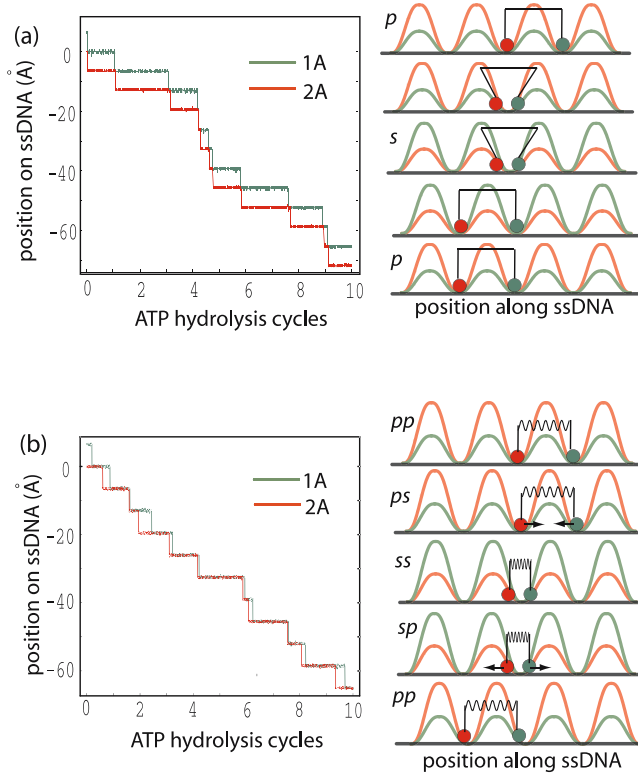


Fig. 10 Langevin simulations of ssDNA translocation in PcrA. Depicted are the trajectories for the weak (a) and the strong coupling scenario (b). *Left* Shown are trajectories of the two translocating domains, 1A (green) and 2A (red), moving along ssDNA in each scenario; the time is given in units of ATP hydrolysis cycles (one cycle lasts about 20 ms). *Right* Illustrated are the individual potentials $U_{i\sigma}$ experienced by domain 1A (green) and domain 2A (red) moving along ssDNA in different states (p , s in the weak coupling scenario or pp , ps , ss , and sp in the strong coupling scenario, defined in [51]) in each scenario. In the weak coupling scenario, two domains are shown as being connected by a rod, corresponding to the geometric constraint; in the strong coupling scenario (b), the domains are shown as being connected by an elastic spring with variable equilibrium lengths, corresponding to the non-vanishing interaction potential $V_{\sigma'}$ (adapted from [51])

schematically the sequence of configurations corresponding to the motion. In the p state, the two domains separate first; domain 1A experiences a low barrier (A_{1p}) and can move readily, while 2A experiences a high barrier (A_{2p}) and is essentially “stuck”. When 1A has moved forward close to 2A, ATP binds into the cleft between 1A and 2A and the system transits to the s state. In the s state, the potentials U_{is} differ qualitatively from U_{ip} , in that now 2A is ready to move (low barrier A_{2s}) and 1A becomes “stuck” (high barrier A_{1s}). When 2A moves forward far enough, ADP and P_i quickly dissociate and the system transits back to the p state. Overall, after one cycle, PcrA has shifted one nt distance compared to the starting state on the ssDNA.

Strong Coupling Scenario

In this scenario, domain 1A and 2A move under the influence of a non-vanishing interaction potential $V_{\sigma'}(x_1, x_2)$, modeled by a harmonic spring. Hence, the barrier-crossing events of each domain happen under the combined potential of $U_{i\sigma}$ and $V_{\sigma'}$. The states in the strong coupling scenario, accordingly, can be defined through two labels, σ and σ' . This leads to four possible states, namely, $(\sigma\sigma') = (ss), (sp), (ps),$ and (pp) as shown in Fig. 10b. Since $A_{1p} < A_{2p}$ as well as $A_{1s} > A_{2s}$ still holds, the barrier experienced by 1A/2A is lower than that experienced by 2A/1A during ATP binding/ADP + P_i dissociation. Therefore, the unidirectional translocation of PcrA is maintained, as demonstrated in Fig. 10b. The domain movement itself, however, can happen fast in this scenario and we assume that the rate limiting steps are waiting for ATP arrival and ATP hydrolysis; the rate constants of these events can also be fitted to yield a translocation speed of about $6.5 \text{ \AA}/20 \text{ ms}$.

From our analysis, it is not clear which of the two limiting scenarios is the physiologically relevant one for PcrA. Experimental evidence suggests that PcrA may actually work in a mixed scenario, in which ATP binding serves as a power-stroke [57, 58], facilitating the forward movement of domain 1A and assisting PcrA conformational change so that 1A can not move backward, while ATP hydrolysis or ADP + P_i dissociation triggers the thermally activated forward movement of domain 2A (Myong et al., 2006, personal communication), assisting PcrA conformational change to prevent 2A from moving backward.

No matter what the exact scenario is, one can recognize that the unidirectional translocation of PcrA stems from the fact that in the state without ATP bound, domain 1A experiences a lower energy barrier than 2A; while in the ATP-bound state, domain 2A experiences a lower energy barrier than 1A, that is $A_{1p} < A_{2p}$ and $A_{1s} > A_{2s}$. One can imagine that the opposite relationship $A_{1p} > A_{2p}$ and $A_{1s} < A_{2s}$ will lead to helicase translocation in the reversed direction.

Since the above Langevin dynamics simulations of PcrA translocation were conducted with the potentials $U_{i\sigma}$ derived from atomic-scale simulations, and

recognizing that it is the difference in barrier height $|A_{1\sigma} - A_{2\sigma}|$ that directs the unidirectional movement, it is interesting to evaluate the contributions of individual residues to this difference term. This reveals that two key amino acids, R260 and K385, contribute prominently to $A_{1s} - A_{2s}$ and $A_{2p} - A_{1p}$, respectively [51], and are therefore proposed to be most important in affecting the unidirectional translocation of PcrA.

6

Summary

Combining QM/MM calculation, nanosecond MD simulation, and millisecond stochastic modeling, the ATP hydrolysis coupled ssDNA translocation of a monomeric PcrA helicase has been investigated.

Figure 11 summarizes the combined results and depicts a single PcrA translocation cycle involving five steps. In configuration 1, domains 1A and 2A are separated without ATP or ADP bound. The arrangement of the bases of the ssDNA are shown in the insert to the left, revealing a single base intercalated between side groups of Y257 and F64. Initially, domain 1A can move more easily along ssDNA than domain 2A. At this point ATP arrives at the binding site and initiates binding. The binding of ATP causes an attractive force between the domains which consequently start approaching each other through motion of domain 1A, leading to configuration 3. The close approach of domains 1A and 2A pushes Y257 and F64 together and, thereby, squeezes the ssDNA base out, it moving from its former position to a new position as indicated by the arrows in the inserted figures. The close approach of 2A and 1A brings R610 and R287 into contact with the γ -phosphate of ATP; the movement of Y257 is transmitted to Q254. QM/MM calculations have shown that the three mentioned side groups, R610, R287, and Q254 control the hydrolysis of ATP in PcrA [36]. In configuration 3 the juxtaposition of the three side groups is optimal for hydrolysis of ATP to ADP + P_i , leading to configuration 4. At this point the mobility of domains 2A and 1A has reversed, 2A now being able to move more easily. Thermal fluctuations lead to the separation of 2A from 1A giving rise to configuration 5, and permitting release of ADP and P_i , reaching configuration 1'. PcrA has advanced by one step and is ready for the next cycle.

In summary, we have presented a comprehensive study of PcrA helicase function covering length scales ranging from the electronic domain using QM/MM simulations to the molecular level via MD and long time stochastic simulations. The time scales covered span several orders of magnitude from fast femtosecond time scale bond-breaking events to physiological protein motions on the millisecond time scale. The combination of methods employed in our computational modeling revealed a detailed picture of PcrA function, for example, a proton relay mechanism responsible for efficient ATP

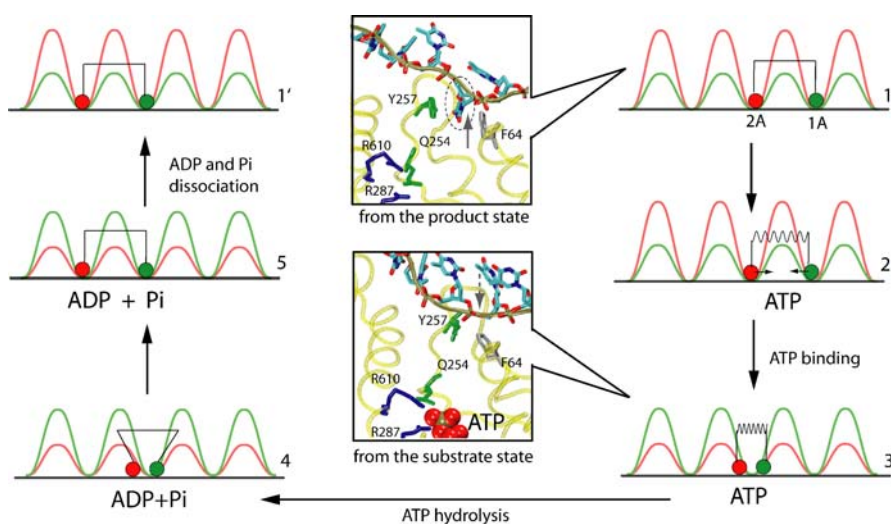


Fig. 11 Five-step PcrA translocation cycle. The figure shows schematically a translocation cycle for a mixed (see text) scenario involving five configurations 1, 2, 3, 4, 5 as well as configuration 1' that is equivalent to configuration 1, except that the system has advanced by one base. The mixed scenario involves both a loaded spring (non-zero potential $V_{\sigma}(x_1, x_2)$, configuration 2 $\rightarrow$ 3) and a step with a random thermal motion (vanishing potential $V_{\sigma}(x_1, x_2)$, configuration 4 $\rightarrow$ 5). In configurations 1, 2, 1' domain 1A (green) moves more readily than domain 2A (red), while it is the opposite for configurations 3, 4, 5. The *insert figures* show how the domain and ssDNA base motions are coupled to the chemistry at the ATP binding site: upon the approach of domains 2A and 1A, R287 and R610 move close to the γ -phosphate of ATP; Q254 is linked closely to Y257 that forms a key binding pocket for an ssDNA base, which is squeezed out when the domains approach each other in binding ATP; Q254 was identified as a key participant in ATP hydrolysis along with the mentioned arginines [36]. The suggested mechanism involves the binding of ATP that pulls domain 1A towards domain 2A; insertion of R287 and R610 into an optimal (for hydrolysis) position in the ATP binding pocket along with Q254 linked to a key ssDNA interaction site; rapid hydrolysis of ATP that initiates separation of domains 2A and 1A through movement of 2A alone (adapted from [51])

hydrolysis in the catalytic site and the means by which uni-directional motion along ssDNA is achieved, namely by a ratchet mechanism. Employing a similar strategy to investigate other molecular motors can be expected to reveal more about the mechanism by which these remarkable nano-machines perform their physiological function.

Acknowledgements The molecular images in this paper were created with the molecular graphics program VMD [59]. This work is supported by grants from the National Institutes of Health PHS-5-P41-RR05969 and the National Science Foundation MCB02-34938. The authors gladly acknowledge supercomputer time provided by Pittsburgh Supercomputer Center and the National Center for Supercomputing Applications via National Resources Allocation Committee grant MCA93S028.

References

1. Li G, Cui Q (2004) *J Phys Chem B* 108:3342–3357
2. Minehard TJ, Marzari N, Cooke R, Pate E, Kollman PA, Car R (2002) *Biophys J* 82:660–675
3. Zheng W, Brooks BR (2005) *Biophys J* 89:167–178
4. Fischer S, Windshügel B, Horak D, Holmes KC, Smith JC (2005) *Proc Natl Acad Sci USA* 102:6873–6878
5. Kawakubo T, Okada O, Minami T (2005) *Biophys Chem*, 115:77–85
6. Wrighers W, Schulten K (1999) *Proteins: Struct Func Gen* 35:262–273
7. Oster G, Wang H (2000) *Biochim Biophys Acta* 1458:482–510
8. Böckmann RA, Grubmüller H (2002) *Nature Struct Biol* 9:198–202
9. Ma J, Flynn TC, Cui Q, Leslie AGW, Walker JE, Karplus M (2002) *Structure* 10:921–931
10. Dittrich M, Hayashi S, Schulten K (2003) *Biophys J* 85:2253–2266
11. Strajbl M, Shurki A, Warshel A (2003) *Proc Natl Acad Sci USA* 100:14834–14839
12. Dittrich M, Hayashi S, Schulten K (2004) *Biophys J* 87:2954–2967
13. Cox K, Watson T, Soultanas P, Hirst J (2003) *Proteins: Struct Func Gen* 52:254–262
14. Betterton M, Julicher F (2003) *Phys Rev Lett* 91:258103
15. Betterton M, Julicher F (2005) *Phys Rev E* 71:011904
16. Bhattacharjee S, Seno F (2003) *J Phys A: Math Gen* 36:L181–L187
17. Bhattacharjee S (2004) *Europhys Lett* 65:574–580
18. Subramanya H, Bird L, Brannigan J, Wigley D (1996) *Nature* 384:379–383
19. Velankar S, Soultanas P, Dillingham M, Subramanya H, Wigley D (1999) *Cell* 97:75–84
20. Wong I, Lohman TM (1992) *Science* 256:250–256
21. Stasiak A, Tsaneva IR, West SC, Benson CJ, Yu X, Egelman EH (1994) *Proc Natl Acad Sci USA* 91:7618–7622
22. Warshel A (1976) *Nature* 260:679–683
23. Singh UC, Kollman PA (1986) *J Comp Chem* 7:718–730
24. Field MJ, Bash PA, Karplus M (1990) *J Comp Chem* 11:700–733
25. Stanley RJ, Boxer SG (1995) *J Phys Chem* 99:859–863
26. Maseras F, Morokuma K (1995) *J Comp Chem* 16:1170–1179
27. Bakowies D, Thiel W (1996) *J Phys Chem* 100:10580–10594
28. Lyne PD, Hodoscek M, Karplus M (1999) *J Phys Chem A* 103:3462–3471
29. Lyne D, Walsh O (2001) *Computer Simulation of Biochemical Reactions with QM/MM Methods*. In: Becker O, Mac Kerell A Jr, Roux B, Watanabe M (eds) *Computational Biochemistry and Biophysics*. Marcel Dekker Inc., New York
30. Warshel A, Levitt M (1976) *J Mol Biol* 103:227–249
31. Théry V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) *J Comp Chem* 15:269–282
32. Bayly C, Cieplak P, Cornell W, Kollman P (1993) *J Phys Chem* 100:10269–10280
33. Hayashi S, Ohmine I (2000) *J Phys Chem B* 104:10678–10691
34. Warshel A, Weiss RM (1980) *J Am Chem Soc* 102:6218
35. Warshel A (2003) *Ann Rev Biophys Biomol Struct* 32:425–443
36. Dittrich M, Schulten K (2006) *Structure* 14:1345–1353
37. Scheidig AJ, Burmester C, Goody RS (1999) *Structure* 7:1311–1324
38. Ahmad Z, Senior AE (2004) *J Biol Chem* 44:46057–46064
39. Dittrich M, Schulten K (2005) *J Bioener Biomemb* 37:441–444
40. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J Comp Chem* 4:187–217

41. Lindahl E, Hess B, van der Spoel D (2001) *J Mol Mod* 7:306–317
42. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham T, Debolt S, Ferguson D, Seibel G, Kollman P (1995) *Comput Phys Commun* 91:1–41
43. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) *J Comp Chem* 26:1781–1802
44. Freddolino PL, Arkhipov AS, Larson SB, Mc Pherson A, Schulten K (2006) *Structure* 14:437–449
45. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) *J Chem Phys* 81:3684–3690
46. Martyna GJ, Tobias DJ, Klein ML (1994) *J Chem Phys* 101:4177–4189
47. Feller SE, Zhang YH, Pastor RW, Brooks BR (1995) *J Chem Phys* 103:4613–4621
48. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wriggers W, Schulten K (1998) *Steered Molecular Dynamics*. In: Deuffhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, Skeel RD (eds) *Computational Molecular Dynamics: Challenges, Methods, Ideas*. Vol 4. Springer, Berlin Heidelberg New York
49. Isralewitz B, Gao M, Schulten K (2001) *Curr Op Struct Biol* 11:224–230
50. Torrie, Valleau (1977) *J Comp Phys* 23:187–199
51. Yu J, Ha T, Schulten K (2006) *Biophys J* 91:2097–2114
52. Åqvist J (1996) *J Comp Chem* 17:1587–1597
53. Dillingham M, Wigley D, Webb M (2000) *Biochemistry* 39:205–212
54. Kubo R (1966) *Rep Progr Phys* 29:255–284
55. Howard J (2001) *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates, Inc., Sunderland, MA
56. Szabo A, Schulten K, Schulten Z (1980) *J Chem Phys* 72:4350–4357
57. Nishizaka T, Oiwa K, Noji H, Kimura S, Muneyuki E, Yoshida M, Kinosita K Jr (2004) *Nat Struct Mol Biol* 11:142–148
58. Shimabukuro K, Yasuda R, Muneyuki E, Hara KY, Kinosita K Jr, Yoshida M (2003) *Proc Natl Acad Sci USA* 100:14731–14736
59. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graphics* 14:33–38

Author Index Volumes 251–271

Author Index Vols. 26–50 see Vol. 50
Author Index Vols. 51–100 see Vol. 100
Author Index Vols. 101–150 see Vol. 150
Author Index Vols. 151–200 see Vol. 200
Author Index Vols. 201–250 see Vol. 250

The volume numbers are printed in italics

- Ajayaghosh A, George SJ, Schenning APHJ (2005) Hydrogen-Bonded Assemblies of Dyes and Extended π -Conjugated Systems. 258: 83–118
- Albert M, Fensterbank L, Lacôte E, Malacria M (2006) Tandem Radical Reactions. 264: 1–62
- Alberto R (2005) New Organometallic Technetium Complexes for Radiopharmaceutical Imaging. 252: 1–44
- Alegret S, see Pividori MI (2005) 260: 1–36
- Amabilino DB, Veciana J (2006) Supramolecular Chiral Functional Materials. 265: 253–302
- Anderson CJ, see Li WP (2005) 252: 179–192
- Anslyn EV, see Houk RJT (2005) 255: 199–229
- Appukkuttan P, Van der Eycken E (2006) Microwave-Assisted Natural Product Chemistry. 266: 1–47
- Araki K, Yoshikawa I (2005) Nucleobase-Containing Gelators. 256: 133–165
- Armitage BA (2005) Cyanine Dye–DNA Interactions: Intercalation, Groove Binding and Aggregation. 253: 55–76
- Arya DP (2005) Aminoglycoside–Nucleic Acid Interactions: The Case for Neomycin. 253: 149–178
- Bailly C, see Dias N (2005) 253: 89–108
- Balaban TS, Tamiaki H, Holzwarth AR (2005) Chlorins Programmed for Self-Assembly. 258: 1–38
- Balzani V, Credi A, Ferrer B, Silvi S, Venturi M (2005) Artificial Molecular Motors and Machines: Design Principles and Prototype Systems. 262: 1–27
- Barbieri CM, see Pilch DS (2005) 253: 179–204
- Barchuk A, see Daasbjerg K (2006) 263: 39–70
- Bayly SR, see Beer PD (2005) 255: 125–162
- Beer PD, Bayly SR (2005) Anion Sensing by Metal-Based Receptors. 255: 125–162
- Bertini L, Bruschi M, de Gioia L, Fantucci P, Greco C, Zampella G (2007) Quantum Chemical Investigations of Reaction Paths of Metalloenzymes and Biomimetic Models – The Hydrogenase Example. 268: 1–46
- Bier FF, see Heise C (2005) 261: 1–25
- Blum LJ, see Marquette CA (2005) 261: 113–129
- Boiteau L, see Pascal R (2005) 259: 69–122
- Bolhuis PG, see Dellago C (2007) 268: 291–317
- Borovkov VV, Inoue Y (2006) Supramolecular Chirogenesis in Host–Guest Systems Containing Porphyrinoids. 265: 89–146

- Boschi A, Duatti A, Uccelli L (2005) Development of Technetium-99m and Rhenium-188 Radiopharmaceuticals Containing a Terminal Metal–Nitrido Multiple Bond for Diagnosis and Therapy. 252: 85–115
- Braga D, D’Addario D, Giaffreda SL, Maini L, Polito M, Grepioni F (2005) Intra-Solid and Inter-Solid Reactions of Molecular Crystals: a Green Route to Crystal Engineering. 254: 71–94
- Brebion F, see Crich D (2006) 263: 1–38
- Brizard A, Oda R, Huc I (2005) Chirality Effects in Self-assembled Fibrillar Networks. 256: 167–218
- Bruce IJ, see del Campo A (2005) 260: 77–111
- Bruschi M, see Bertini L (2007) 268: 1–46
- del Campo A, Bruce IJ (2005) Substrate Patterning and Activation Strategies for DNA Chip Fabrication. 260: 77–111
- Carney CK, Harry SR, Sewell SL, Wright DW (2007) Detoxification Biomaterials. 270: 155–185
- Chaires JB (2005) Structural Selectivity of Drug–Nucleic Acid Interactions Probed by Competition Dialysis. 253: 33–53
- Chiorboli C, Indelli MT, Scandola F (2005) Photoinduced Electron/Energy Transfer Across Molecular Bridges in Binuclear Metal Complexes. 257: 63–102
- Cölfen H (2007) Bio-inspired Mineralization Using Hydrophilic Polymers. 271: 1–77
- Collin J-P, Heitz V, Sauvage J-P (2005) Transition-Metal-Complexed Catenanes and Rotaxanes in Motion: Towards Molecular Machines. 262: 29–62
- Collyer SD, see Davis F (2005) 255: 97–124
- Commeyras A, see Pascal R (2005) 259: 69–122
- Coquerel G (2007) Preferential Crystallization. 269
- Correia JDG, see Santos I (2005) 252: 45–84
- Costanzo G, see Saladino R (2005) 259: 29–68
- Credi A, see Balzani V (2005) 262: 1–27
- Crestini C, see Saladino R (2005) 259: 29–68
- Crich D, Brebion F, Suk D-H (2006) Generation of Alkene Radical Cations by Heterolysis of β -Substituted Radicals: Mechanism, Stereochemistry, and Applications in Synthesis. 263: 1–38
- Cuerva JM, Justicia J, Oller-López JL, Oltra JE (2006) Cp_2TiCl in Natural Product Synthesis. 264: 63–92
- Daasbjerg K, Svith H, Grimme S, Gerenkamp M, Mück-Lichtenfeld C, Gansäuer A, Barchuk A (2006) The Mechanism of Epoxide Opening through Electron Transfer: Experiment and Theory in Concert. 263: 39–70
- D’Addario D, see Braga D (2005) 254: 71–94
- Danishesky SJ, see Warren JD (2007) 267
- Darmency V, Renaud P (2006) Tin-Free Radical Reactions Mediated by Organoboron Compounds. 263: 71–106
- Davis F, Collyer SD, Higson SPJ (2005) The Construction and Operation of Anion Sensors: Current Status and Future Perspectives. 255: 97–124
- Deamer DW, Dworkin JP (2005) Chemistry and Physics of Primitive Membranes. 259: 1–27
- Dellago C, Bolhuis PG (2007) Transition Path Sampling Simulations of Biological Systems. 268: 291–317
- Deng J-Y, see Zhang X-E (2005) 261: 169–190
- Dervan PB, Poulin-Kerstien AT, Fechter EJ, Edelson BS (2005) Regulation of Gene Expression by Synthetic DNA-Binding Ligands. 253: 1–31

- Dias N, Vezin H, Lansiaux A, Bailly C (2005) Topoisomerase Inhibitors of Marine Origin and Their Potential Use as Anticancer Agents. 253: 89–108
- DiMauro E, see Saladino R (2005) 259: 29–68
- Dittrich M, Yu J, Schulten K (2007) PcrA Helicase, a Molecular Motor Studied from the Electronic to the Functional Level. 268: 319–347
- Dobrawa R, see You C-C (2005) 258: 39–82
- Du Q, Larsson O, Swerdlow H, Liang Z (2005) DNA Immobilization: Silanized Nucleic Acids and Nanoprinting. 261: 45–61
- Duatti A, see Boschi A (2005) 252: 85–115
- Dworkin JP, see Deamer DW (2005) 259: 1–27
- Edelson BS, see Dervan PB (2005) 253: 1–31
- Edwards DS, see Liu S (2005) 252: 193–216
- Ernst K-H (2006) Supramolecular Surface Chirality. 265: 209–252
- Ersmark K, see Wannberg J (2006) 266: 167–197
- Escudé C, Sun J-S (2005) DNA Major Groove Binders: Triple Helix-Forming Oligonucleotides, Triple Helix-Specific DNA Ligands and Cleaving Agents. 253: 109–148
- Van der Eycken E, see Appukkuttan P (2006) 266: 1–47
- Fages F, Vögtle F, Žinić M (2005) Systematic Design of Amide- and Urea-Type Gelators with Tailored Properties. 256: 77–131
- Fages F, see Žinić M (2005) 256: 39–76
- Faigl F, Schindler J, Fogassy E (2007) Advantages of Structural Similarities of the Reactants in Optical Resolution Processes. 269
- Fantucci P, see Bertini L (2007) 268: 1–46
- Fechter EJ, see Dervan PB (2005) 253: 1–31
- Fensterbank L, see Albert M (2006) 264: 1–62
- Fernández JM, see Moonen NNP (2005) 262: 99–132
- Fernando C, see Szathmáry E (2005) 259: 167–211
- Ferrer B, see Balzani V (2005) 262: 1–27
- De Feyter S, De Schryver F (2005) Two-Dimensional Dye Assemblies on Surfaces Studied by Scanning Tunneling Microscopy. 258: 205–255
- Flood AH, see Moonen NNP (2005) 262: 99–132
- Fogassy E, see Faigl F (2007) 269
- Fricke M, Volkmer D (2007) Crystallization of Calcium Carbonate Beneath Insoluble Monolayers: Suitable Models of Mineral–Matrix Interactions in Biomineralization? 270: 1–41
- Fujimoto D, see Tamura R (2007) 269
- Fujiwara S-i, Kambe N (2005) Thio-, Seleno-, and Telluro-Carboxylic Acid Esters. 251: 87–140
- Gansäuer A, see Daasbjerg K (2006) 263: 39–70
- Garcia-Garibay MA, see Karlen SD (2005) 262: 179–227
- Gelinck GH, see Grozema FC (2005) 257: 135–164
- Geng X, see Warren JD (2007) 267
- George SJ, see Ajayaghosh A (2005) 258: 83–118
- Gerenkamp M, see Daasbjerg K (2006) 263: 39–70
- Giaffreda SL, see Braga D (2005) 254: 71–94
- de Gioia L, see Bertini L (2007) 268: 1–46
- Greco C, see Bertini L (2007) 268: 1–46
- Grepioni F, see Braga D (2005) 254: 71–94

- Grimme S, see Daasbjerg K (2006) 263: 39–70
- Grozema FC, Siebbeles LDA, Gelinck GH, Warman JM (2005) The Opto-Electronic Properties of Isolated Phenylenevinylene Molecular Wires. 257: 135–164
- Guisseppi-Elie A, Lingerfelt L (2005) Impedimetric Detection of DNA Hybridization: Towards Near-Patient DNA Diagnostics. 260: 161–186
- Di Giusto DA, King GC (2005) Special-Purpose Modifications and Immobilized Functional Nucleic Acids for Biomolecular Interactions. 261: 131–168
- Haase C, Seitz O (2007) Chemical Synthesis of Glycopeptides. 267
- Hansen SG, Skrydstrup T (2006) Modification of Amino Acids, Peptides, and Carbohydrates through Radical Chemistry. 264: 135–162
- Harry SR, see Carney CK (2007) 270: 155–185
- Heise C, Bier FF (2005) Immobilization of DNA on Microarrays. 261: 1–25
- Heitz V, see Collin J-P (2005) 262: 29–62
- Herrmann C, Reiher M (2007) First-Principles Approach to Vibrational Spectroscopy of Biomolecules. 268: 85–132
- Higson SPJ, see Davis F (2005) 255: 97–124
- Hirayama N, see Sakai K (2007) 269
- Hirst AR, Smith DK (2005) Dendritic Gelators. 256: 237–273
- Holzwarth AR, see Balaban TS (2005) 258: 1–38
- Houk RJT, Tobey SL, Anslyn EV (2005) Abiotic Guanidinium Receptors for Anion Molecular Recognition and Sensing. 255: 199–229
- Huc I, see Brizard A (2005) 256: 167–218
- Ihmels H, Otto D (2005) Intercalation of Organic Dye Molecules into Double-Stranded DNA – General Principles and Recent Developments. 258: 161–204
- Imai H (2007) Self-Organized Formation of Hierarchical Structures. 270: 43–72
- Indelli MT, see Chiorboli C (2005) 257: 63–102
- Inoue Y, see Borovkov VV (2006) 265: 89–146
- Ishii A, Nakayama J (2005) Carbodithioic Acid Esters. 251: 181–225
- Ishii A, Nakayama J (2005) Carboselenothioic and Carbodiselenoic Acid Derivatives and Related Compounds. 251: 227–246
- Ishi-i T, Shinkai S (2005) Dye-Based Organogels: Stimuli-Responsive Soft Materials Based on One-Dimensional Self-Assembling Aromatic Dyes. 258: 119–160
- James DK, Tour JM (2005) Molecular Wires. 257: 33–62
- Jones W, see Trask AV (2005) 254: 41–70
- Justicia J, see Cuerva JM (2006) 264: 63–92
- Kambe N, see Fujiwara S-i (2005) 251: 87–140
- Kano N, Kawashima T (2005) Dithiocarboxylic Acid Salts of Group 1–17 Elements (Except for Carbon). 251: 141–180
- Kappe CO, see Kremsner JM (2006) 266: 233–278
- Kaptein B, see Kellogg RM (2007) 269
- Karlen SD, Garcia-Garibay MA (2005) Amphidynamic Crystals: Structural Blueprints for Molecular Machines. 262: 179–227
- Kato S, Niyomura O (2005) Group 1–17 Element (Except Carbon) Derivatives of Thio-, Seleno- and Telluro-Carboxylic Acids. 251: 19–85
- Kato S, see Niyomura O (2005) 251: 1–12

- Kato T, Mizoshita N, Moriyama M, Kitamura T (2005) Gelation of Liquid Crystals with Self-Assembled Fibers. 256: 219–236
- Kaul M, see Pilch DS (2005) 253: 179–204
- Kaupp G (2005) Organic Solid-State Reactions with 100% Yield. 254: 95–183
- Kawasaki T, see Okahata Y (2005) 260: 57–75
- Kawashima T, see Kano N (2005) 251: 141–180
- Kay ER, Leigh DA (2005) Hydrogen Bond-Assembled Synthetic Molecular Motors and Machines. 262: 133–177
- Kellogg RM, Kaptein B, Vries TR (2007) Dutch Resolution of Racemates and the Roles of Solid Solution Formation and Nucleation Inhibition. 269
- King GC, see Di Giusto DA (2005) 261: 131–168
- Kirchner B, see Thar J (2007) 268: 133–171
- Kitamura T, see Kato T (2005) 256: 219–236
- Kniep R, Simon P (2007) Fluorapatite-Gelatine-Nanocomposites: Self-Organized Morphogenesis, Real Structure and Relations to Natural Hard Materials. 270: 73–125
- Komatsu K (2005) The Mechanochemical Solid-State Reaction of Fullerenes. 254: 185–206
- Kremsner JM, Stadler A, Kappe CO (2006) The Scale-Up of Microwave-Assisted Organic Synthesis. 266: 233–278
- Kriegisch V, Lambert C (2005) Self-Assembled Monolayers of Chromophores on Gold Surfaces. 258: 257–313
- Lacôte E, see Albert M (2006) 264: 1–62
- Lahav M, see Weissbuch I (2005) 259: 123–165
- Lambert C, see Kriegisch V (2005) 258: 257–313
- Lansiaux A, see Dias N (2005) 253: 89–108
- Larhed M, see Nilsson P (2006) 266: 103–144
- Larhed M, see Wannberg J (2006) 266: 167–197
- Larsson O, see Du Q (2005) 261: 45–61
- Leigh DA, Pérez EM (2006) Dynamic Chirality: Molecular Shuttles and Motors. 265: 185–208
- Leigh DA, see Kay ER (2005) 262: 133–177
- Leiserowitz L, see Weissbuch I (2005) 259: 123–165
- Lhoták P (2005) Anion Receptors Based on Calixarenes. 255: 65–95
- Li WP, Meyer LA, Anderson CJ (2005) Radiopharmaceuticals for Positron Emission Tomography Imaging of Somatostatin Receptor Positive Tumors. 252: 179–192
- Liang Z, see Du Q (2005) 261: 45–61
- Lingerfelt L, see Guiseppi-Elie A (2005) 260: 161–186
- Liu S (2005) 6-Hydrazinonicotinamide Derivatives as Bifunctional Coupling Agents for ^{99m}Tc -Labeling of Small Biomolecules. 252: 117–153
- Liu S, Robinson SP, Edwards DS (2005) Radiolabeled Integrin $\alpha_v\beta_3$ Antagonists as Radiopharmaceuticals for Tumor Radiotherapy. 252: 193–216
- Liu XY (2005) Gelation with Small Molecules: from Formation Mechanism to Nanostructure Architecture. 256: 1–37
- Luderer F, Walschus U (2005) Immobilization of Oligonucleotides for Biochemical Sensing by Self-Assembled Monolayers: Thiol-Organic Bonding on Gold and Silanization on Silica Surfaces. 260: 37–56
- Maeda K, Yashima E (2006) Dynamic Helical Structures: Detection and Amplification of Chirality. 265: 47–88
- Magnera TF, Michl J (2005) Altitudinal Surface-Mounted Molecular Rotors. 262: 63–97
- Maini L, see Braga D (2005) 254: 71–94

- Malacria M, see Albert M (2006) 264: 1–62
- Marquette CA, Blum LJ (2005) Beads Arraying and Beads Used in DNA Chips. 261: 113–129
- Mascini M, see Palchetti I (2005) 261: 27–43
- Matsumoto A (2005) Reactions of 1,3-Diene Compounds in the Crystalline State. 254: 263–305
- McGhee AM, Procter DJ (2006) Radical Chemistry on Solid Support. 264: 93–134
- Meyer B, Möller H (2007) Conformation of Glycopeptides and Glycoproteins. 267
- Meyer LA, see Li WP (2005) 252: 179–192
- Michl J, see Magnera TF (2005) 262: 63–97
- Milea JS, see Smith CL (2005) 261: 63–90
- Mizoshita N, see Kato T (2005) 256: 219–236
- Möller H, see Meyer B (2007) 267
- Moonen NNP, Flood AH, Fernández JM, Stoddart JF (2005) Towards a Rational Design of Molecular Switches and Sensors from their Basic Building Blocks. 262: 99–132
- Moriyama M, see Kato T (2005) 256: 219–236
- Murai T (2005) Thio-, Seleno-, Telluro-Amides. 251: 247–272
- Murakami H (2007) From Racemates to Single Enantiomers – Chiral Synthetic Drugs over the last 20 Years. 269
- Mutule I, see Suna E (2006) 266: 49–101
- Naka K (2007) Delayed Action of Synthetic Polymers for Controlled Mineralization of Calcium Carbonate. 271: 119–154
- Nakayama J, see Ishii A (2005) 251: 181–225
- Nakayama J, see Ishii A (2005) 251: 227–246
- Neese F, see Sinnecker S (2007) 268: 47–83
- Nguyen GH, see Smith CL (2005) 261: 63–90
- Nicolau DV, Sawant PD (2005) Scanning Probe Microscopy Studies of Surface-Immobilised DNA/Oligonucleotide Molecules. 260: 113–160
- Nilsson P, Olofsson K, Larhed M (2006) Microwave-Assisted and Metal-Catalyzed Coupling Reactions. 266: 103–144
- Niyomura O, Kato S (2005) Chalcogenocarboxylic Acids. 251: 1–12
- Niyomura O, see Kato S (2005) 251: 19–85
- Nohira H, see Sakai K (2007) 269
- Oda R, see Brizard A (2005) 256: 167–218
- Okahata Y, Kawasaki T (2005) Preparation and Electron Conductivity of DNA-Aligned Cast and LB Films from DNA-Lipid Complexes. 260: 57–75
- Okamura T, see Ueyama N (2007) 271: 155–193
- Oller-López JL, see Cuerva JM (2006) 264: 63–92
- Olofsson K, see Nilsson P (2006) 266: 103–144
- Oltra JE, see Cuerva JM (2006) 264: 63–92
- Onoda A, see Ueyama N (2007) 271: 155–193
- Otto D, see Ihmels H (2005) 258: 161–204
- Palchetti I, Mascini M (2005) Electrochemical Adsorption Technique for Immobilization of Single-Stranded Oligonucleotides onto Carbon Screen-Printed Electrodes. 261: 27–43
- Pascal R, Boiteau L, Commeyras A (2005) From the Prebiotic Synthesis of α -Amino Acids Towards a Primitive Translation Apparatus for the Synthesis of Peptides. 259: 69–122
- Paulo A, see Santos I (2005) 252: 45–84
- Pérez EM, see Leigh DA (2006) 265: 185–208

- Pilch DS, Kaul M, Barbieri CM (2005) Ribosomal RNA Recognition by Aminoglycoside Antibiotics. 253: 179–204
- Pividori MI, Alegret S (2005) DNA Adsorption on Carbonaceous Materials. 260: 1–36
- Piwnica-Worms D, see Sharma V (2005) 252: 155–178
- Polito M, see Braga D (2005) 254: 71–94
- Poulin-Kerstien AT, see Dervan PB (2005) 253: 1–31
- Procter DJ, see McGhee AM (2006) 264: 93–134
- Quiclet-Sire B, Zard SZ (2006) The Degenerative Radical Transfer of Xanthates and Related Derivatives: An Unusually Powerful Tool for the Creation of Carbon–Carbon Bonds. 264: 201–236
- Ratner MA, see Weiss EA (2005) 257: 103–133
- Raymond KN, see Seeber G (2006) 265: 147–184
- Rebek Jr J, see Scarso A (2006) 265: 1–46
- Reckien W, see Thar J (2007) 268: 133–171
- Reiher M, see Herrmann C (2007) 268: 85–132
- Renaud P, see Darmency V (2006) 263: 71–106
- Robinson SP, see Liu S (2005) 252: 193–216
- Saha-Möller CR, see You C-C (2005) 258: 39–82
- Sakai K, Sakurai R, Hirayama N (2007) Molecular Mechanisms of Dielectrically Controlled Resolution (DCR). 269
- Sakai K, Sakurai R, Nohira H (2007) New Resolution Technologies Controlled by Chiral Discrimination Mechanisms. 269
- Sakamoto M (2005) Photochemical Aspects of Thiocarbonyl Compounds in the Solid-State. 254: 207–232
- Sakurai R, see Sakai K (2007) 269
- Sakurai R, see Sakai K (2007) 269
- Saladino R, Crestini C, Costanzo G, DiMauro E (2005) On the Prebiotic Synthesis of Nucleobases, Nucleotides, Oligonucleotides, Pre-RNA and Pre-DNA Molecules. 259: 29–68
- Santos I, Paulo A, Correia JDG (2005) Rhenium and Technetium Complexes Anchored by Phosphines and Scorpionates for Radiopharmaceutical Applications. 252: 45–84
- Santos M, see Szathmáry E (2005) 259: 167–211
- Sato K (2007) Inorganic–Organic Interfacial Interactions in Hydroxyapatite Mineralization Processes. 270: 127–153
- Sauvage J-P, see Collin J-P (2005) 262: 29–62
- Sawant PD, see Nicolau DV (2005) 260: 113–160
- Scandola F, see Chiorboli C (2005) 257: 63–102
- Scarso A, Rebek Jr J (2006) Chiral Spaces in Supramolecular Assemblies. 265: 1–46
- Scheffer JR, Xia W (2005) Asymmetric Induction in Organic Photochemistry via the Solid-State Ionic Chiral Auxiliary Approach. 254: 233–262
- Schenning APHJ, see Ajayaghosh A (2005) 258: 83–118
- Schmidtchen FP (2005) Artificial Host Molecules for the Sensing of Anions. 255: 1–29 Author Index Volumes 251–255
- Schindler J, see Faigl F (2007) 269
- Schoof S, see Wolter F (2007) 267
- De Schryver F, see De Feyter S (2005) 258: 205–255
- Schulten K, see Dittrich M (2007) 268: 319–347

- Seeber G, Tiedemann BEF, Raymond KN (2006) Supramolecular Chirality in Coordination Chemistry. 265: 147–184
- Seitz O, see Haase C (2007) 267
- Senn HM, Thiel W (2007) QM/MM Methods for Biological Systems. 268: 173–289
- Sewell SL, see Carney CK (2007) 270: 155–185
- Sharma V, Piwnica-Worms D (2005) Monitoring Multidrug Resistance P-Glycoprotein Drug Transport Activity with Single-Photon-Emission Computed Tomography and Positron Emission Tomography Radiopharmaceuticals. 252: 155–178
- Shinkai S, see Ishi-i T (2005) 258: 119–160
- Sibi MP, see Zimmerman J (2006) 263: 107–162
- Siebbeles LDA, see Grozema FC (2005) 257: 135–164
- Silvi S, see Balzani V (2005) 262: 1–27
- Simon P, see Kniep R (2007) 270: 73–125
- Sinnecker S, Neese F (2007) Theoretical Bioinorganic Spectroscopy. 268: 47–83
- Skrydstrup T, see Hansen SG (2006) 264: 135–162
- Smith CL, Milea JS, Nguyen GH (2005) Immobilization of Nucleic Acids Using Biotin-Strept(avidin) Systems. 261: 63–90
- Smith DK, see Hirst AR (2005) 256: 237–273
- Specker D, Wittmann V (2007) Synthesis and Application of Glycopeptide and Glycoprotein Mimetics. 267
- Stadler A, see Kremsner JM (2006) 266: 233–278
- Stibor I, Zlatušková P (2005) Chiral Recognition of Anions. 255: 31–63
- Stoddart JF, see Moonen NNP (2005) 262: 99–132
- Strauss CR, Varma RS (2006) Microwaves in Green and Sustainable Chemistry. 266: 199–231
- Suk D-H, see Crich D (2006) 263: 1–38
- Suksai C, Tuntulani T (2005) Chromogenetic Anion Sensors. 255: 163–198
- Sun J-S, see Escudé C (2005) 253: 109–148
- Suna E, Mutule I (2006) Microwave-assisted Heterocyclic Chemistry. 266: 49–101
- Süssmuth RD, see Wolter F (2007) 267
- Svith H, see Daasbjerg K (2006) 263: 39–70
- Swerdlow H, see Du Q (2005) 261: 45–61
- Szathmáry E, Santos M, Fernando C (2005) Evolutionary Potential and Requirements for Minimal Protocells. 259: 167–211
- Taira S, see Yokoyama K (2005) 261: 91–112
- Takahashi H, see Tamura R (2007) 269
- Takahashi K, see Ueyama N (2007) 271: 155–193
- Tamiaki H, see Balaban TS (2005) 258: 1–38
- Tamura R, Takahashi H, Fujimoto D, Ushio T (2007) Mechanism and Scope of Preferential Enrichment, a Symmetry-Breaking Enantiomeric Resolution Phenomenon. 269
- Thar J, Reckien W, Kirchner B (2007) Car-Parrinello Molecular Dynamics Simulations and Biological Systems. 268: 133–171
- Thayer DA, Wong C-H (2007) Enzymatic Synthesis of Glycopeptides and Glycoproteins. 267
- Thiel W, see Senn HM (2007) 268: 173–289
- Tiedemann BEF, see Seeber G (2006) 265: 147–184
- Tobey SL, see Houk RJT (2005) 255: 199–229
- Toda F (2005) Thermal and Photochemical Reactions in the Solid-State. 254: 1–40
- Tour JM, see James DK (2005) 257: 33–62

- Trask AV, Jones W (2005) Crystal Engineering of Organic Cocrystals by the Solid-State Grinding Approach. 254: 41–70
- Tuntulani T, see Suksai C (2005) 255: 163–198
- Uccelli L, see Boschi A (2005) 252: 85–115
- Ueyama N, Takahashi K, Onoda A, Okamura T, Yamamoto H (2007) Inorganic–Organic Calcium Carbonate Composite of Synthetic Polymer Ligands with an Intramolecular $\text{NH} \cdots \text{O}$ Hydrogen Bond. 271: 155–193
- Ushio T, see Tamura R (2007) 269
- Varma RS, see Strauss CR (2006) 266: 199–231
- Veciana J, see Amabilino DB (2006) 265: 253–302
- Venturi M, see Balzani V (2005) 262: 1–27
- Vezin H, see Dias N (2005) 253: 89–108
- Vögtle F, see Fages F (2005) 256: 77–131
- Vögtle M, see Žinić M (2005) 256: 39–76
- Volkmer D, see Fricke M (2007) 270: 1–41
- Vries TR, see Kellogg RM (2007) 269
- Walschus U, see Luderer F (2005) 260: 37–56
- Walton JC (2006) Unusual Radical Cyclisations. 264: 163–200
- Wannberg J, Ersmark K, Larhed M (2006) Microwave-Accelerated Synthesis of Protease Inhibitors. 266: 167–197
- Warman JM, see Grozema FC (2005) 257: 135–164
- Warren JD, Geng X, Danishefsky SJ (2007) Synthetic Glycopeptide-Based Vaccines. 267
- Wasielewski MR, see Weiss EA (2005) 257: 103–133
- Weiss EA, Wasielewski MR, Ratner MA (2005) Molecules as Wires: Molecule-Assisted Movement of Charge and Energy. 257: 103–133
- Weissbuch I, Leiserowitz L, Lahav M (2005) Stochastic “Mirror Symmetry Breaking” via Self-Assembly, Reactivity and Amplification of Chirality: Relevance to Abiotic Conditions. 259: 123–165
- Williams LD (2005) Between Objectivity and Whim: Nucleic Acid Structural Biology. 253: 77–88
- Wittmann V, see Specker D (2007) 267
- Wright DW, see Carney CK (2007) 270: 155–185
- Wolter F, Schoof S, Süßmuth RD (2007) Synopsis of Structural, Biosynthetic, and Chemical Aspects of Glycopeptide Antibiotics. 267
- Wong C-H, see Thayer DA (2007) 267
- Wong KM-C, see Yam VW-W (2005) 257: 1–32
- Würthner F, see You C-C (2005) 258: 39–82
- Xia W, see Scheffer JR (2005) 254: 233–262
- Yam VW-W, Wong KM-C (2005) Luminescent Molecular Rods – Transition-Metal Alkynyl Complexes. 257: 1–32
- Yamamoto H, see Ueyama N (2007) 271: 155–193
- Yashima E, see Maeda K (2006) 265: 47–88
- Yokoyama K, Taira S (2005) Self-Assembly DNA-Conjugated Polymer for DNA Immobilization on Chip. 261: 91–112

- Yoshikawa I, see Araki K (2005) 256: 133–165
- Yoshioka R (2007) Racemization, Optical Resolution and Crystallization-Induced Asymmetric Transformation of Amino Acids and Pharmaceutical Intermediates. 269
- You C-C, Dobrawa R, Saha-Möller CR, Würthner F (2005) Metallosupramolecular Dye Assemblies. 258: 39–82
- Yu J, see Dittrich M (2007) 268: 319–347
- Yu S-H (2007) Bio-inspired Crystal Growth by Synthetic Templates. 271: 79–118
- Zampella G, see Bertini L (2007) 268: 1–46
- Zard SZ, see Quiclet-Sire B (2006) 264: 201–236
- Zhang W (2006) Microwave-Enhanced High-Speed Fluorous Synthesis. 266: 145–166
- Zhang X-E, Deng J-Y (2005) Detection of Mutations in Rifampin-Resistant *Mycobacterium Tuberculosis* by Short Oligonucleotide Ligation Assay on DNA Chips (SOLAC). 261: 169–190
- Zimmerman J, Sibi MP (2006) Enantioselective Radical Reactions. 263: 107–162
- Žinić M, see Fages F (2005) 256: 77–131
- Žinić M, Vögtle F, Fages F (2005) Cholesterol-Based Gelators. 256: 39–76
- Zipse H (2006) Radical Stability—A Theoretical Perspective. 263: 163–190
- Zlatušková P, see Stibor I (2005) 255: 31–63

Subject Index

- Active site models 10
Additive schemes 181
Adiabatic dynamics 219
Amino acids 155
Antiferromagnetic coupling, hydrogenases 2
Aquaporins 155
ATP hydrolysis 319, 320, 323
- Bacteriorhodopsin 155
Basis sets 145
Bennett–Chandler 293
Bioinorganic spectroscopy 50
Biomimetic models 7
Biomimetics 158
–, nitrogen fixation 159
Biomolecular isomerization 306
Biomolecular reaction modelling 220
Biomolecules, large 115
Blue copper proteins, spectroscopy 70
Born–Oppenheimer approximation 91
Boundary atoms 197
Boundary schemes 189
Broken symmetry approach 17
- Carbon monoxide, [Fe] hydrogenase 28
Car–Parrinello simulations 133, 149
Caspase-3 157
Chirality, VCD/ROA 110
Chromatin formation 292
CO photolysis, TDDFT 28
Coordination compounds 1
–, bio-inspired 12
Copper complexes, binuclear mixed valence 78
CREP-binding protein (CBP) 155
Cysteine radical, *g*-tensor anisotropy 67
- Density functional theory (DFT) 1, 33, 48
Difference-IR/-Raman 105
Dispersion corrections 143
DMSO, diffusion in water 135
DNA 157
–, base pairs, binding/unbinding 307
DNA helicase 319
DNA replication 292
- Electron localization function (ELF), water dimer 153
Electronic potential energy hypersurface 95
Electronic resonance, Raman 118
Electronic structure 141
Embedding 182, 325
ENDOR spectra, DFT 78
Environment effects, molecules in solution 97
Enzymatic mechanisms 173
Enzyme catalysis 15, 224, 311
–, cycle 15
Exchange correlation functional 143
Excited state hypersurfaces 96
Excited states MD 158
- F1-ATPase 320
Fe(V) 76
Fictitious mass μ 150
First-principles methods 86
Fock operator, solvated 328
Free-energy methods 173
Free-energy perturbation 215
Frozen localized orbitals 199
- Galactose oxidase 156
Generalized gradient approximation (GGA) 144
Glucanase 157

- Ground state hypersurfaces 96
 GTP hydrolysis 156
 Guanidinium receptor 135

 H-cluster, modeling 17
 Harmonic approximation 92
 Helicases 319
 HIV-1 protease 156
 Hohenberg–Kohn theorems 141
 Hydrogenases 1, 5
 –, [Fe] hydrogenases 2, 5
 –, [NiFe] hydrogenases 5
 Hyperplanes 295
 Hypersurfaces 96
 HYSORE spectra simulations, DFT 78

 Importance sampling 299
 IR 105
 Iron complexes, mononuclear 71

 Kinesin 320
 Kohn–Sham equations 142

 Langevin simulations, PcrA function 339
 Leap frog algorithm 332
 Linear response theory, ground state 53
 Link atoms 191, 326
 Lipid bilayers 312
 Lipids 292
 Local contributions 118
 Local density approximation (LDA) 143

 Magnetic spectroscopy, ground state 50
 Metadynamics 219
 Metalloenzymes 1, 156
 Molecular conformation 114
 Molecular dynamics 173
 –, simulations 137
 –, –, PcrA 319
 Molecular motor 319, 322
 Molecular simulations 173
 Molecular structures, optimization 9
 Monte Carlo simulations 214
 Mössbauer spectroscopy 60
 Motion, equations 137
 Myoglobin 156
 Myosin 320

 Nitrogen fixation 159
 Non-equilibrium thermodynamic integration 217
 Nose–Hoover thermostat 139, 297
 NRVS 124
 Nucleosomes transitions 292

 Optical spectroscopy 62
 Optimization 173
 –, large systems 210
 Organic radicals, high precision studies 64
 Orientational averaging 102

 Path ensemble 296
 –, Monte Carlo 299
 PcrA helicase 319
 Peptides 155
 Phenoxyl radicals, molecular g-tensors 65
 Photochemical reactions, [Fe] hydrogenases 2
 –, modeling 26
 Plane waves 145
 Protein folding 308
 Proteins 155
 Pseudopotentials 147

 QM/MM 151, 319
 –, biomolecular 227
 –, bonded interactions 188
 –, boundary, covalent bonds 189
 –, combined 173
 –, energy expressions 179
 –, free-energy perturbation 215
 –, implementations 205
 –, interactions, long-range electrostatic 186
 –, method 178
 –, molecular-dynamics 214
 –, optimization techniques 211
 –, packages 205
 –, reaction-path potential 219
 –, studies, biomolecular 227
 –, van der Waals interactions 187
 QM/MM-related approaches 207
 Quantum chemistry 1
 Quinone radical anion–solvent interactions 66

- Raman 105
Reaction-path techniques 213
Reactivity, modeling 9
Resonance CARS 122
Resonance ROA 124
Restrained electrostatic potential (RESP) charge method 326
Rhodopsin 155, 159
Ribozymatic reactions 292
RNA 157
–, folding 292
ROA theory 113

Serine protease 156
Solvated Fock operator 328
Solvent effects 116
Spontaneous resonance Raman scattering 118
Stable states 298
Stationary points 209
Stochastic modeling, PcrA 319

Subtractive schemes 179
Switches, biochemical networks 312

TDDFT modeling 26
Thermostats 139, 297
Time-dependent density functional theory (TDDFT) 26, 38, 158
Transcription 292
Transition state theory (TST) 293
Transition-path sampling 218, 296
Translocation 292
Trans-membrane transport 293
Triosephosphate isomerase 157

VCD theory 112
Vibrational normal modes 91
Vibrational spectroscopy 86
–, intensities 98

Water clusters, cooperative effects 136
Weighted density approximation (WDA) 144