

TOPICS IN CURRENT CHEMISTRY

307

Multiscale Molecular Methods in Applied Chemistry

Topics in Current Chemistry

Editorial Board:

**K.N. Houk • C.A. Hunter • M.J. Krische • J.-M. Lehn
S.V. Ley • M. Olivucci • J. Thiem • M. Venturi • P. Vogel
C.-H. Wong • H. Wong • H. Yamamoto**

Topics in Current Chemistry

Recently Published and Forthcoming Volumes

Multiscale Molecular Methods in Applied Chemistry

Volume Editors: Barbara Kirchner, Jadran Vrabec
Vol. 307, 2012

Solid State NMR

Volume Editor: Jerry C. C. Chan
Vol. 306, 2012

Prion Proteins

Volume Editor: Jörg Tatzelt
Vol. 305, 2011

Microfluidics: Technologies and Applications

Volume Editor: Bingcheng Lin
Vol. 304, 2011

Photocatalysis

Volume Editor: Carlo Alberto Bignozzi
Vol. 303, 2011

Computational Mechanisms of Au and Pt Catalyzed Reactions

Volume Editors: Elena Soriano,
José Marco-Contelles
Vol. 302, 2011

Reactivity Tuning in Oligosaccharide Assembly

Volume Editors: Bert Fraser-Reid,
J. Cristóbal López
Vol. 301, 2011

Luminescence Applied in Sensor Science

Volume Editors: Luca Prodi, Marco Montalti,
Nelsi Zaccheroni
Vol. 300, 2011

Chemistry of Opioids

Volume Editor: Hiroshi Nagase
Vol. 299, 2011

Electronic and Magnetic Properties of Chiral Molecules and Supramolecular Architectures

Volume Editors: Ron Naaman,
David N. Beratan, David H. Waldeck
Vol. 298, 2011

Natural Products via Enzymatic Reactions

Volume Editor: Jörn Piel
Vol. 297, 2010

Nucleic Acid Transfection

Volume Editors: Wolfgang Bielke,
Christoph Erbacher
Vol. 296, 2010

Carbohydrates in Sustainable Development II

Volume Editors: Amélia P. Rauter,
Pierre Vogel, Yves Queneau
Vol. 295, 2010

Carbohydrates in Sustainable Development I

Volume Editors: Amélia P. Rauter,
Pierre Vogel, Yves Queneau
Vol. 294, 2010

Functional Metal-Organic Frameworks: Gas Storage, Separation and Catalysis

Volume Editor: Martin Schröder
Vol. 293, 2010

C-H Activation

Volume Editors: Jin-Quan Yu, Zhangjie Shi
Vol. 292, 2010

Asymmetric Organocatalysis

Volume Editor: Benjamin List
Vol. 291, 2010

Ionic Liquids

Volume Editor: Barbara Kirchner
Vol. 290, 2010

Orbitals in Chemistry

Volume Editor: Satoshi Inagaki
Vol. 289, 2009

Multiscale Molecular Methods in Applied Chemistry

Volume Editors: Barbara Kirchner · Jadran Vrabec

With Contributions by

R. Abrol · D. Bratko · C.D. Daub · L. Della Site · P.J. di Dio ·
W.A. Goddard III · G. Guevara-Carrion · H. Hasse · C. Holm ·
J. Hutter · A. Jaramillo-Botero · H.A. Karimi-Varzaneh · F.J. Keil ·
B. Kirchner · A. Luzar · J. Mueller · F. Müller-Plathe · R. Nielsen ·
T. Pascal · J.L. Rafferty · G.C. Schatz · M.R. Schure · J.I. Siepmann ·
J. Su · J. Vrabec · N.F.A. van der Vegt · S. Yockel

Editors

Prof. Barbara Kirchner
Wilhelm-Ostwald Institute of Physical
and Theoretical Chemistry
University of Leipzig
Linnéstr. 2
04103 Leipzig
Germany
bkirchner@uni-leipzig.de

Prof. Jadran Vrabec
Thermodynamics and Energy Technology
University of Paderborn
Warburger Str. 100
33098 Paderborn
Germany
Jadran.vrabec@uni-paderborn.de

ISSN 0340-1022 e-ISSN 1436-5049
ISBN 978-3-642-24967-9 e-ISBN 978-3-642-24968-6
DOI 10.1007/978-3-642-24968-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011940291

© Springer-Verlag Berlin Heidelberg 2012

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 to prosecution under the German Copyright Law.

The use of general descriptive names, 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.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Volume Editors

Prof. Barbara Kirchner

Wilhelm-Ostwald Institute of Physical
and Theoretical Chemistry
University of Leipzig
Linnéstr. 2
04103 Leipzig
Germany
bkirchner@uni-leipzig.de

Prof. Jadran Vrabec

Thermodynamics and Energy Technology
University of Paderborn
Warburger Str. 100
33098 Paderborn
Germany
Jadran.vrabec@uni-paderborn.de

Editorial Board

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. Steven V. Ley

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

Prof. Dr. Christopher A. Hunter

Department of Chemistry
University of Sheffield
Sheffield S3 7HF, United Kingdom
c.hunter@sheffield.ac.uk

Prof. Dr. Massimo Olivucci

Università di Siena
Dipartimento di Chimica
Via A De Gasperi 2
53100 Siena, Italy
olivucci@unisi.it

Prof. Michael J. Krische

University of Texas at Austin
Chemistry & Biochemistry Department
1 University Station A5300
Austin TX, 78712-0165, USA
mkrische@mail.utexas.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. Dr. Jean-Marie Lehn

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

Prof. Dr. Margherita Venturi

Dipartimento di Chimica
Università di Bologna
via Selmi 2
40126 Bologna, Italy
margherita.venturi@unibo.it

Prof. Dr. Pierre Vogel

Laboratory of Glycochemistry
and Asymmetric Synthesis
EPFL – Ecole polytechnique fédérale
de Lausanne
EPFL SB ISIC LGSA
BCH 5307 (Bat.BCH)
1015 Lausanne, Switzerland
pierre.vogel@epfl.ch

Prof. Dr. Chi-Huey Wong

Professor of Chemistry, Scripps Research
Institute
President of Academia Sinica
Academia Sinica
128 Academia Road
Section 2, Nankang
Taipei 115
Taiwan
chwong@gate.sinica.edu.tw

Prof. Dr. Henry Wong

The Chinese University of Hong Kong
University Science Centre
Department of Chemistry
Shatin, New Territories
hncwong@cuhk.edu.hk

Prof. Dr. Hisashi Yamamoto

Arthur Holly Compton Distinguished
Professor
Department of Chemistry
The University of Chicago
5735 South Ellis Avenue
Chicago, IL 60637
773-702-5059
USA
yamamoto@uchicago.edu

Topics in Current Chemistry

Also Available Electronically

Topics in Current Chemistry is included in Springer's eBook package *Chemistry and Materials Science*. If a library does not opt for the whole package the book series may be bought on a subscription basis. Also, all back volumes are available electronically.

For all customers with a print standing order we offer free access to the electronic volumes of the series published in the current year.

If you do not have access, you can still view the table of contents of each volume and the abstract of each article by going to the SpringerLink homepage, clicking on "Chemistry and Materials Science," under Subject Collection, then "Book Series," under Content Type and finally by selecting *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 by typing in *Topics in Current Chemistry*.

Color figures are published in full color in the electronic version on SpringerLink.

Aims and Scope

The series *Topics in Current Chemistry* presents critical reviews of the present and future trends in modern chemical research. The scope includes all areas of chemical science, including the interfaces with related disciplines such as biology, medicine, and materials science.

The objective of each thematic volume is to give the non-specialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights of interest to a larger scientific audience are emerging.

Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5–10 years are presented, using selected examples to illustrate the principles discussed. A description of the laboratory procedures involved is often useful to the reader. The coverage is not exhaustive in data, but rather conceptual, concentrating on the methodological thinking that will allow the non-specialist reader to understand the information presented.

Discussion of possible future research directions in the area is welcome.

Review articles for the individual volumes are invited by the volume editors.

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

Impact Factor 2010: 2.067; Section “Chemistry, Multidisciplinary”: Rank 44 of 144

Preface

Driven by advances in simulation methodology and computer hardware, an increasing spectrum of topics in applied chemistry is becoming accessible via the use of computational methods. In recent years, multiscale molecular simulations of complete and realistic processes have thereby emerged. This volume of *Topics in Current Chemistry* focuses on molecular methods for large and complex systems, such as technical chemical processes. It spans the spectrum from representative methodological approaches containing static quantum chemical calculations, ab initio molecular simulations, and traditional force field methods, to coarse-grained simulations from a multiscale perspective. Each field of theoretical chemistry is highly advanced, and although there is still room for further developments, these do not seem as tremendous as ten years ago if only one scale is considered. Current developments are often concerned with the refinement of old methods rather than with introducing new ones. Because the considered systems have become larger and more complex, the next step towards their accurate description lies in combining the advantages of more than one method, i.e. in multiscale approaches.

The multiscale aspect comes into play on different levels; one level is given by the well-known hybrid approach, i.e. combining existing methods in a concurrent calculation. Separate calculations applying different methods to the same system provide another approach. Coarser methods can be refined by more accurate methods and more accurate methods speeded up by making them more coarse. The investigated systems range from a single molecule to industrial processes. On the level of fluid properties, a scale-bridging ansatz considers molecular properties such as electronic energies, as well as thermodynamic quantities such as pressure. Thus, a connection between different levels is established. Furthermore, dynamic heterogeneity is accessible, and therefore a broader scale range in terms of dynamics can be covered. As microscopic movements on the femtosecond scale may substantially influence entire processes, the consequences for the macroscopic level are also taken into account.

The contributions to this volume cover applied topics such as hierarchically structured materials, molecular reaction dynamics, chemical catalysis, thermodynamics of aggregated phases, molecular self-assembly, chromatography, nanoscale

electrowetting, polyelectrolytes, charged colloids and macromolecules. Throughout, the authors have aimed at quantitative and qualitative predictions for complex systems in technical chemistry and thus in real-world applications. The nine chapters are structured in three groups: 1. *From first-principle calculations to complex systems via several routes* (Jaramillo-Botero et al., Yockel and Schatz, Keil, and Kirchner et al.), 2. *Making molecular dynamics simulations larger and accessing more complex situations* (Daub et al., Rafferty et al., and Guevara-Carrion et al.) and 3. *Coarse grained modelling reaching out afar* (Delle Site et al., and Karimi-Varzaneh and Müller-Plathe).

We would like to thank all the authors as well as all those who have facilitated this volume, and hope that readers will consider it as a helpful tool for obtaining an overview of the recent developments in the field of multiscale molecular methods in applied chemistry.

Leipzig and Paderborn

Barbara Kirchner
Jadran Vrabec

Contents

First-Principles-Based Multiscale, Multiparadigm Molecular Mechanics and Dynamics Methods for Describing Complex Chemical Processes	1
Andres Jaramillo-Botero, Robert Nielsen, Ravi Abrol, Julius Su, Tod Pascal, Jonathan Mueller, and William A. Goddard III	
Dynamic QM/MM: A Hybrid Approach to Simulating Gas-Liquid Interactions	43
Scott Yockel and George C. Schatz	
Multiscale Modelling in Computational Heterogeneous Catalysis	69
F.J. Keil	
Real-World Predictions from Ab Initio Molecular Dynamics Simulations	109
Barbara Kirchner, Philipp J. di Dio, and Jürg Hutter	
Nanoscale Wetting Under Electric Field from Molecular Simulations ..	155
Christopher D. Daub, Dusan Bratko, and Alenka Luzar	
Molecular Simulations of Retention in Chromatographic Systems: Use of Biased Monte Carlo Techniques to Access Multiple Time and Length Scales	181
Jake L. Rafferty, J. Ilja Siepmann, and Mark R. Schure	
Thermodynamic Properties for Applications in Chemical Industry via Classical Force Fields	201
Gabriela Guevara-Carrion, Hans Hasse, and Jadran Vrabec	

Multiscale Approaches and Perspectives to Modeling Aqueous Electrolytes and Polyelectrolytes	251
Luigi Delle Site, Christian Holm, and Nico F.A. van der Vegt	
Coarse-Grained Modeling for Macromolecular Chemistry	295
Hossein Ali Karimi-Varzaneh and Florian Müller-Plathe	
Index	323

First-Principles-Based Multiscale, Multiparadigm Molecular Mechanics and Dynamics Methods for Describing Complex Chemical Processes

Andres Jaramillo-Botero, Robert Nielsen, Ravi Abrol, Julius Su, Tod Pascal, Jonathan Mueller and William A. Goddard III

Abstract We expect that systematic and seamless computational upscaling and downscaling for modeling, predicting, or optimizing material and system properties and behavior with atomistic resolution will eventually be sufficiently accurate and practical that it will transform the mode of development in the materials, chemical, catalysis, and Pharma industries. However, despite truly dramatic progress in methods, software, and hardware, this goal remains elusive, particularly for systems that exhibit inherently complex chemistry under normal or extreme conditions of temperature, pressure, radiation, and others. We describe here some of the significant progress towards solving these problems via a general multiscale, multiparadigm strategy based on first-principles quantum mechanics (QM), and the development of breakthrough methods for treating reaction processes, excited electronic states, and weak bonding effects on the conformational dynamics of large-scale molecular systems. These methods have resulted directly from filling in the physical and chemical gaps in existing theoretical and computational models, within the multiscale, multiparadigm strategy. To illustrate the procedure we demonstrate the application and transferability of such methods on an ample set of challenging problems that span multiple fields, system length- and timescales, and that lay beyond the realm of existing computational or, in some case, experimental approaches, including understanding the solvation effects on the reactivity of organic and organometallic structures, predicting transmembrane protein structures, understanding carbon nanotube nucleation and growth, understanding the effects of electronic excitations in materials subjected to extreme conditions of temperature and pressure, following the dynamics and energetics of long-term conformational evolution of DNA

A. Jaramillo-Botero (✉), R. Nielsen, R. Abrol, J. Su, T. Pascal, J. Mueller,
and W.A. Goddard III (✉)

Chemistry and Chemical Engineering, California Institute of Technology, Mail code 139-74,
1200 E California Blvd, Pasadena, CA 91125, USA

e-mail: ajaramil@wag.caltech.edu, wag@wag.caltech.edu

macromolecules, and predicting the long-term mechanisms involved in enhancing the mechanical response of polymer-based hydrogels.

Keywords Multiscale modeling, Nanotube growth, Non-adiabatic molecular dynamics, Organometallic structures, Protein structure prediction, Reactive molecular dynamics

Contents

1	First Principles-Based Multiscale, Multiparadigm Simulations	2
2	The Role of QM in Multiscale Modeling	4
2.1	The Wave Equation for Matter	4
2.2	Approximations to Schrödinger's Equation	6
3	From QM to Molecular Mechanics/Dynamics: Force Fields	11
3.1	Conventional Force Fields	11
3.2	Simulating Complex Chemical Processes with FFs	16
4	Bridging MM/MD with the Mesoscale	26
4.1	Constrained and Coarse-Grain MD	26
5	Concluding Remarks	35
	References	37

1 First Principles-Based Multiscale, Multiparadigm Simulations

The computations required for accurate modeling and simulation of large-scale systems with atomistic resolution involve a hierarchy of levels of theory: quantum mechanics (QM) to determine the electronic states; force fields to average the electronics states and to obtain atom based forces (FF), molecular dynamics (MD) based on such an FF; mesoscale or coarse grain descriptions that average or homogenize atomic motions; and finally continuum level descriptions (see Fig. 1).

By basing computations on first principles QM it is possible to overcome the lack of experimental data to carry out accurate predictions with atomistic resolution, which would otherwise be impossible. Furthermore, QM provides the fundamental information required to describe quantum effects, electronically excited states, as well as reaction paths and barrier heights involved in chemical reactions processes. However, the practical scale for accurate QM today is <1,000 atoms per molecule or periodic cell (a length scale of a few nanometers) whereas the length scale for modeling supramolecular systems in biology may be in the tens of nanometers, while elucidating the interfacial effects between grains in composite materials may require hundreds of nanometers, and modeling turbulent fluid flows or shock-induced instabilities in multilayered materials may require micrometers. Thus, simulations of engineered materials and systems may require millions to billions of atoms, rendering QM methods impractical.

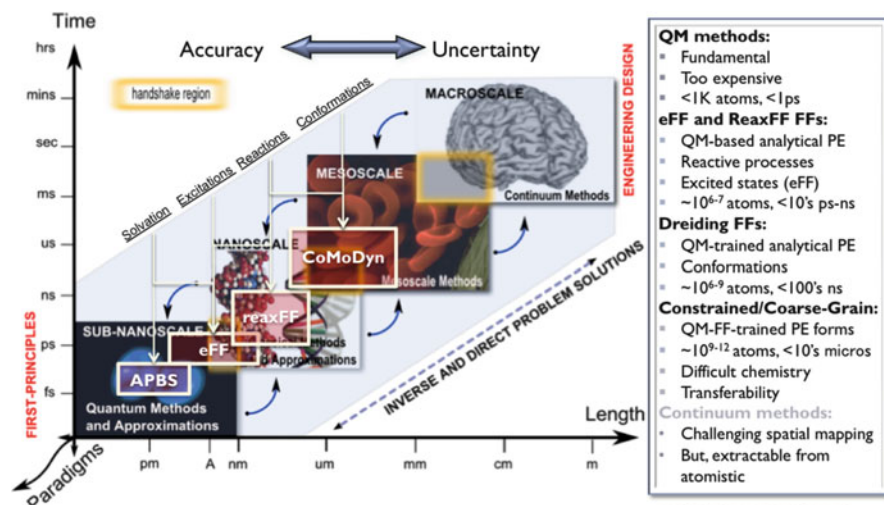


Fig. 1 Hierarchical multiscale, multiparadigm approach to materials modeling, from QM to the mesoscale, incorporating breakthrough methods to handle complex chemical processes (eFF, ReaxFF). Adapted from our multiscale group site <http://www.wag.caltech.edu/multiscale>

Nonetheless, QM methods are essential for accurately describing atomic-level composition, structure and energy states of materials, considering the influence of electronic degrees of freedom. By incorporating time-dependent information, the dynamics of a system under varying conditions may be explored from QM-derived forces, albeit within a limited timescale (<1 ps). The prominent challenge for theory and computation involves efficiently bridging, from QM first-principles, into larger length scales with predominantly heterogeneous spatial and density distributions, and longer timescales of simulation – enough to connect into engineering-level design variables – while retaining physicochemical accuracy and certainty. Equally challenging remains the inverse top-down engineering design problem, by which macroscopic material/process properties would be tunable from optimizing its atomic-level composition and structure. Our approach to this challenge has been to develop breakthrough methods to staple and extend hierarchically over existing ones, as well as to develop the necessary tools to enable continuous lateral (multiparadigm) and hierarchical (multiscale) couplings, between the different theories and models as a function of their length- and timescale range – a strategy referred to here as *First-Principles-Based Multiscale-Multiparadigm Simulation*.

The ultimate goal is a reversible bottom-up, top-down approach, based on first principles QM, to characterize properties of materials and processes at a hierarchy of length and timescales. This will improve our ability to design, analyze, and interpret experimental results, perform model-based prediction of phenomena, and to control precisely the multi-scale nature of material systems for multiple applications. Such an approach is now enabling us to study problems once thought to be intractable, including reactive turbulent flows, composite material instabilities,

dynamics of warm-dense-matter and plasma formation, functional molecular biology, and protein structure prediction, among others.

In this chapter, we describe some of our progress in theory, methods, computational techniques, and tools towards first-principles-based multiscale, multiparadigm simulations, in particular, for systems that exhibit intricate chemical behavior. We map the document over the hierarchical framework depicted in Fig. 1, threading the description from QM up through mesoscale classical approximations, presenting significant and relevant example applications to different fields at each level.

2 The Role of QM in Multiscale Modeling

QM relies solely on information about the atomic structure and composition of matter to describe its behavior. Significant progress has been made in the development of QM theory and its application, since its birth in the 1920s. The following sections present an overview of some parts of this evolution, describing how it provides the foundations for our first-principles-based multiscale, multiparadigm strategy to materials modeling and simulation.

2.1 The Wave Equation for Matter

Circa 1900 Max Planck suggested that light was quantized, and soon after, in 1905, Albert Einstein interpreted Planck’s quantum to be photons, particles of light, and proposed that the energy of a photon is proportional to its frequency. In 1924, Louis de Broglie argued that since light could be seen to behave under some conditions as particles [1] (e.g., Einstein’s explanation of the photoelectric effect) and at other times as waves (e.g., diffraction of light), one could also consider that matter has the same ambiguity of possessing both particle and wave properties. Starting with de Broglie’s idea that particles behave as waves and the fundamental (Hamilton’s) equations of motion (EOM) from classical mechanics, Erwin Schrödinger [2] developed the electronic wave equation that describes the space- (and time-) dependence of quantum mechanical systems [3], for an n -particle system as

$$\left[-\hbar^2 \sum_{i=1}^n \frac{\nabla^2}{2m_i} + V(r_1, r_2 \dots r_n, t) \right] \Psi(r_1, r_2 \dots r_n, t) = i\hbar \frac{\partial}{\partial t} \Psi(r_1, r_2 \dots r_n, t), \quad (1)$$

where the term in brackets corresponds to a linear operator that involves the kinetic (first term) and potential (second term, V) energy operators that act over the systems’ wavefunction, Ψ , and the right-hand side the quantized energy operator, corresponding to the full energy of the system, acting on the same wavefunction.

The wavefunction is interpreted as the probability amplitude for different configurations, r , of the system at different times, i.e., it describes the dynamics of the n -particles as a function of space, r , and time, t . In more abstract terms, (1) may also be written as

$$\hat{H}\psi = E\psi, \quad (2)$$

and take several different forms, depending on the physical situation.

In principle, all properties of all materials, with known atomic structure and composition, can be accurately described using (1) and one could then replace existing empirical methods used to model materials properties by a first principles or de novo computational approach design of materials and devices. Unfortunately, direct first principles applications of QM is highly impractical with current methods, mainly due to the computational complexity of solving (1) in three dimensions for a large number of particles, i.e., for systems relevant to the materials designer, with a gap of $\sim 10^{20}$!

There are numerous approaches to approximate solutions for (1), most of which involve finding the system's total ground state energy, E , including methods that treat the many-body wavefunction as an antisymmetric function of one-body orbitals (discussed in later sections), or methods that allow a direct representation of many-body effects in the wave function such as Quantum Monte Carlo (QMC), or hybrid methods such as coupled cluster (CC), which adds multi-electron wavefunction corrections to account for the many-body (electron) correlations.

QMC can, in principle, provide energies to within chemical accuracy (≈ 2 kcal/mol) [4] and its computational expense scales with system size as $O(N^3)$ or better [5, 6], albeit with a large prefactor, while CC tends to scale inefficiently with the size of the system, generally $O(N^6 \text{ to } N!)$ [7].

Nevertheless, we have shown how QMC performance can be significantly improved using short equilibration schemes that effectively avoid configurations that are not representative of the desired density [8], and through efficient data parallelization schemes amenable to GPU processing [9]. Furthermore, in [10] we also showed how QMC can be used to obtain high quality energy differences, from generalized valence bond (GVB) wave functions, for an intuitive approach to capturing the important sources of static electronic correlation. Part of our current drive involves using the enhanced QMC methods to obtain improved functionals for Density Functional Theory (DFT) calculations, in order to enhance the scalability and quality of solutions to (1).

But for the sake of brevity, we will focus here on methods and applications that are unique for integrating multiple paradigms and spanning multiple length- and timescales, while retaining chemical accuracy, i.e., beyond direct use of conventional QM approaches. The following section describes the general path to classical approximations to (1), in particular to interatomic force fields and conventional MD, which sacrifice electronic contributions that drive critical chemical properties, and our departure from conventionalism to recover the missing physicochemical details.

2.2 Approximations to Schrödinger’s Equation

A number of simplifications to Schrödinger’s equation are commonly made to ease the computational costs; some of these are reviewed below in order to explain the nature of our methods.

2.2.1 Adiabatic Approximation (Treat Electrons Separately from the Nuclei)

An important approximation is to factor the total wavefunction in terms of an electronic wavefunction, which depends parametrically on the stationary nuclear positions, and a nuclear wavefunction, as

$$\psi_{\text{total}} = \psi^{\text{electronic}} \times \psi^{\text{nuclear}}. \quad (3)$$

This is also known as the Born–Oppenheimer [11] approximation. The underlying assumption is that since nuclei are much heavier than electrons (e.g., the proton to electron mass ratio is ~ 1836.153), they will also move in a much lower timescale. For a set of fixed nuclear positions, (1) is used to solve for the corresponding electronic wavefunction and electronic energies (typically in their lowest or ground-state). A sufficient set of electronic solutions, at different nuclear positions, leads to the systems’ nuclei-only dependent Potential Energy Surface (PES). Modern codes can also lead directly to the inter-atomic forces, from the negative gradient of the potential energies, required for understanding the dynamics of systems.

Methods for solving the electronic equation (1) have evolved into sophisticated codes that incorporate a hierarchy of approximations that can be used as “black boxes” to achieve accurate descriptions for the PES for ground states of molecular systems. Popular codes include Gaussian [12], GAMESS [13], and Jaguar [14] for finite molecules and VASP [15], CRYSTAL [16], CASTEP [17], and Sequest [18] for periodic systems.

The simplest wavefunction involves a product of one-particle functions, or spin-orbitals, antisymmetrized to form a (Slater) determinant that satisfies the Pauli (exclusion) principle, i.e., two electrons with the same spin orbital result in no wavefunction. Optimizing these spin-dependent orbitals leads to the Hartree–Fock (HF) method, with the optimum orbitals described as molecular orbitals (MO). HF is excellent for ground state geometries and good for vibrational frequencies, but its neglect of electron correlation [19] leads to problems in describing bond breaking and chemical reactions. In addition, it cannot account for the London dispersion forces responsible for van der Waals attraction of molecular complexes. A hierarchy of methods has been developed to improve the accuracy of HF. Some of the popular methods include second-order Moller–Plesset perturbation theory (MP2) [20], CC with multiple perturbative excitations, multireference selfconsistent field (MC-SCF), and multireference configuration interaction (MR-CI) [21] methods

(see [22] for a recent review). A form of MC-SCF useful for interpreting electron correlation and bonding is the GVB method, [23–25] which leads to the best description in which every orbital is optimized for a single electron. These are referred to as *ab initio* methods as they are based directly on solving (1), without any empirical data. Many methods, which rely on empirical data to obtain approximate descriptions for systems too large for *ab initio* methods, have also been proved useful. [26]

A non-empirical alternative to *ab initio* methods that now provides the best compromise between accuracy and cost for solving Schrödinger’s equation of large molecules is DFT. The original concept was the demonstration by Hohenberg and Kohn [27] that the ground state properties of a many-electron system are uniquely determined by the density, ρ , as a function of nuclear coordinates, r , and hence all the properties of a (molecular) system can be deduced from a functional of $\rho(r)$, i.e.,

$$E = \varepsilon[\rho(r)]. \quad (4)$$

DFT has evolved dramatically over the years, with key innovations including the formulation of the Kohn–Sham equations [28] to develop a practical one-particle approach, while imposing the Pauli principle, the Local Density Approximation (LDA) based on the exact solution of the correlation energy of the uniform electron gas, the generalized gradient approximation (GGA) to correct for the gradients in the density for real molecules, incorporating exact exchange into the DFT. This has led to methods such as B3LYP and X3LYP that provide accurate energies (~ 3 kcal/mol) and geometries [29] for solids, liquids, and large molecules [30, 31]. Although generally providing high accuracy, there is no prescription for improving DFT when it occasionally leads to large errors. Even so, it remains the method of choice for electronic structure calculations in chemistry and solid-state physics. We recently demonstrated improved accuracy in DFT by introducing a universal damping function to correct empirically the lack of dispersion [32].

An important area of application for QM methods has been determining and describing reaction pathways, energetics, and transition states for reaction processes between small species. QM-derived first and second derivatives of energy calculated at stable and saddle points on PES can be used under statistical mechanics formulations [33, 34] to yield enthalpies and free energies of structures in order to determine their reactivity. Transition state theory and idealized thermodynamic relationships (e.g., $\Delta G[P_0 \rightarrow P] = kT \ln[P/P_0]$) allow temperature and pressure regimes to be spanned when addressing simple gas phase and gas-surface interactions.

On the other hand, many applications involve interactions between solutes and solvent, which utterly distinguish the condensed phase from *in vacuo*, free energy surfaces. To tackle this challenge, we describe below a unique multiparadigm strategy to incorporate the effects of a solvent when using QM methods to determine reactivity in organic and organometallic systems.

Application Example: Solvent and pH Effects on Reactivity

Interactions critical to the rate and selectivity of reactions include the relaxation of a wavefunction or zwitter-ionic geometry in response to a polarizable solvent, hydrogen bonding, and reversible proton transfer. It is necessary in these cases to introduce solvation effects explicitly through the inclusion of solvent molecules, and/or implicitly through a continuum representation of the medium. Adding explicit solvent molecules increases the cost of already expensive QM calculations, while implicit solvation models vary in their degree of parameterization and generality.

One approach assigns an empirical surface free energy to each exposed atom or functional group in a solute. More general algorithms combine an electrostatic term based on atomic charges and solvent dielectric constant with empirical corrections specific to functional groups and solvent cavitation energies. In the Poisson–Boltzmann (PB) model [25], solvent is represented as a polarizable continuum (with dielectric ϵ) surrounding the solute at an interface constructed by combining atomic van der Waal radii with the effective probe radius of the solvent. Charges are allowed to develop on this interface according to the electrostatic potential of the solute and ϵ through the solution of the Poisson–Boltzmann equation. Charges representing the polarized solvent are then included in the QM Hamiltonian, such that the wavefunction of the complex is relaxed self-consistently with the solvent charges via iterative solution of the PB and Schrödinger equations. Implicit models offer the advantage over explicit solvation that degrees of freedom corresponding to solvent motion are thermally averaged; thus the number of particles in a QM simulation (which typically scales as N^3 or worse) is not significantly increased.

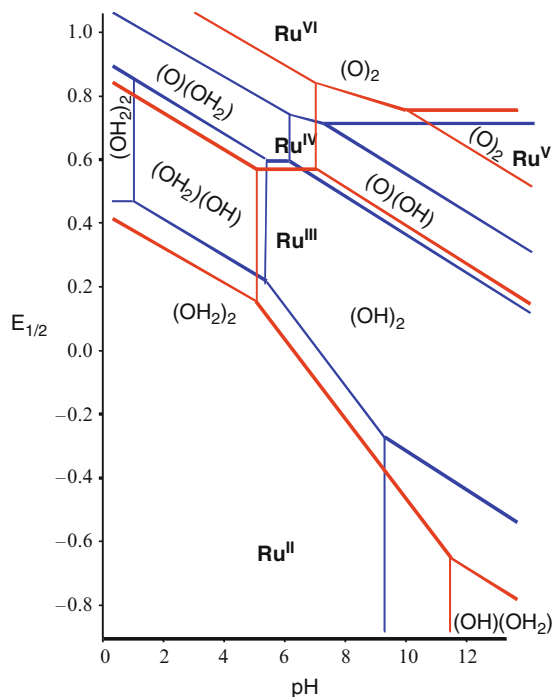
In spite of the success of implicit solvation models, it is often easier and more precise to take advantage of the tabulated free energies of solvation of small, common species such as proton, hydroxide, halide ions, and so on [35, 36]. To screen new potential homogeneous catalysts for favorable kinetics and elucidate mechanisms of existing systems, we have typically employed the following expression for free energies of species in solution:

$$G = E_{\text{elec}} + ZPE + H_{\text{vib}} - TS_{\text{vib}} + G_{\text{solv}}, \quad (5)$$

which includes an electronic energy, E_{elec} , a temperature-dependent enthalpy, TS_{vib} , entropy contributions, H_{vib} , the zero-point-energy, ZPE, and a solvation free energy, G_{solv} , provided by a PB continuum description [14].

An example of fundamental transformations that cannot be modeled without accurate accounting of changes in electronic structure (on the order of 100 kcal/mol), solvation of multiply charged species (~ 100 kcal/mol), and the macroscopic concentration of protons (~ 10 kcal/mol) is the pH-dependent oxidation of acidic metal complexes. Figure 2 compares experimentally determined pKas and oxidation potentials [33] of *trans*-(bpy)₂Ru(OH₂)₂²⁺ to values computed with (5). Maximum errors are 200 mV and 2 pH units, despite the large changes in the components

Fig. 2 Pourbaix diagram for $\text{trans-(bpy)}_2\text{Ru(OH}_2)_2^{2+}$ showing pKas (vertical lines) and oxidation potentials (bold lines) determined by cyclic voltammetry [33] and our calculated pKas and oxidation potentials using DFT(MO6) and Poisson–Boltzmann continuum solvation (red). $\text{Ru}^{\text{III}}/(\text{OH})_2$ denotes, for example, the region of stability of $\text{trans-(bpy)}_2\text{Ru}^{\text{III}}(\text{OH})_2^{1+}$



of free energy. The changes in free energy associated with redox processes determine the driving force behind many catalytic cycles. Coupled with the energies of transition states between intermediates, these tools allow *predictive* work in applications of homogeneous catalysis to problems in synthetic and energy-related reactions. Given that spin–orbit coupling corrections are important for open-shell wavefunctions of heavy elements and have been computed to useful accuracy [37], such corrections may be incorporated into (5).

Having described a hybrid approach that integrates a first-level QM-DFT approximation with a continuum-level implicit APBS solvation model, as a multiparadigm stratagem to study the effects of solvation on reactivity, we now return to describing further approximations to (1).

2.2.2 Treat the Nuclei as Classical Particles Moving on a PES

The PES found via the adiabatic approximation described in the previous section portrays the hyper landscape over which a nucleus moves, in the classical sense, while under the influence of other nuclei of a particular system. This is useful for describing vibrations or reactions. Electronic contributions have been averaged into each point on the PES, and their effect considered for that particular nuclear

conformation; therefore one might consider replacing (1) by Newton's ordinary differential equation of motion, i.e.,

$$F = -\frac{\partial V}{\partial R} = m \frac{d^2 R}{dt^2}, \quad (6)$$

where F represents the forces (obtained from the negative gradient of the PES with respect to nuclear positions) and m the corresponding atomic mass. Integrating (6) with respect to time leads to particle trajectories, and this is conventionally referred to as MD. Since only nuclei motions are considered, all information about the electrons is gone (e.g., quantum effects like electron tunneling, excited electronic states, and so on). Such calculations in which the forces come directly from a QM computed PES are often referred to as Car–Parrinello calculations [38]. Unfortunately, the costs of QM-MD limit such calculations to $\sim 1,000$ atoms, and at best < 1 ps, so an additional simplification is to find an alternative mean to compute the PES. This is discussed next.

2.2.3 Approximate the PES with Inexpensive Analytical Forms: Force Fields

A practical solution for large systems, requiring long-term dynamics, is to describe the PES, U , in terms of a force field (FF), a superposed set of analytic functions describing the potential energy between the interacting particles (and its negative gradient, corresponding to the inter-atomic forces, F) as a function of atomic (nuclear) coordinates (x):

$$F = m_i \ddot{x}_i = -\nabla_i U(x_1, x_2, \dots, x_n), \quad (7)$$

where U is conventionally portioned in terms of valence, or bond functions, and non-bond functions, as follows:

$$U = [U_r + U_\theta + U_\varphi + U_\psi]_{\text{bond}} + [U_{\text{vdW}} + U_{\text{Coulomb}}]_{\text{non-bond}}. \quad (8)$$

Integrating (7) with respect to time, leads to a description of nuclear trajectories as a function of time.

U can take numerous forms, and since it is the key element affecting the accuracy and transferability of a force field we discuss this further below, but first a few words about the validity of the classical approximations to (1) discussed thus far.

When the thermal de Broglie wavelength is much smaller than the interparticle distance, a system can be considered to be a classical or Maxwell–Boltzmann gas (the thermal de Broglie wavelength is roughly the average de Broglie wavelength of the particles in an ideal gas at the specified temperature). On the other hand, when

the thermal de Broglie wavelength is on the order of, or larger than, the interparticle distance, quantum effects will dominate and the gas must be treated as a Fermi gas or a Bose gas, depending on the nature of the gas particles; in such a case the classical approximations discussed are unsuitable. Their use is also not recommended for very light systems such as H_2 , He, Ne, or systems with vibrational frequencies $h\nu > K_{\text{B}}T$, systems in extreme conditions of temperature and pressure, with high energy or a large number of excited electronic states, nor for systems with two different electronic states but close nuclear energy (i.e., different ψ^n).

3 From QM to Molecular Mechanics/Dynamics: Force Fields

As mentioned previously, the definition of an empirical potential establishes its physical accuracy; those most commonly used in chemistry embody a classical treatment of pairwise particle–particle and n-body bonded interactions that can reproduce structural and conformational changes. Potentials are useful for studying the molecular mechanics (MM), e.g., structure optimization, or dynamics (MD) of systems whereby, from the ergodic hypothesis from statistical mechanics, the statistical ensemble averages (or expectation values) are taken to be equal to time averages of the system being integrated via (7).

In the following sections, we outline our first-principles-based Dreiding [39] potential, to exemplify regular force fields, which usually cannot reproduce chemical reactions, and follow up with an introduction to two of our unique force field approaches, which overcome most of the limitations in the conventional approach. In each case, we present unique applications to demonstrate their usefulness.

3.1 Conventional Force Fields

Traditionally, the bonded components are treated harmonically (see expressions in Fig. 3). There are generally two non-valence or non-bonded terms: the van der Waals term (U_{vdW}) which accounts for short-range repulsion, arising from the Pauli Principle and interacting dipoles, and for long range attractions arising from the weak London dispersions, expressed generally as

$$U_{\text{vdW}} = \sum_{\substack{R_{ij} > R_{\text{cut}} \\ [\text{excl}(1-2, 1-3)]}} \hat{U}_{\text{vdW}}(R) \cdot S(R_{ij}, R_{\text{on}}, R_{\text{off}}), \quad (9)$$

where U_{vdW} can represent different forms, and the electrostatic or Coulomb interactions, which account for the charged interactions between every ij pair of atoms

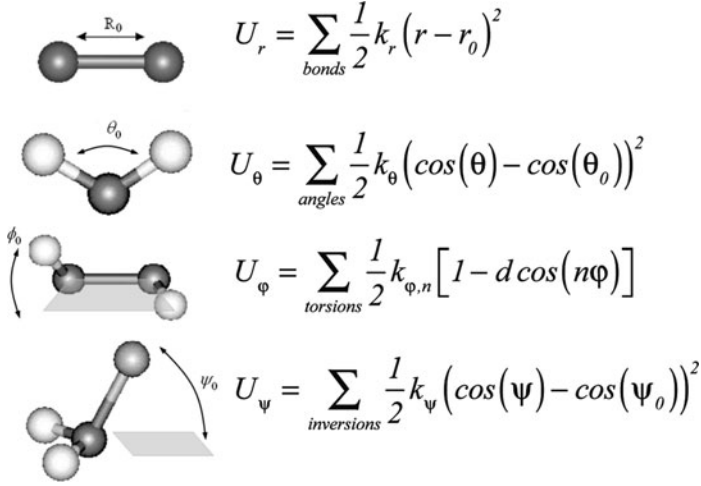


Fig. 3 Conventional (Dreiding) valence interatomic potentials. Sub-indices 0 indicate equilibrium values, k constants are related to force constants for vibrational frequencies, c constants are related to an energy barriers, and n refers to periodicity

flowing within a dielectric medium ($\epsilon = 1$ in a vacuum but larger values are used for various media), expressed conventionally as

$$U_{\text{Coulomb}} = C_0 \sum_{i>j} \frac{Q_i Q_j}{\epsilon R_{ij}} S(R_{ij}, R_{\text{on}}, R_{\text{off}}), \quad (10)$$

where C_0 corresponds to a unit conversion scalar (e.g., for energy in kcal/mol, distances in Å, and charge in electron units, $C_0 = 332.0637$), Q_{ij} to the pairwise point charges, R_{ij} to the interparticle distance, and S to a cutoff function.

One additional term included in Dreiding accounts for weak hydrogen bonded interactions, as a mixture of 3-body angles (between an H atom, and H donor and acceptor atoms) and non-bonded terms (between donor and acceptor atoms), and is given by

$$E_{\text{HB}}(R, q_{\text{AHD}}) = E_{\text{b}}(R) E_{\text{a}}(\cos(q_{\text{AHD}})). \quad (11)$$

The most time-consuming aspect of MD simulations for large systems corresponds to the calculation of long-range non-bond interactions, (7) and (8), which decrease slowly with R . This scales as $O(N^2)$ for an N particle system (e.g., a protein with 600 residues would have ~6,000 atoms requiring ~18 million terms to be evaluated every time step). One way to reduce this cost is to allow the long-range terms to be cut off smoothly after a threshold value (S function in (9) and (10)). Alternatively, our Cell Multipole Method (CMM) [40] (and the Reduced CMM [41]) enable linear scaling, reducing the computational cost while retaining accuracy over large-scale systems.

Many useful FF have been developed over the last 30 years ([42] provides a recent review), a significant number of which are aimed at biological systems. Commonly used FF include AMBER [43], CHARMM [44], Dreiding [39], and OPLS [45]. Most of the parameters in these FF were adjusted to fit a combination of results from theory and experiments.

A key strategy in our multiscale approach has been to parameterize force fields (e.g., non-reactive Dreiding) from QM calculations on small representative systems, adjusting the FF descriptions to reproduce the structures, energetics, and dynamics from QM on nanoscale systems. This favors transferability and the predictive capability, in particular for systems with little or no existing empirical data.

With an FF it is practical to apply MD simulations to the atomic-level dynamics of large-scale systems (e.g., proteins [46]) interacting with other nanoscale components or external fields under complex conditions. Force fields allow one to carry MD simulations on systems $\sim 10^{6-9}$ times larger than for QM. It is no surprise then that a particularly well-suited niche for the application of force fields is in the prediction of protein structures, in particular for membrane proteins that are otherwise impossible to crystallize in their active form using existing experimental methods. One class of membrane proteins with significant relevance is that of G-protein coupled receptors, mainly because they are involved in many diseases. This fact makes them a target of approximately 30–50% of all modern medicinal drugs. GPCRs are seven-transmembrane protein domain receptors, that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. GPCRs are found only in eukaryotes, choanoflagellates, and animals. The ligands that bind and activate these receptors include light-sensitive compounds, odors, pheromones, hormones, and neurotransmitters, and vary in size from small molecules to peptides to large proteins.

The following section describes our multiscale, multiparadigm modeling approach to protein structure prediction, and in particular to GPCRs.

3.1.1 Application Example: Structure Prediction of GPCRs

The activation related conformational changes in G protein-coupled receptors (GPCRs) allow cells to sense their environment and convert extracellular signals (e.g., ligand binding) into intracellular signals (through G protein and β arrestin pathways), leading to physiological responses. They are activated by a variety of molecules (including biogenic amines, peptides, lipids, nucleotides, hormones, proteins, tastants, odorants, among others) and non-molecular sensory signals (such as light, touch, and others), and thus play an important role in all major disease areas including cardiovascular, metabolic, neurodegenerative, psychiatric, cancer, and infectious diseases. There are ~ 370 non-sensory human GPCRs (out of ~ 800 human GPCRs [47, 48] but experimental crystal structures are available only for two (human β_2 and human A_{2A}), both in the inactive form. We can expect additional structures for human GPCRs to become available slowly over the next

few years, but most will also be in the inactive form. This lack of structures is contributing to the shortage of safe and efficacious drugs that target GPCRs. Availability of GPCR structures, experimental or predicted, can lay the foundation for rational structure-based drug design. So, a fast but accurate computational approach is needed that can generate structures for all important conformations of a target receptor and any other receptors implicated for off-target therapeutic side-effects and determine their ligand binding efficacies for developing highly selective drug candidates with potentially minimal side-effects. We have developed one such FF based method.

The structural topology of GPCRs consists of seven transmembrane (TM) α -helices that span the membrane and are connected by both intracellular and extracellular loops. To characterize this topology quantitatively with respect to a common reference frame, the middle of the membrane is assumed to correspond to the $z = 0$ plane or the hydrophobic plane that cuts the 7-helix bundle into two halves. Each GPCR structure can then be characterized by the six orientation parameters of the seven helices shown in Fig. 4, which shows how the helix position and tilt are defined. Helix position on the hydrophobic plane is then given by x and y . Value h corresponds to the hydrophobic center residue from the helix that will be positioned on the hydrophobic plane. Two angles, θ and ϕ , specify the tilt angles of the helix and the angle η corresponds to the helix rotation angle about its axis. The two tilt angles (θ , ϕ) and the rotation angle (η) require a definition of the helical axis which needs to account for the reality of bent helices as prolines are commonly found in the TM helices. We use a helical axis that corresponds to the least moment of inertia vector for the helix obtained by eigensolution of the moment of inertia matrix for the helix using only heavy backbone atoms.

The structural analysis of available experimental structures shows large variations in helix tilts and rotations. Considering a $\pm 10^\circ$ sampling of the θ tilt angle, and $\pm 30^\circ$ sampling of the ϕ and η angles, for each of the seven helices, leads to $(3 \times 5 \times 5)^7$, ~ 10 trillion possible conformations, for each of which the amino acid side chains must be optimized. To make such a huge sampling computationally feasible, we developed the SuperBiHelix sampling (BiHelix sampling only sampled the η rotation angle) procedure. As indicated by 12 double arrows in Fig. 5 (left) a typical class A GPCR template has 12 important pair-wise interactions. For each such pair of helices, we will sample all combinations of θ , ϕ , η over some grid. During this sampling, the other five helices are ignored, as indicated in Fig. 5

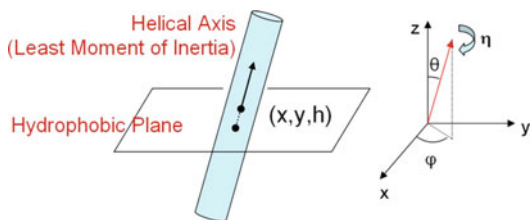


Fig. 4 Definitions of orientation parameters of a transmembrane helix

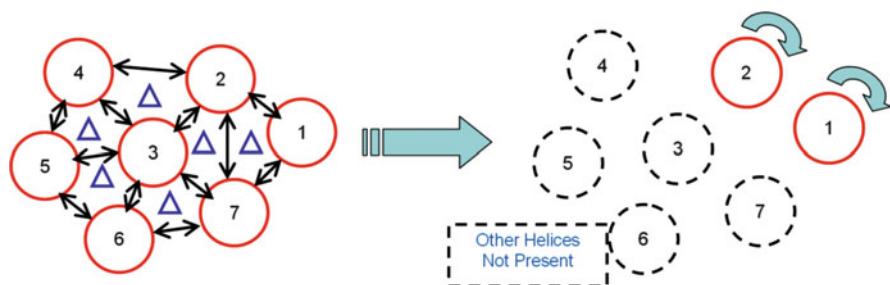


Fig. 5 *Left:* The 12 important helix pair interactions. *Right:* The BiHelix concept in which the interhelical interactions are optimized for a pair of helix rotations, while ignoring the other 5

(right) for helix 1–2 pair. For each of the $(3 \times 5 \times 5)^2 = 5,625$ rotational-tilt combinations of each of the 12 pairs, we optimize the side chains using SCREAM [49] with all-atom DREIDING force field [39]. SCREAM uses a library of residue conformations ranging from a CRMS diversity of 0.4–1.6 Å in conjunction with a Monte Carlo sampling using full valence, hydrogen bond, and electrostatic interactions from D3FF, but with a special flat bottom van der Waals (vdW) potential that reduces the penalty for contacts that are slightly too short while retaining the normal attractive interactions at full strength.

The total energies for each of the $(12) \times (5,625)$ helix pair combinations are used to estimate the energy for all 10 trillion 7-helix bundle conformational combinations. In a procedure called SuperComBiHelix, the top 2,000 of these helical bundles are explicitly built and the side chains reassigned, given that they will take different conformations compared to those in the bihelical model. Then the structure is minimized for ten steps. The energy ranking will be different in SuperComBiHelix than SuperBiHelix because all seven helices are present instead of just two at a time. This procedure results in an ensemble of low-lying structures. Examination of the low-lying structures shows the helix packing preferred by the receptor, which should also include conformations relevant for understanding function and activation of these proteins. This procedure was applied to A_{2A} receptor where its helices were placed in the β_2 template. The starting structure was 2.0 Å from the A_{2A} crystal structure (see Fig. 6). After the SuperBiHelix/SuperComBiHelix optimization (sampling ~ 10 trillion conformations), the lowest energy structure was 1.3 Å from the A_{2A} crystal structure (Fig. 6), an improvement that was also critical in the identification of correct ligand binding mode in the protein. The procedure has been applied to all available crystallized receptors (humBeta2, turBeta1, bovRhod, humA2A). For each of these systems we first sampled only the helix rotation angle η over full 360° range in increments of 15° , which needs the sampling of 24^7 (~ 4 billion) conformations. This resulted in crystal structure ranking number 1 at the end of the procedure for all cases. The results for turBeta1 are shown in Table 1, which shows the crystal structure as ranked number 1 and other near-native structures also ranked in the top ten list.

Fig. 6 Visual comparison of the starting crystal structure (*green*) and the predicted tertiary structure [before (*red*) and after optimization (*blue*)]

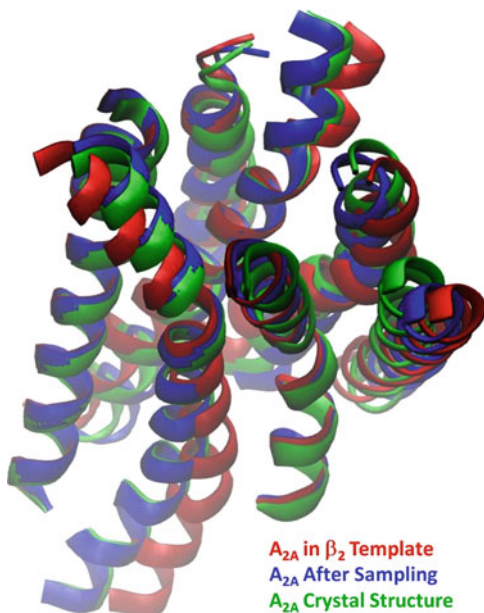


Table 1 Top ten predicted conformations for Turkey $\beta 1$ adrenergic receptor

Eta	H1	H2	H3	H4	H5	H6	H7	TotEnergy	RMSD
Eta	0	0	0	0	0	0	0	-385	0.0
Eta	0	0	0	345	0	0	0	-379	0.4
Eta	0	0	0	0	345	0	0	-377	0.4
Eta	0	0	0	0	15	0	0	-376	0.4
Eta	0	0	0	345	15	0	0	-365	0.4
Eta	0	0	345	345	15	0	0	-364	0.5
Eta	0	0	0	15	0	0	0	-349	0.4
Eta	0	0	0	15	345	0	0	-349	0.4
Eta	0	0	0	0	0	0	345	-346	0.4
Eta	75	0	0	345	0	0	0	-341	1.1

These results show that the two-helices-at-a-time sampling method is robust, but, more important, the FF-based energies used for scoring the conformations can be trusted to resolve the near-native structures of proteins.

3.2 Simulating Complex Chemical Processes with FFs

A major drawback with MD using conventional FFs is that they are unable to describe chemical reaction processes, or other electronic structure dependent processes such as electronic excitations, and as we've already discussed, QM is not

practical for systems larger than $\sim 1,000$ atoms, and timescales longer than 1 ps. We have made significant breakthroughs in recent years towards addressing these problems, through the development of the reaxFF reactive force field [50] for describing “ground-state” reaction processes, and the electron force field (eFF) [51] for describing systems with explicit electrons in their ground or excited states.

3.2.1 The ReaxFF Force Field for Studying Reactive Processes

ReaxFF [50] provides a generally valid and accurate way to capture the barriers for various chemical reaction processes (allowed and forbidden reactions) into the force fields needed for large-scale MD simulation. ReaxFF is parameterized exclusively from QM calculations, and has been shown to reproduce the energy surfaces, structures, and reaction barriers for reactive systems at nearly the accuracy of QM but at costs nearly as low as conventional FFs.

Applications of ReaxFF have been reported for a wide range of materials, including hydrocarbons [50], nitramines [52], ceramics [53], metals and metal oxides [54, 55], metal/hydrocarbon interactions [56], and metal hydrides [57]. ReaxFF has been used to simulate chemical events in many systems, including nanotube deformation and buckyball polymerization [58, 59], thermal decomposition of polymers [60], high-energy materials initiation [61, 62], crack propagation [63], reactive oxygen and hydrogen migration in fuel cells [64], and metal/metal oxides surface catalysis [65].

Salient features of reaxFF include: (a) *Environmentally dependent charge distributions on atoms*. The charges on the atoms adjust in response to the local environment allowing them to change as bonds are broken and formed and to shield the Coulomb interaction between atoms at small distances; (b) *Bond order dependent valence terms*. A general relation is provided between bond distance and bond order and between bond order and bond energy (and hence forces). The bond orders gradually go to zero as the bond lengths increase and they gradually increase for shorter distances, finally saturating for the smallest distances (e.g., BO = 3 for CC bonds). This provides a smooth description of the valence terms during chemical reactions; (c) *Non-bond or van der Waals interactions*. ReaxFF uses a simple Morse function to account for the short-range repulsion and steric interactions arising from the Pauli principle (between *every* atom pair). The long range attraction accounts for vdW attraction; (d) *No cutoffs*. All interactions change smoothly during reactions (which are allowed to occur at any time and place) so that ReaxFF can be used with general conditions of temperature and pressure; (e) *Transferable potential*. Simple FFs provide different parameters for different atomic environments (e.g., single vs double bonds, sp³ vs sp² geometries). ReaxFF eschews such description using only a single atom type for each element, which is necessary since bond orders and geometries change during reactions. This leads to good transferability of the FF; (f) It is *QM-based*. All parameters are optimized/derived directly from QM studies on a large number of reactions. This allows extensions to new materials where there may be no experimental data.

Recently, we have demonstrated the use of reaxFF to the challenging problem of elucidating the growth process of carbon nanotubes (CNTs). Understanding this process is critical for determining the control variables that lead to chiral-specific (with semiconducting or metallic electrical conductivity behavior) mass production of CNTs. These results are summarized in the following section.

Application Example: Dynamics of CNT Growth

Since their discovery in 1991 [66], CNTs have been widely studied. Researchers have proposed CNT applications to an ample set of technologies [67] including interconnects, transistors, and diodes for microelectronics [68], as well as electrochemical transducers [69], sensor components [70], field emission devices [71], and even gas sensors [72]. The mass production of uniform, well-characterized CNTs is crucial for realizing many of these applications. However, while CNT synthesis has been demonstrated for numerous catalysts, and a wide range of reaction conditions, complete product control has remained elusive [73]. Thus, multiple investigations aimed at elucidating the key mechanism or mechanisms of CNT growth are still being carried out, in the hope that a more fundamental understanding of the growth process will result in better synthetic control [74]. Experimental observations have shed some light on CNT growth mechanisms. Atomic force microscopy (AFM), scanning electron microscopy (SEM), and tunneling electron microscopy (TEM) have been used to support instances of tip and base growth mechanism in different synthesis procedures [75–77]. More recently time-resolved, high-resolution in situ TEM studies have highlighted the role of catalyst deformation in SWNT growth and provided direct experimental validation for a Yarmulke mechanism for nucleation [78, 79]. Nevertheless, these cutting edge techniques provide overarching, general descriptions rather than detailed, atomistic mechanisms for each stage of CNT synthesis.

To fill in these experimentally inaccessible details, mechanistic studies often appeal to atomistic simulations. DFT is now widely used to explore catalytic systems, and has been applied to simplistic models of CNT growth [74, 80, 81]. Nevertheless, the usefulness of DFT is hampered by stringent limitations on the number of atoms and especially the number of structural iterations that it is feasible to consider with current computer technology [82]. Tight binding (TB) methods, which use approximations (i.e., simplified integrals) to reduce the computational cost of handling electron–electron interactions explicitly, have been used in conjunction with MD simulations to study this problem [83]; however, the timescales necessary for observing the growth process are still beyond the reach of this approach – even though TB calculations are typically a couple orders of magnitude faster than DFT [83]. Monte Carlo methods have provided another popular means of “simulating” CNT growth [82, 84]. At best, however, Monte Carlo methods show a succession of possible snapshots from the growth process, leaving the mechanistic details hidden.

As reported previously [85, 86], we have developed a set of ReaxFF parameters describing hydrocarbon chemistry catalyzed by nickel and nickel carbide catalyst particles. This ReaxFF potential is capable of treating the adsorption and decomposition of both saturated and unsaturated hydrocarbon species on several different nickel surfaces. Of particular relevance for studying CNT growth is that a single set of ReaxFF parameters accurately describes carbon in all hybridization states and a variety of chemical environments. These states include sp , sp^2 , and sp^3 hybridized carbon in various hydrocarbon molecules, carbon binding at and migrating between interstitial sites in bulk nickel, and carbon bonded to nickel surfaces strongly as an adsorbed lone adatom or in a small hydrocarbon molecule, or weakly as part of a graphene layer.

While the vast majority of theoretical studies of CNT growth starts with lone carbon atoms, assuming that decomposition has already taken place, there are conditions (e.g., low temperature growth) under which decomposition is believed to be the rate-limiting step [87]. Thus we have utilized this ReaxFF force field in a reactive dynamics (RD) study of the early stages of CNT growth. In [85] we reported on the chemisorption and decomposition of various hydrocarbon species on a nickel nanoparticle. Over the course of 100 ps of RD simulations performed, we were able map out the preferred reaction pathways for the decomposition of each hydrocarbon species studied.

The synthesis of CNTs can be broken down into three or four distinct stages. The first stage is feedstock decomposition, as discussed above. Under low temperature growth conditions, experiments suggest that feedstock decomposition is the rate-limiting step [87]. Thus our analysis of hydrocarbon decomposition pathways on nickel nanoparticles shows how the selection of different hydrocarbon species for the feedstock influences the chemisorption rate, surface coverage, and extent of carbide formation during the nanotube growth process. In particular, because we find that chemisorption is the rate limiting decomposition step for saturated hydrocarbons, the selection of unsaturated hydrocarbon species, with very small chemisorption barriers, for the feedstock, is expected to improve the growth rate where feedstock decomposition is rate limiting.

Following feedstock decomposition is the carbon transport stage, in which a hydrocarbon or carbon species is either transported along the catalyst surface or else diffuses through the catalyst bulk as carbide. Because a constant supply of carbon is needed for both nucleation and growth, carbon transport likely occurs during both the nucleation and growth stages and so is most naturally treated as a part of each of these stages taken separately. It is also possible that a partially decomposed species migrates to the nucleation or growth site where it further decomposes into the activated species. In any case, experiments indicate that there are growth conditions under which surface diffusion is the rate-limiting step [88]. ReaxFF RD simulations demonstrate the formation of nickel carbide following acetylene chemisorptions and decomposition, lending plausibility to either mechanism.

It is believed that nucleation occurs when enough carbon material accumulates on the surface for the formation of surface ring structures. The ring structures develop into a graphene island on the particle which, when it becomes large enough,

lifts its center off the particle surface in the experimentally observed Yarmulke mechanism [78, 88]. Currently, ReaxFF RD simulations beginning with several hundred gas-phase acetylene molecules surrounding a nickel nanoparticle support the early stages of this picture. Initially, as acetylene chemisorbs and decomposes on the Ni nanoparticle, the C atoms formed migrate into the bulk of the catalyst and forming carbide. After a couple nanoseconds of dynamics, the chemical potential gradient reverses and carbon begins segregating to the surface, forming carbon chains. As more carbon moves to the surface, ring structures form and clump together to form larger ring structures, resulting in multi-ring structures with tens of rings formed from a couple of hundred carbon atoms (see Fig. 7). Thus, the trajectories from these RD simulations provide an atomistically detailed picture of the early stages of CNT growth.

Following nucleation is the nanotube growth stage in which carbon is added to the end of the growing nanotube. This stage likely lasts significantly longer than the previous stages, which means that ReaxFF RD simulations of the entire growth stage are probably not computationally feasible at present. Nevertheless, a couple different strategies are available for overcoming this difficulty. The first is to use an already growing nanotube as the initial structure for ReaxFF simulations, and study just a part of the growth process. As a simple model we have used ReaxFF to consider the barriers for adding small hydrocarbon species to the edge of a graphene sheet laying on a Ni(111) surface. These simulations find the lowest carbon addition barriers for C_2 hydrocarbon species, suggesting that C_2 may be the activate form of carbon responsible for CNT growth. Unconstrained ReaxFF RD on a full-scale model of a growing CNT will provide further validation for this hypothesis.

The second option for circumventing the time limitations on ReaxFF RD is the use of a kinetic Monte Carlo procedure to bypass long periods of quasi-equilibrium dynamics between reaction events using principles from statistical mechanics and

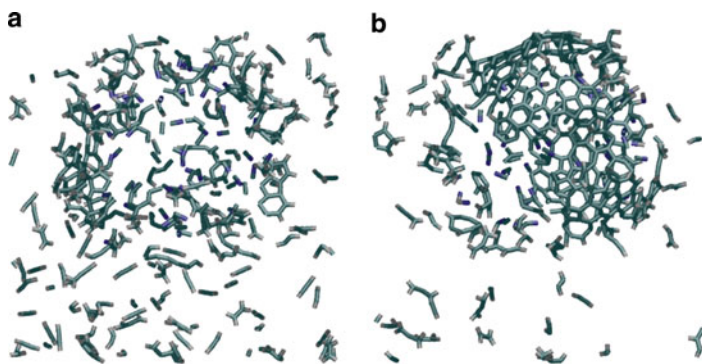


Fig. 7 ReaxFF RD simulations of acetylene adsorption and decomposition on a 468-atom nickel particle [not shown]: (a) after 1 ns a limited number of structured rings have formed and (b) after 2 ns of ReaxFF RD simulations a clear ring pattern formation appears. Simulations were performed using the parallel prototype reaxFF implementation from collaboration with H.M. Aktulga and A. Grama at Purdue, and A.C.T van Duin at Penn State

transition state theory. Because traditional kinetic Monte Carlo methods require predefined reactions and make the lattice approximation, they are not directly applicable to a complex process such as CNT growth. Nevertheless, alternative schemes have been proposed for circumventing the lattice approximation by calculating reaction barriers on the fly [89]. The bond order/bond distance relationship already present in ReaxFF would provide a natural tool for the development of an automated reaction search procedure, enabling kinetic Monte Carlo simulations within the ReaxFF framework. Such simulations would be capable of looking at CNT growth over a significantly longer timescale than ReaxFF RD.

As effective as reaxFF is for handling reactive systems and processes in their ground-state, it is unable to describe the dynamics of electrons and systems with excited electronic states. QM-MD is also limited mostly to ground-state dynamics or to a very small number of excited electronic states (see [90] for further discussion on this). The following section presents our progress in addressing this problem with a mixed quantum-classical force field method, the eFF.

3.2.2 Non-Adiabatic Excited Electronic State Dynamics with an FF

A significant number of processes involve excited electronic states, whose character – and representation in a theoretical method – depends strongly on the degree of excitation involved. Low-level electron excitations of molecules can initiate radical reactions, isomerize bonds, and induce transfers of electrons. Such processes can be studied effectively using conventional QM, using a wavefunction formed from Hartree–Fock or Kohn–Sham orbitals. At the other extreme, high-level excitations result in the formation of a weakly-coupled plasma, where bonding and chemistry vanishes, and electrons act as point particles interacting with nuclei via classical electrostatics. Such systems can be studied using classical plasma simulations techniques, i.e., particle-in-cell codes.

However, in between low and high extremes of electron excitation lies a rich variety of phenomena where the electrons are far removed from the ground state, yet remain strongly coupled to the nuclei, so that remnants of bonding and chemistry persist.

Understanding the properties of warm dense matter present in moderately excited systems is of crucial importance to developing a range of new technological enterprises. For example, in inertial confinement fusion, liquid deuterium is compressed by a shock wave, causing molecules to dissociate into atoms, atoms to ionize into plasma, and metallic conducting phases to form. Knowledge of how these phases interact could contribute to the design of improved fuel pellets.

Other examples come from the semiconductor industry, where electron beams are used to etch ultra-fine features (<35 nm) into silicon, the nuclear industry, where the interior of reactors must be protected from the passage of fast charged particles, and the biological community, where synchrotron radiation could enable single molecule X-ray diffraction, if the dynamics of highly excited and ionized biomolecules could be understood. In the above cases, theory could play a critical

role by elucidating the fundamental properties of chemical bonds and relating it to the performance of materials under extreme conditions.

To study such systems, our group has developed the eFF [59, 90, 91] approximation to QM, which can simulate moderate excitations (tens to hundreds of electron volts) that vary sharply over space and time, in large systems (tens of thousands of atoms) where strong couplings between nuclei and electrons exist, and chemistry occurs. In eFF, electrons are represented by wave packets, and nuclei by classical point charges moving in the time-varying field of the electrons (Ehrenfest dynamics). The overall electronic wavefunction is represented by a Hartree product of spin orbitals, where each orbital is a single Gaussian wave packet with size (s) and position (x):

$$\Psi(r) \propto \prod_i \exp \left[- \left(\frac{1}{s_i^2} - \frac{2p_{s,i}}{s_i} \right) (r - x_i)^2 \right] \cdot \exp[ip_{x,i}r], \quad (12)$$

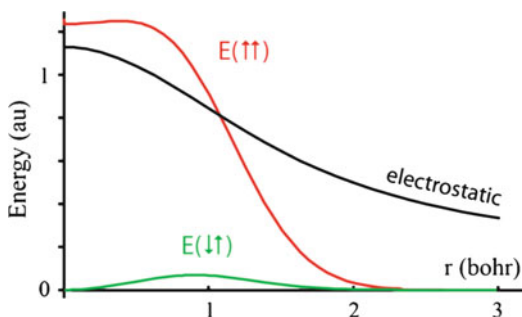
where p_x and p_s correspond to the conjugate translational and radial momenta, respectively, which represent the translational motion and radial expansion/shrinking of the wave packet over time. To account for the orbital orthogonality imposed by the Pauli principle, a spin-dependent Pauli repulsion potential is included in the Hamiltonian, which increases with the overlap between the wave packets (see Fig. 8).

The eFF method is similar to previous wave packet molecular dynamics (WPMD) approaches, but differs in the form of the Pauli potential, which we derived to account for the effect of pairwise orthogonalization on the kinetic energy of orbitals; with this change, the scope and accuracy of previous approaches is greatly extended. The only other terms present in eFF are interactions between charge distributions from classical electrostatics, and a kinetic energy term for the electron wave packets derived from QM, which provides the “kinetic energy pressure” that prevents electrons from collapsing to a point:

$$E_{\text{KE}} = \frac{\hbar^2}{m_e} \sum_i \frac{3}{2s_i^2}, \quad (13)$$

where m_e corresponds to the electron mass. From these simple terms, a rich array of chemical phenomena emerges – separation of core and valence electrons into

Fig. 8 Pauli repulsion between two electrons with size $s = 1$ bohr, as a function of their separation, r , and spin. These curves are described with three universal parameters adjusted to give reasonable structures and energies for CH_4 , C_2H_6 , LiH , and B_2H_6



separate shells, atomic hybridization, covalent, ionic, multicenter, and/or metallic bonds, and steric repulsions between bonds.

Since the interactions between particles in eFF are simply pairwise forces, the overall method is extremely fast and scales well computationally. We have simulated on a single processor tens of thousands of electrons, and on multiple processors millions of electrons [90].

Application Examples: Material Shock Hugoniot and Auger Decay

In one application of eFF we studied the thermodynamics of shock-compressed liquid hydrogen, characterizing molecular, atomic, plasma, and metallic phases at temperatures up to 200,000 K and compressions up to fivefold liquid density (see Fig. 9). We found reasonable agreement with data from both static compression (diamond anvil) and dynamic compression (shocks from explosives, magnetically pinched wires, lasers) experiments.

We have also demonstrated the capabilities of eFF for computing single-shock Hugoniot for lithium metal from dynamic shock wave experiments, via the shock wave and piston kinematics and initial and final densities of a 640,000-particle system (see Fig. 10). We also reported on the degree of ionization suffered by the material, a function of the explicit nuclear delocalization of electrons [90]. A simpler depiction of such dynamic shock experiments is shown in Fig. 11, wherein

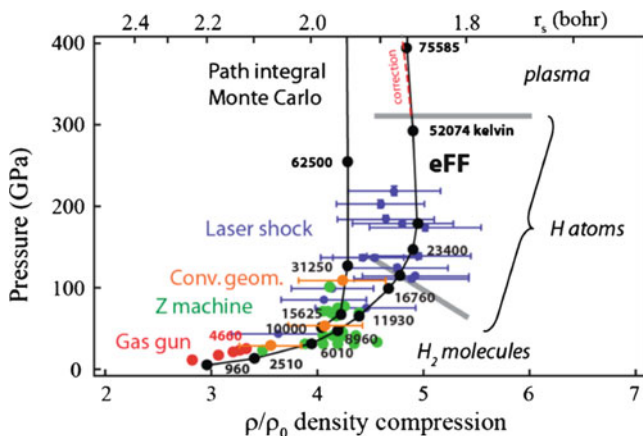


Fig. 9 Shock Hugoniot curve for liquid D2. We show here that eFF agrees well with most experiments: gas gun (red dots), Z machine (green dots), convergence geometry (orange), and the more recent laser data (blue dots) from LLNL. The PIMC results agree with eFF up to a compression of 4.2, but leads to a lower limiting compression than eFF. To compute the Hugoniot curve, we perform NVE simulations of D2, interpolating to temperatures such that the internal energy, volume, and pressure satisfy the Rankine–Hugoniot relationship. We note that the eFF Hugoniot curve connects to an eFF low temperature starting point, while the PIMC Hugoniot curve connects to a U_0 from a separate calculation

Fig. 10 Shock Hugoniot for lithium calculated directly from the planar shock velocity U_s , particle or piston velocity U_p , and initial and final densities obtained from our simulations compared to existing experiments. From [90]

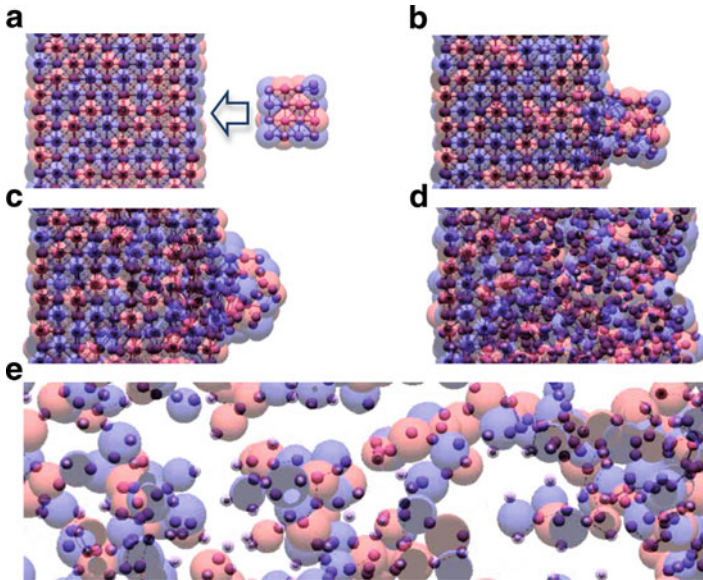
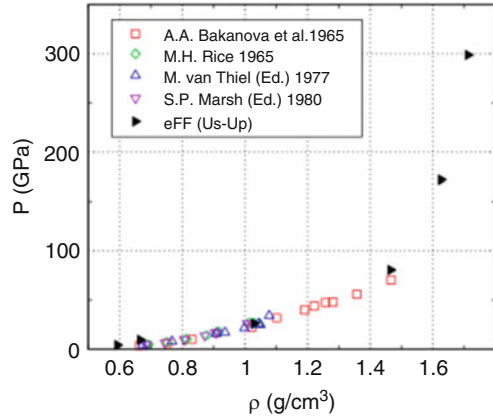


Fig. 11 Hypervelocity impact of a Li cluster on a Li metal slab. (a) Initial state. (b) Impact at $v = 2$ km/s leads to welding and no ionization. (c) Impact at $v = 5$ km/s leads to melting, and scattered ionized valence electrons. (d) Impact at $v = 10$ km/s leads to a fluid, and ~ 0.25 fraction of ionized valence electrons. (e) Impact at $v = 20$ km/s leads to a plasma with ~ 0.7 fraction of ionized electrons. *Dark small spheres* represent nuclei, *red/blue spheres* depict electron up/down spin, and their size represents the degree of de/localization

a small lithium cluster impacts onto a lithium metal slab at different speeds. Different material phases are observed as a function of impact velocity, as well as degrees of electron ionizations. Dynamic shock experiments enable higher compressions than static compression experiments.

In another application we have examined the Auger decay process in a diamond nanoparticle and in silicon, relevant to the etching of semiconductor substrates using low energy electron-enhanced etching processes [92]. We found, for the diamond case, that ionizing core electrons induced selective breaking of bonds via a variety of mechanisms, i.e., through direct excitation and ionization of valence electrons, or through indirect heating, or even in a small subset of cases, a billiard-ball like scattering away of valence electrons through ejection of neighboring bonding electrons (see Fig. 12). Our results were consistent with ion ejection data from photon-simulated desorption experiments performed on diamond films.

Our current development of eFF involves adding explicit electron exchange-correlation potentials, core pseudo-potentials, and extended support for systems with significant p- and d-character. Using eFF, we're now able to study the effect of highly excited electrons in the dynamics of material subjected to extreme conditions, including those described before, as well as other open problems in interfacial shock instabilities, radiation damage, to name a few.

As simulation requirements shift to larger length scales and longer times and system properties are amenable to homogenization in space and averaging in time, for example in characterizing the conformational behavior of supramolecular systems, coarse-grain methods tuned from finer scale ones (e.g., QM, MM) represent a suitable and more efficient alternative for evolving the dynamics of systems with reduced degrees of freedom. The following section discusses our progress in developing coarse-grain force fields and time-lower bound solutions to the resulting rigid multibody EOM.

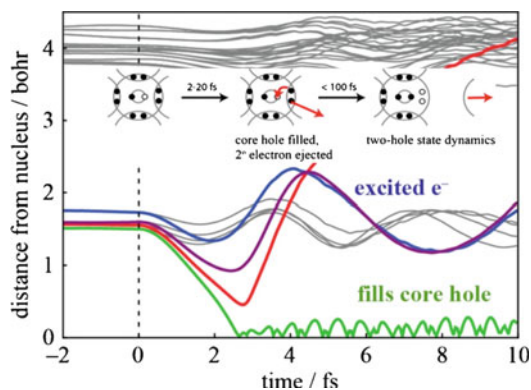


Fig. 12 Single Auger trajectory after ionization of a carbon core electron at the center of the diamond nanoparticle ($m_{\text{elec}} = m_p$). Valence electrons surrounding the core hole with the same spin as the ionized core electron are highlighted in red, green, blue, and purple. Distance of valence electrons from the core hole, showing the green electron filling the core hole, the red electron being ejected (and trapped after 20 fs, not shown), and the blue and purple electrons being excited. From [93]

4 Bridging MM/MD with the Mesoscale

Successful applications of first principles methods to supramolecular modeling requires a scale lying in between the molecular or atomistic scale (where it is convenient to describe molecules in terms of a collection of bonded atoms) and the continuum or macroscale (where it is convenient to describe systems as continuous with a finite element mesh basis) [94]. This coarse grain or meso-scale level is most important for determining the properties and performance of a wide range of different materials, including “soft condensed matter,” amphiphilic fluids, colloids and polymers, gels, liquid crystals; proteins, and DNA. An important class of problems that need to be described at these scales include biological processes such as protein activation, enzymatic transformations, ribosome activity, and general diffusive motions of biomolecules on timescales of microseconds and longer [95].

Several approaches can reduce the computational costs of solving (7) for systems with a large number of atoms, including the use of explicit constraints on fast atomic motions, bead representations which join several atoms into pseudo-particles with no rotational inertia, or representations which cluster collections of atoms into rigid bodies with inertia, among others. We will refer here to those treating clusters of atoms as rigid bodies that interact with others through net forces and torques, and which use coarse-grain force fields to solve the system’s dynamic behavior.

4.1 *Constrained and Coarse-Grain MD*

By imposing constraints on fast atomic motions, one can effectively increase the timescales of integration of the EOM, thereby enabling longer simulation times. Conventional methods for doing this on Cartesian atomistic degrees of freedom include those that compensate relative restraint forces into the particle EOM, such as SHAKE [96, 97], RATTLE [98], and the like [99–101]. Unfortunately, these methods are limited to low temperature dynamics [102] and to relatively small systems, due to the added cost of solving for the explicit constraints. To overcome this, alternative approaches have focused on simplifying the description of the system through EOM that operate only on the degrees of motion of a system, mostly using internal coordinate representations that treat clusters of atoms as rigid bodies.

Mazur et al. [103, 104] demonstrated the conformational dynamics of bio-macromolecules. However, their method scaled exponentially with size and relied on an expensive expression for the inter-atomic potentials in internal coordinates. Subsequently, our group pioneered the development of internal coordinate constrained MD methods, based on ideas initially developed by the robotics community [102, 105–107], reaching $O(n)$ serial implementations, using the Newton–Euler Inverse Mass Operator or NEIMO [108–110] and Comodyn [111] based on a variant of the Articulated Body Inertia algorithm [112], as well as a parallel implementation of $O(\log n)$ in $O(n)$ processors using the Modified Constraint Force Algorithm

or MCFA [107, 113]. These methods can selectively handle implicit constraints through appropriate projection matrices in the EOM. Nonetheless, most have focused on the torsional degrees of freedom (DOF) which affect the conformation of a system. The general state space EOM for internal coordinate constrained MD can be written as

$$\tau = M(Q)\ddot{Q} + C(Q, \dot{Q})\dot{Q}, \quad (14)$$

where τ corresponds to the vector of generalized forces (e.g., torques), M denotes the articulated body inertia matrix, C denotes the nonlinear velocity dependent terms of force (e.g., Coriolis, centrifugal and gyroscopic forces), and $Q, \dot{Q}, \ddot{Q}$ correspond to the generalized coordinates that define the state of the system. It then follows that the dynamics of motion for a microcanonical ensemble is obtained by solving for the hinge accelerations, access to increased integration time-steps, faster exchange between low- and high-frequency modes for high temperature dynamics, and faster and smoother sampling of the PES (conformational space), among others. Our rigid body MD approaches, with atomistic and coarse-grain force fields, are currently used to predict the conformational evolution of helical domains in GPCR protein bundles (see Fig. 13):

$$\ddot{Q} = M^{-1}(Q)[\tau - C(Q, \dot{Q})\dot{Q}]. \quad (15)$$

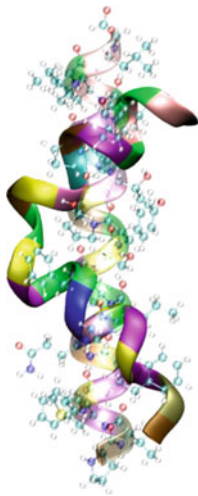


Fig. 13 Beta2 GPCR helix 7 final structure (shown in ribbons representation) after 100 ps of constrained MD-NVT, using Comodyn with a 5 fs timestep and the Dreiding force field, shows a kink about a proline amino-acid group. The kink is also observed during full-atom MD-NVT at roughly the same timescale. The original starting structure is shown in transparency, for both the full-atomistic and ribbon representations. The coarse-grain representation involved 127 clusters for a total of 133 DOF in the internal coordinate representation of the equations of motion (compared to the 1,170 DOF for the atomistic model)

These constrained MD methods rely on efficient solutions to the rigid-body EOM and on the use of atomistic, simplified atomistic (i.e., no intra-cluster valence or non-bond interactions) or coarse-grain force fields, discussed next.

Coarse grain (CG) models must carry enough information about the atomistic behavior while at the same time be efficient to scale in both time (>1 ms) and length (>100 s nm). For example, accurate models to represent solute-solvent interactions should account for solvent momentum so that its behavior can be consistent with hydrodynamics, have the correct density at the desired temperature, and be able to maintain a liquid/vapor interface over such a temperature [114].

To this end, several approaches have been used and demonstrated for phospholipids [115–123], oligosaccharides and their water mixtures [124], proteins [125], and dendrimers and polymers [126]. Coarse-grain force fields have been developed from heuristically simplified models of biomolecules (e.g., water, alkanes, lipids, etc.) and by systematic optimization procedures of a set of interaction potentials between collections of atoms treated as rigid bodies. In the former case, the fitting process relies on results from the finer atomistic MD using Monte Carlo schemes [114, 127], random search algorithms including genetic algorithms (GAs) [128, 129], and hybrid algorithms to accelerate convergence using artificial neural networks (ANNs) [128] and gradient based algorithms near local minima.

Our effort focuses on a first-principles-based strategy in order to provide not only the accuracy from finer grain calculations, but improved scalability and a seamless coupling. We exemplify this next with a coarse-grain approach for representing the dynamics of DNA.

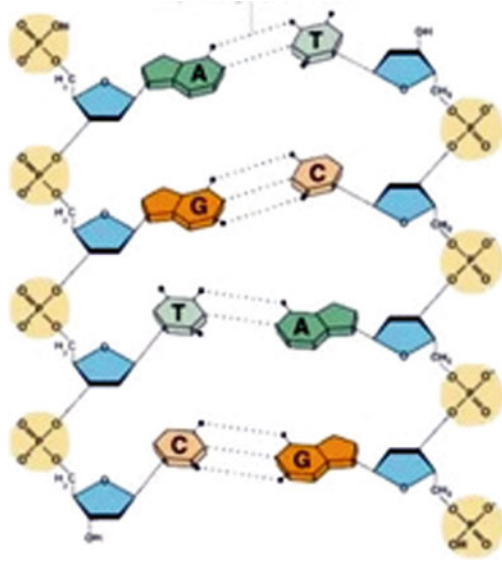
4.1.1 Application Examples: Meso-Scale Dynamics of DNA

Our meso-scale model of DNA corresponds to a backbone-based structure, where three pseudo-atoms, with two or three pseudo-beads, represent each nucleotide. For each nucleotide, one pseudo-atom represents the phosphate group and another represents the sugar group of the sugar-phosphate backbone. There is then a pseudo-atom representing each of the different nucleosides (adenine, cytosine, guanine and thymine). Additionally, guanine and cytosine have three mass-less pseudo-beads, representing the three hydrogen bonding sites. Similarly, thymine and adenine have two pseudo-beads (see Fig. 14). These pseudo-beads move as rigid bodies with their parent pseudo-atom. This removes the need to define bond and angle terms for them.

Statistics of the potential of mean force (between sets of atoms representing each bead) is computed from various full atomistic DNA simulations. These distributions are then fitted to standard potentials in order to obtain the interaction parameters for every bead. The resulting force-field uniquely describes the interaction of each bead with every other bead in system, whose total energy is given by

$$E_{\text{Total}} = E_{\text{Bonds}} + E_{\text{Angles}} + E_{\text{Torsions}} + E_{\text{nonBond}} + E_{\text{hBond}}, \quad (16)$$

Fig. 14 Meso-scale model of DNA



where E_{Bonds} , E_{Angles} , and E_{Torsions} take the same form as those in Fig. 3, while the E_{nonBond} corresponds to a Morse type potential:

$$E_{\text{nonBond}} = D_0 \left[e^{-\alpha(-1+r_{ij}/r_0)} - 2e^{-0.5\alpha(-1+r_{ij}/r_0)} \right], \quad (17)$$

and the E_{hBond} :

$$E_{\text{hBond}} = D_0 \left[5 \left(\frac{r_0}{r} \right)^{12} - 6 \left(\frac{r_0}{r} \right)^6 \right] \cos^4 \theta, \quad (18)$$

where θ corresponds to the angle between a hydrogen and the corresponding donor–acceptor pair.

Coarse-Grain Parameter Optimization

The snapshots used when computing the nearest neighbor parameters are analyzed over 1,049 structures. For each snapshot, atoms comprising each pseudo-atom were determined. The center of mass of these atoms was then computed and the corresponding pseudo-atoms were placed at this center of mass.

Valence Parameters

Each snapshot was iteratively traversed and the two body bond lengths, three body angles, and four body dihedrals calculated from the positions of the pseudo-atoms.

These were tabulated and histograms of the distribution of these valence terms, for each unique combination of pseudo-atoms, were generated (see Fig. 15a). These distributions were shown to be a single-model Gaussian distribution for the bonds and angles, and multi-modal for dihedrals. The only exception occurred when considering the bond distribution between the phosphate (PHO) and sugar (SUG) pseudo-atoms, which was bi-modal. This anomaly is explained by considering that starting at the 5' end, the SUG $\rightarrow$ PHO bond length (4.4 Å) is longer than the next PHO $\rightarrow$ SUG bond length (4.0 Å). When constructing the meso-model from the atomistic representation, if the backbone atoms are alternatively labeled as PHO-SUG-PHP-SUS-PHO, this anomaly is effectively resolved and the PHO-SUG, PHO-SUS, PHP-SUG, and PHP-SUS distributions are uni-modal.

The parameters were obtained by performing a least squares fitting of a Gaussian curve of the desired functional form. The fitting function is an exponential with exponent equal to the relevant potential/ $2RT$. For the bonds, the fitting function was therefore $y = a_0 e^{-0.5\kappa(r-r_0)^2/2RT}$. Table 2 gives the parameters obtained for the various bond-stretch terms.

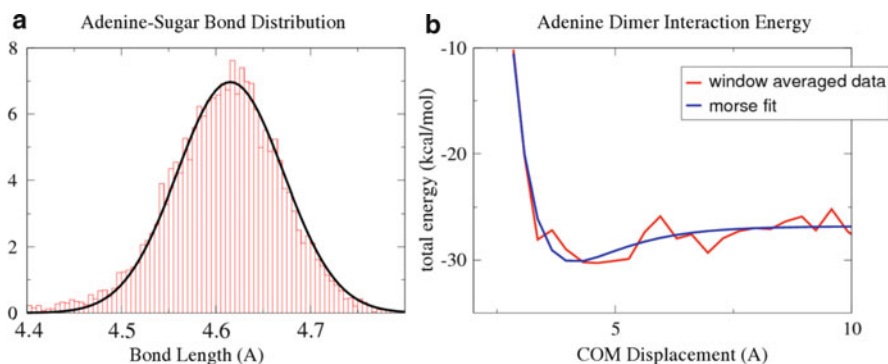


Fig. 15 Plots of (a) adenine–sugar bond distribution from atomistic MD and (b) adenine dimer interaction energy. This includes the binding energy of the dimer as well as the interaction with the water molecules

Table 2 Bond parameters for pseudo-beads

	k_0 (kcal mol ⁻¹ Å ⁻²)	r_0 (Å)
SUG-THY	167.848	4.2136
SUG-ADE	155.974	4.6175
SUG-GUA	143.419	4.7412
SUG-CYT	193.31	4.1101
SUG-PHO	46.0059	4.0053
PHP-SUG	18.2605	4.4283

Non-Bond Parameters

Two copies of the atoms comprising each pseudo-atom were obtained, displaced by 4 Å in the *z*-direction and individually immersed in a pre-equilibrated box of TIP3 water. The charges were maintained, and hence the only set of atoms with a net charge were the five atoms comprising the phosphate (PHO/PHP) bead, with a net -1 charge per set of atoms. If necessary, the system was neutralized with Na⁺ ions and subjected to the same simulation protocol as before. First 100 ps of NPT dynamics was performed, with snapshots of the last 20 ps (every 2 ps) saved. The ten snapshots were then superimposed on each other and the average MD solution structure obtained as previously discussed.

A constrained MD simulation was then run, where the center of mass of each of the two sets of atoms are constrained by a harmonic potential. Every time step, the distance between the two centers of mass is checked, and if they are not equal to a specified distance, a restoring spring of magnitude $-k\Delta x$ is applied to all of the atoms in the second set. Here, Δx is the differential distance from equilibrium and k is a force constant of $500 \text{ kcal mol}^{-1} \text{ Å}^{-1}$. The center of mass of the two sets of atoms are constrained, starting at 2.0 Å, in 0.1 Å increments, until 10 Å. After each increment, 40 ps of NPT dynamics is performed to equilibrate the structure at the restraint, followed by a further 20 ps of NPT dynamics, during which time the non-bond energy of both sets of atoms is calculated and tabulated.

The average energy per center of mass separation was then computed and plotted vs the center of mass (Fig. 15b). This function was then shifted so that its plateau was at 0, and fitted to the Morse potential in (17) using the least squares fitting procedure previously discussed. This fitting is the effective potential of mean force between both sets of atoms, in the presence of water. This potential has both electrostatic and van der Waals contributions, which is critical since none of the pseudo-atoms are charged.

Meso-Scale Simulation of B-DNA Dodecamer

Simulation Protocol

Bond stretches are the highest frequency modes in any MD simulation. The largest time step of any MD simulation is inversely proportional to the highest frequency modes, which is related to the force constant by $f = 1/2\pi\sqrt{k/\mu}$ (assuming harmonic bonds), where μ is the reduced mass of the parameter atoms being considered. A rule of thumb is that the largest time step possible should be $1/(6*f)$. The pseudo-atoms of this meso-scale force field are much heavier than those of regular atoms and the largest bond force constant is one-fifth that of the typical largest atomistic force constant, which allowed the use of a 10 fs timestep. Furthermore, the meso-scale system contains fewer particles than its atomistic counterpart (each nucleotide is composed of 3 pseudo-atoms, compared to about 40 for an atomistic nucleotide) and it avoids electrostatic interactions because the pseudo-atoms were

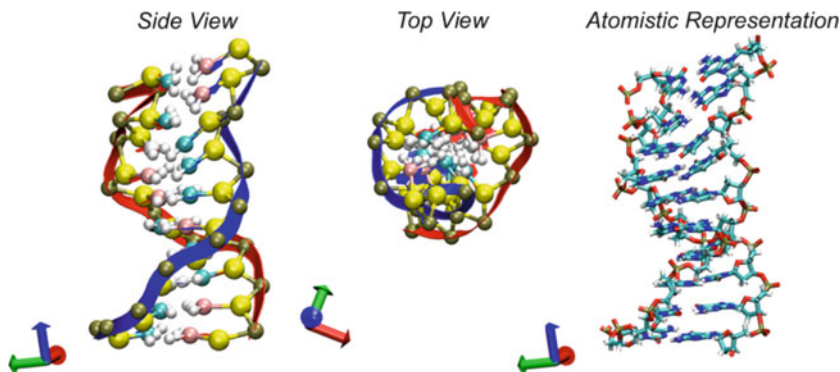


Fig. 16 Meso-dodecamer (*left and center*) and its atomistic reconstructed model (*right*)

chosen to be neutral, which led to an overall increase in the times spanned during simulations.

A 1.4-Å crystal structure of the B-DNA dodecamer (355D) was solvated in a TIP3 water box and neutralized. This fully atomistic structure was simulated using the usual procedure for 2 ns of NPT dynamics, with snapshots of the last 500 ps (every 5 ps) saved for thermodynamic and structural analysis. The average MD structure was calculated as before and the ions were removed. It was then converted into its mesoscale description, with the waters modeled as inflated van der Waals spheres. This water model is the same as the M3B model from our previous work in Molinero et al. [124], and was parameterized to reproduce the density, diffusivity, and cohesive energy of experimental bulk water. The mesoscale dodecamer is then simulated for 2×10^6 steps, representing 2 μ s of total simulation time. Snapshots of the system during the last 50 ns of simulation were saved and used for thermodynamic and structural analysis. The average MD structure during the last 50 ns was calculated and the atomistic level description reconstructed from this average structure. This reconstructed atomistic structure was then minimized and simulated for 1,000 steps of NPT dynamics (Fig. 16).

Comparison of Meso-Scale and Atomistic Dodecamer Simulations

Timing tests indicate that the atomistic level simulation took twice as long to complete 2 ns than the meso-scale did to complete 2 μ s (150 vs 69 CPU hours). This is remarkable, and represents a 2,000 $\times$ speedup for the meso-scale model. This opens the door for studying DNA system in the micro-second timescale.

As a measure of the similarities between the meso, atomistic, and crystal structures, we calculate the CRMS. We find that the simulation structures are quite different from the crystal structure (5.5 Å and 3.2 Å for meso and atomistic, respectively), as well as from each other (4.7 Å difference between the two). This is further illustrated in Fig. 17, where the average MD structures of both the meso-scale and the atomistic structures are compared. A smooth backbone profile can be

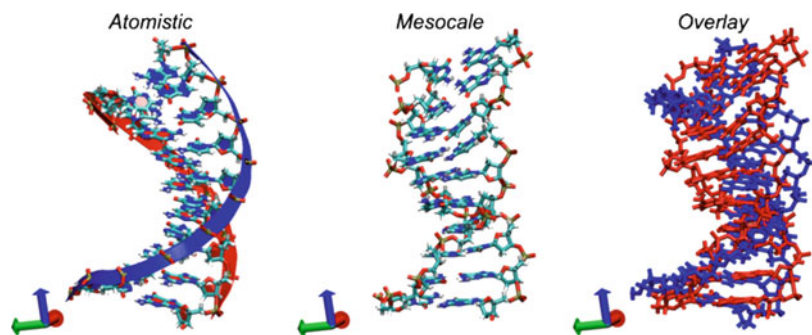


Fig. 17 Comparison of average MD structures between meso-scale and atomistic

observed in the case of the atomistic DNA, contrasted with the irregular profile for the meso-scale case. The overlay of the two structures illustrated the sharp differences in the backbone torsions of the atomistic (blue) and the meso scale (red) structures. A considerably larger distortion can be seen for the meso structure, in particular the presence of several backbone kinks. The backbone parameters for both sets of simulation are within the acceptable range of B-DNA, except for the backbone twist, where the average MD structures are under-twisted (this is a known issue with the AMBER force field). There is also significant deviation in the backbone twist for the meso dodecamer, which is a consequence of the aforementioned kinks in the structure.

From these helical analyses it is clear that further optimization of the meso-scale force field is required. In particular, the Phosphate-Sugar-Phosphate backbone twist angle and the Phosphate-Sugar-Phosphate-Sugar backbone dihedral angles need to be optimized to prevent the under-twisting of the helix seen in the dodecamer simulation. The helical rise in the meso-scale dodecamer is outside the accepted range for B-DNA (3.4 ± 0.2) which, when combined with the under-twisting of the helix, points to unwinding of the helix.

The following section describes the application of QM-parameterized atomistic and coarse-grain potentials to tissue engineering.

4.1.2 Application Example: Mechanoregulation in Polymer-Based Hydrogel Networks for Tissue Engineering

We are currently studying the nanoscale properties that lead to improved micro-scale mechanoregulatory response of polymer-based hydrogel networks for cartilage tissue scaffolding. The structural similarity of synthetic polymer-based hydrogels to the collagen ECM found in human cartilage and their ease of processability makes them ideal candidates for cartilage scaffold-supported cell therapies [131]. These hydrogels can act as a temporary artificial extra cellular matrix (ECM) to provide mechanical support, or provide the ambience for new cells to grow towards

the desired lineage, as well as enable transport or confinement of cells to/within defect sites. Yet, in spite of the significant progress in polymer-based tissue engineering during the past decade, important challenges remain to be addressed in order to restore tissues that serve a predominantly biomechanical function [132–134]. Recent experiments demonstrate the existence of high strength double network polymer-based combinations [135] that may serve as potential candidates for cartilage tissue scaffold development, yet little is known about the atomic composition and the nanometric structures that lead to the enhanced response. We have developed a multiscale approach to characterize the nature of this phenomena [130], and to provide an *in silico* framework for developing improved materials for cell-therapies with high mechanical loading requirements.

Our approach to characterizing the mechanisms and for tuning the mechanical response of polymer-based hydrogels involves a first principles QM-derived Dreiding force field to investigate the thermodynamic and composition conditions for the gel point near the Flory–Stockmayer transition in selected double network hydrogel combinations (poly(acrylamide)-PAAm and poly(2-acrylamido-2-methylpropane-sulfonic acid)-PAMPS), with *N,N'*-methylene bisacrylamide (NNMBA), and a first-principles coarse-grain model for describing the long-term dynamics of percolation based on composition and density ratios between the polymeric and crosslinking agents. We found that percolation events at particular threshold ratios (between polymer components and crosslinker concentrations) are a key atomistic mechanism to promote enhanced mechanical strength in crosslinked hydrogel networks. We used atomic charges and torsional potential energy curves from QM to calculate single chain statistics (including radius of gyration as a function of degree of polymerization). Substituting the radius of gyration in an adjusted continuum model for percolation from [136] led to an estimate of the number of polymer molecules required to achieve percolation using the critical number density found from our atomistic simulations (see Fig. 18). Using our constrained MD with a coarse-grain potential based on a QM parameterized finite extensible nonlinear elastic (FENE) [137], we have been able to determine percolation thresholds as a function of composition (i.e., solvent, crosslinking agent concentration, molar ratio between polymer component, among others) on single and double solvated networks. Figure 19 shows two snapshots of our coarse-grain representation, on the left a set of linear poly(acrylamide) chains mixed at a particular proportion with solvent (water), starter, and crosslinking molecules, and on the right the appearance after running MD-NVE (from a starting temperature of 300K) of a percolated structure of poly(acrylamide). Our results are consistent with experimentally measured gel points and help explain the precipitous loss of the high fracture energy in double network hydrogels at particular crosslink densities [138]. We have also confirmed the mechanical strengthening process via calculation of elastic constants and viscoelastic response of the final networked structures, relative to the single network components. These results will be reported in a separate publication. We were able to determine the critical cross-link concentrations as a function of starting monomer concentrations and degree of polymerization required for improved

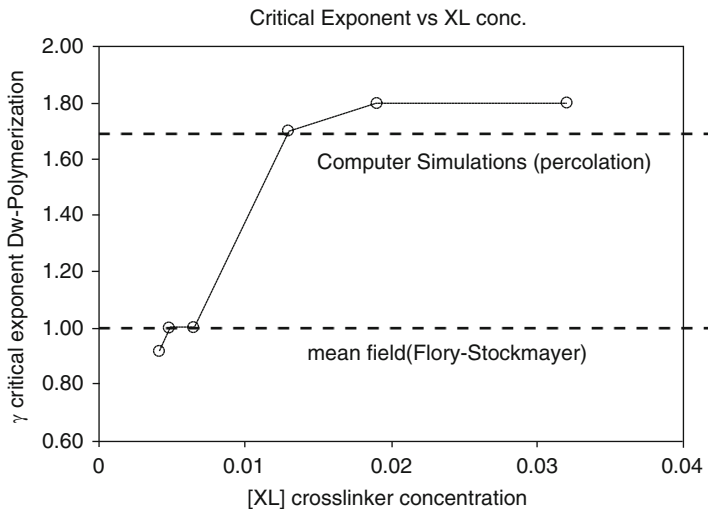


Fig. 18 Critical-exponent for the degree of polymerization in AAM gel networks. From [130]

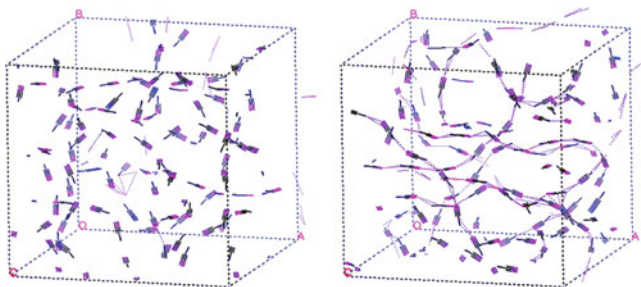


Fig. 19 Snapshots of coarse-grain unit cell model of PAAm and crosslinker mixture (*left*), and PAAm percolated structure after a few ns of MD-NVT (*right*)

mechanical strength. These findings suggest a strategy for systematic characterization and optimization of polymer-based hydrogel molecular network architectures, using design-level tuning of the structural and compositional variables that correlate to their mechanical and thermodynamic response.

5 Concluding Remarks

Seamless integration of paradigms and simulation time- and length-scales, using first-principles-based methods poses tremendous challenges that go far beyond parameter coupling between methods and scales, requiring new physical models

and novel computational schemes capable of systematic upscaling or downscaling. Nevertheless, the hierarchical strategy we have followed has spurred important breakthroughs in the development of methods that bridge paradigms and scales, without sacrificing the ability to model complex chemical processes, including the mixed QM-APBS scheme to understand the solvation effects on molecular reactivity, our reactive and non-adiabatic explicit eFF, reaxFF and eFF, and our efficient coarse-grain methods to bridge, using finer details from atomistic force fields, into the meso scale for conformational studies of systems that are defined through complex chemical interactions (including H bonding). We have demonstrated how these methods are applied to problems that have been untreatable using conventional QM or MM/MD methods.

As described for our simulations of CNT growth, ReaxFF extends our *ab initio* knowledge of the reactivity of small, model systems to complex, extended phenomena involving thousands of atoms throughout nanoseconds of dynamics. This sets us closer to the goal of elucidating and optimizing the fundamental mechanisms that control the chirality of CNTs during their growth. Our results using the eFF methodology, on the hypervelocity shock-induced and electron-induced chemistry of materials, demonstrate its unique ability to depict and predict accurately processes that are highly dependent on electronic contributions, enabling us to understand the real-time dynamics of systems with excited electronic states, all of which is beyond existing QM, QM-approximations (e.g., tight-binding), or MM/MD methods, and in some cases, experimental observation.

Last but not least, our parameterization of non-reactive atomistic and coarse-grain force fields has proven invaluable to understanding and predicting the structure and dynamics of conformations of large biological complexes, such as GPCRs and DNA, both of which lay beyond the capabilities of experiments, considering GPCRs have not been crystallized in their native form (embedded in the cellular membrane), and that activation and critical conformation events in these macromolecules stretch to the micro-second, and beyond, timescales.

We also described the use of our coarse-grain and rigid body MD approaches to elucidate a long-standing problem, associated with the chemical processes that lead to enhanced mechanical strength in polymer-based double network hydrogels (i.e., percolation thresholds as a function of polymer molar ratios and cross-linking concentrations near the Flory–Stockmayer transition limit in polymers gels), of significant importance in the development of scaffolds for load-bearing tissue engineering (e.g., cartilage).

In spite of our progress in methods development and the advent of more powerful computing resources, *in-silico de novo* design and optimization will still have to rely on the correct chemical and physical insight from modelers, especially considering the combinatorial explosion of interactions and the increased number of degrees of freedom as one approaches the continuum macro-scale. As we have shown, a first-principles-based approach does not need to propagate all degrees of freedom during upscaling, but those that are pertinent to the properties and phenomena of interest.

Acknowledgements The material on single shock Hugoniot is based upon work supported by the Department of Energy's National Nuclear Security Administration under Award Number DE-FC52-08NA28613. The material on hydrogel mechanics for tissue engineering scaffolding is based upon work supported by the National Science Foundation (CMMI 0727870). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author/s and do not necessarily reflect the views of the National Science Foundation. The early developments of these multiscale multiparadigm methods were initiated with support by DOE under the ECUT program (Prog. Mgr. Minoo Dastoor) and continued DAPRA under the PROM and ONR programs (Prog. Mgr. Carey Swartz, Judah Goldwasser, and Steve Wax). Substantial support was provided by ARO-MURI, ONR-MURI, DURIP, and ASCI projects along with Chevron, Dow-Corning, Aventis, Asahi Kasei, Intel, PharmSelex, and many other industrial labs.

References

1. de Broglie L (1924) Recherches sur la théorie des quanta
2. Schrodinger E (1926) Quantification of the eigen-value problem. *Ann Phys* 79(6):489–527
3. Messiah A (ed) (1999) Quantum mechanics, vol 1. Reprinted by Dover Publications, Mineola, NY
4. Anderson JB (1976) Quantum chemistry by Random-Walk – H₂p, H+3d3h1a'1, H-23-Sigma+U, H-41-Sigma+G, Be1s. *J Chem Phys* 65(10):4121–4127
5. Williamson AJ, Hood RQ, Grossman JC (2001) Linear-scaling quantum Monte Carlo calculations. *Phys Rev Lett* 8724(24):246406-(4)
6. Reboredo FA, Williamson AJ (2005) Optimized nonorthogonal localized orbitals for linear scaling quantum Monte Carlo calculations. *Phys Rev B* 71(12):121105-(4)
7. Morokuma K et al (2001) Model studies of the structures, reactivities, and reaction mechanisms of metalloenzymes. *Ibm J Res Dev* 45(3–4):367–395
8. Fisher DR et al (2008) An optimized initialization algorithm to ensure accuracy in quantum Monte Carlo calculations. *J Comput Chem* 29(14):2335–2343
9. Anderson AG, Goddard WA, Schroder P (2007) Quantum Monte Carlo on graphical processing units. *Comput Phys Commun* 177(3):298–306
10. Anderson AG, Goddard WA (2010) Generalized valence bond wave functions in quantum Monte Carlo. *J Chem Phys* 132(16):164110-(10)
11. Born M, Oppenheimer R (1927) Quantum theory of molecules. *Ann Phys* 84(20):0457–0484
12. Frisch MJ et al (2004) Gaussian. Gaussian, Inc., Wallingford CT
13. Schmidt MW et al (1993) General atomic and molecular electronic-structure system. *J Comput Chem* 14(11):1347–1363
14. Jaguar (1991–2000) Jaguar. Schrodinger, Inc., Portland, OR
15. Kresse G, Hafner J (1993) Ab initio molecular-dynamics for liquid-metals. *Phys Rev B* 47(1):558–561
16. Dovesi R et al (2006) Crystal, U.o. Torino, Editor. 2006: Torino
17. Segall MD et al (2002) First-principles simulation: ideas, illustrations and the CASTEP code. *J Phys Condens Matter* 14(11):2717–2744
18. Schultz PA (2007) SeqQuest electronic structure code. Sandia National Laboratories, Albuquerque
19. Harvey J (2001) Molecular electronic structure. Available from: http://www.chm.bris.ac.uk/pt/harvey/elstruct/beyond_hf.html
20. Moller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. *Phys Rev* 46(7):0618–0622
21. Carter EA, Goddard WA (1988) Correlation-consistent configuration-interaction – accurate bond-dissociation energies from simple wave-functions. *J Chem Phys* 88(5):3132–3140

22. Friesner RA (2005) Ab initio quantum chemistry: methodology and applications. *Proc Natl Acad Sci U S A* 102(19):6648–6653
23. Goddard WA et al (1973) Generalized valence bond description of bonding in low-lying states of molecules. *Acc Chem Res* 6(11):368–376
24. Greeley BH et al (1994) New pseudospectral algorithms for electronic-structure calculations – length scale separation and analytical 2-electron integral corrections. *J Chem Phys* 101(5): 4028–4041
25. Tannor DJ et al (1994) Accurate first principles calculation of molecular charge-distributions and solvation energies from ab-initio quantum-mechanics and continuum dielectric theory. *J Am Chem Soc* 116(26):11875–11882
26. Bredow T, Jug K (2005) Theory and range of modern semiempirical molecular orbital methods. *Theor Chem Acc* 113(1):1–14
27. Hohenberg P, Kohn W (1964) *Phys Rev B* 136(3):864
28. Kohn W, Sham LJ (1965) *Phys Rev A* 140(4):1133
29. Li ZY, He W, Yang JL (2005) Recent progress in density functional theory and its numerical methods. *Prog Chem* 17(2):192–202
30. Foulkes WMC et al (2001) Quantum Monte Carlo simulations of solids (review article). *Rev Mod Phys* 73:33
31. Chen XJ, Langlois JM, Goddard WA (1995) Dual-space approach for density-functional calculations of 2-dimensional and 3-dimensional crystals using Gaussian-basis functions. *Phys Rev B* 52(4):2348–2361
32. Liu Y, Goddard WA (2009) A universal damping function for empirical dispersion correction on density functional theory. *Mater Trans* 50(7):1664–1670
33. Dobson JC, Meyer TJ (1988) Redox properties and ligand loss chemistry in aqua hydroxo complexes derived from CIS-(BPY)2RU2(OH2)2 2+ and TRANS- (BPY)2RUII(OH2) 2 2+. *Inorg Chem* 27(19):3283–3291
34. Hill T (1960) An introduction to statistical thermodynamics. Addison-Wesley, Reading
35. Tissandier MD et al (1998) The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *J Phys Chem A* 102(40):7787–7794
36. Kelly CP, Cramer CJ, Truhlar DG (2006) Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J Phys Chem B* 110(32):16066–16081
37. Srncic M et al (2008) Effect of spin-orbit coupling on reduction potentials of octahedral ruthenium(II/III) and osmium(II/III) complexes. *J Am Chem Soc* 130(33):10947–10954
38. Car R, Parrinello M (1985) Unified approach for molecular-dynamics and density-functional theory. *Phys Rev Lett* 55(22):2471–2474
39. Mayo SL, Olafson BD, Goddard WA III (1990) DREIDING: a generic force field for molecular simulations. *J Phys Chem* 94:8897–8909
40. Ding HQ, Karasawa N, Goddard WA (1992) Atomic level simulations on a million particles – the cell multipole method for Coulomb and London nonbond interactions. *J Chem Phys* 97(6):4309–4315
41. Ding HQ, Karasawa N, Goddard WA (1992) The reduced cell multipole method for Coulomb interactions in periodic-systems with million-atom unit cells. *Chem Phys Lett* 196(1–2):6–10
42. Ponder JW, Case DA (2003) Force fields for protein simulations. *Protein Simul* 66:27–85
43. Weiner SJ et al (1986) An all atom force-field for simulations of proteins and nucleic-acids. *J Comput Chem* 7(2):230–252
44. MacKerell AD et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616
45. Jorgensen WL, Maxwell DS, TiradoRives 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
46. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 9(9):646–652

47. Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery (vol 7, p 339, 2008). *Nat Rev Drug Discov* 7(6):542
48. Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* 7(4):339–357
49. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theory Comput* 4(12):2160–2169
50. van Duin ACT et al (2001) ReaxFF: a reactive force field for hydrocarbons. *J Phys Chem A* 105(41):9396–9409
51. Su JT, Goddard WA (2007) Excited electron dynamics modeling of warm dense matter. *Phys Rev Lett* 99(18):185003
52. Strachan A et al (2005) Thermal decomposition of RDX from reactive molecular dynamics. *J Chem Phys* 122(5):054502
53. van Duin ACT et al (2003) ReaxFF(SiO) reactive force field for silicon and silicon oxide systems. *J Phys Chem A* 107(19):3803–3811
54. Han SS et al (2005) Optimization and application of lithium parameters for the reactive force field, ReaxFF. *J Phys Chem A* 109(20):4575–4582
55. Zhang Q et al (2004) Adhesion and nonwetting-wetting transition in the Al/ α -Al₂O₃ interface. *Phys Rev B* 69(4):045423–(11)
56. Nielson KD et al (2005) Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes. *J Phys Chem A* 109(3):493–499
57. Cheung S et al (2005) ReaxFF(MgH) reactive force field for magnesium hydride systems. *J Phys Chem A* 109(5):851–859
58. Chen N et al (2005) Mechanical properties of connected carbon nanorings via molecular dynamics simulation. *Phys Rev B* 72(8):085416–(9)
59. Su HB et al (2007) Simulations on the effects of confinement and Ni-catalysis on the formation of tubular fullerene structures from peapod precursors. *Phys Rev B* 75(13):134107–(5)
60. Chenoweth K et al (2005) Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field. *J Am Chem Soc* 127(19):7192–7202
61. Strachan A et al (2003) Shock waves in high-energy materials: the initial chemical events in nitramine RDX. *Phys Rev Lett* 91(9):098301–(4)
62. van Duin ACT et al (2005) Atomistic-scale simulations of the initial chemical events in the thermal initiation of triacetone triperoxide. *J Am Chem Soc* 127(31):11053–11062
63. Buehler MJ, van Duin ACT, Goddard WA (2006) Multiparadigm modeling of dynamical crack propagation in silicon using a reactive force field. *Phys Rev Lett* 96(9):095505–(4)
64. Goddard WA et al (2006) Development of the ReaxFF reactive force field for mechanistic studies of catalytic selective oxidation processes on BiMoO_x. *Top Catal* 38(1–3):93–103
65. Ludwig J et al (2006) Dynamics of the dissociation of hydrogen on stepped platinum surfaces using the ReaxFF reactive force field. *J Phys Chem B* 110(9):4274–4282
66. Iijima S (1991) Helical microtubules of graphitic carbon. *Nature* 354(6348):56–58
67. Endo M, Strano MS, Ajayan PM (2008) Potential applications of carbon nanotubes. *Carbon Nanotubes* 111:13–61
68. Kreupl F (2008) Carbon nanotubes in microelectronic applications. In: Hierold C (ed) *Carbon nanotube devices*, 2nd edn. Wiley VCH, Weinheim
69. Stampfer C (2008) Electromechanical carbon nanotube transducers. In: Hierold C (ed) *Carbon nanotube devices*, 2nd edn. Wiley VCH, Weinheim
70. Roman C (2008) Modeling the properties of carbon nanotubes for sensor-based devices. In: Hierold C (ed) *Carbon nanotube devices*, 2nd edn. Wiley VCH, Weinheim
71. Robertson J (2008) Carbon nanotube field emission devices. In: Hierold C (ed) *Carbon nanotube devices*, 2nd edn. Wiley VCH, Weinheim
72. Yeow JT (2008) Carbon nanotube gas sensors. In: Hierold C (ed) *Carbon nanotube devices*, 2nd edn. Wiley VCH, Weinheim

73. Joselevich E (2008) Carbon nanotube synthesis and organization. In: Jorio A, Dresselhaus G, Dresselhaus M (eds) Carbon nanotubes, 2nd edn. Springer, Berlin
74. Bolton K et al (2009) Density functional theory and tight binding-based dynamical studies of carbon-metal systems of relevance to carbon nanotube growth. *Nano Res* 2(10):774–782
75. Gavillet J et al (2001) Root-growth mechanism for single-wall carbon nanotubes. *Phys Rev Lett* 87(27):275504–(4)
76. Huang SM et al (2004) Growth mechanism of oriented long single walled carbon nanotubes using “fast-heating” chemical vapor deposition process. *Nano Lett* 4(6):1025–1028
77. Li YM et al (2001) Growth of single-walled carbon nanotubes from discrete catalytic nanoparticles of various sizes. *J Phys Chem B* 105(46):11424–11431
78. Helveg S et al (2004) Atomic-scale imaging of carbon nanofibre growth. *Nature* 427(6973):426–429
79. Hofmann S et al (2007) In situ observations of catalyst dynamics during surface-bound carbon nanotube nucleation. *Nano Lett* 7(3):602–608
80. Raty JY, Gygi F, Galli G (2005) Growth of carbon nanotubes on metal nanoparticles: a microscopic mechanism from ab initio molecular dynamics simulations. *Phys Rev Lett* 95(9):096103–(4)
81. Abild-Pedersen F et al (2006) Mechanisms for catalytic carbon nanofiber growth studied by ab initio density functional theory calculations. *Phys Rev B* 73(11):115419–(13)
82. Amara H et al (2009) Tight-binding potential for atomistic simulations of carbon interacting with transition metals: application to the Ni-C system. *Phys Rev B* 79(1):014109–(17)
83. Ohta Y et al (2008) Rapid growth of a single-walled carbon nanotube on an iron cluster: density-functional tight-binding molecular dynamics simulations. *ACS Nano* 2(7):1437–1444
84. Moors M et al (2009) Early stages in the nucleation process of carbon nanotubes. *ACS Nano* 3(3):511–516
85. Mueller JE, van Duin ACT, Goddard WA (2010) Application of the ReaxFF reactive force field to reactive dynamics of hydrocarbon chemisorption and decomposition. *J Phys Chem C* 114(12):5675–5685
86. Mueller JE, van Duin ACT, Goddard WA (2010) Development and validation of ReaxFF reactive force field for hydrocarbon chemistry catalyzed by nickel. *J Phys Chem C* 114(11):4939–4949
87. Mora E et al (2008) Low-temperature single-wall carbon nanotubes synthesis: feedstock decomposition limited growth. *J Am Chem Soc* 130(36):11840–11841
88. Hofmann S et al (2005) Surface diffusion: the low activation energy path for nanotube growth. *Phys Rev Lett* 95(3):036101–(4)
89. Henkelman G, Jonsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. *J Chem Phys* 115(21):9657–9666
90. Jaramillo-Botero A et al (2010) Large-scale, long-term non-adiabatic electron molecular dynamics for describing material properties and phenomena in extreme environments. *J Comput Chem*. epub ahead of print.
91. Su JT, Goddard WA (2009) The dynamics of highly excited electronic systems: applications of the electron force field. *J Chem Phys* 131(24):244501–(20)
92. Gillis HP et al (1995) Low-energy electron-enhanced etching of Si(100) in hydrogen helium direct-current plasma. *Appl Phys Lett* 66(19):2475–2477
93. Su JT, Goddard WA (2009) Mechanisms of Auger-induced chemistry derived from wave packet dynamics. *Proc Natl Acad Sci U S A* 106(4):1001–1005
94. Goddard W III (1998) Nanoscale theory and simulation: a critical driver for and a critical challenge to commercial nanotechnology. In: WTEC workshop. World Technology Evaluation Center, Arlington, VA
95. Jaramillo-Botero A et al (2008) Multiscale-multiparadigm modeling and simulation of nanometer scale systems and processes for nanomedical applications. In: Zhang M, Xi N (eds) Nanomedicine: a systems engineering approach. Pan Stanford Publishing, Singapore

96. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with constraints – molecular-dynamics of n-alkanes. *J Comput Phys* 23(3):327–341
97. Van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* 34(5):1311–1327
98. Andersen HC (1983) Rattle – a velocity version of the shake algorithm for molecular-dynamics calculations. *J Comput Phys* 52(1):24–34
99. Krautler V, Van Gunsteren WF, Hunenberger PH (2001) A fast SHAKE: algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* 22(5):501–508
100. Hess B et al (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463–1472
101. Miyamoto S, Kollman PA (1992) Settle – an analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* 13(8):952–962
102. Bae DS, Haug EJ (1987) A recursive formulation for constrained mechanical system dynamics. 1. Open loop-systems. *Mech Struct Mach* 15(3):359–382
103. Abagyan RA, Mazur AK (1989) New methodology for computer-aided modeling of biomolecular structure and dynamics. 2. Local deformations and cycles. *J Biomol Struct Dyn* 6(4): 833–845
104. Mazur AK, Abagyan RA (1989) New methodology for computer-aided modeling of biomolecular structure and dynamics. 1. Non-cyclic structures. *J Biomol Struct Dyn* 6(4):815–832
105. Bae DS, Kuhl JG, Haug EJ (1988) A recursive formulation for constrained mechanical system dynamics. 3. Parallel processor implementation. *Mech Struct Mach* 16(2):249–269
106. Bae DS, Haug EJ (1988) A recursive formulation for constrained mechanical system dynamics. 2. Closed-loop systems. *Mech Struct Mach* 15(4):481–506
107. Jaramillo-Botero A, Lorente ACI (2002) A unified formulation for massively parallel rigid multibody dynamics of $O(\log(2) n)$ computational complexity. *J Parallel Distrib Comput* 62(6):1001–1020
108. Vaidehi N, Jain A, Goddard WA (1996) Constant temperature constrained molecular dynamics: The Newton-Euler inverse mass operator method. *J Phys Chem* 100(25): 10508–10517
109. Jain A, Vaidehi N, Rodriguez G (1993) A fast recursive algorithm for molecular-dynamics simulation. *J Comput Phys* 106(2):258–268
110. Mathiowetz AM et al (1994) Protein simulations using techniques suitable for very large systems – the cell multipole method for nonbond interactions and the Newton-Euler inverse mass operator method for internal coordinate dynamics. *Proteins Struct Funct Genet* 20(3):227–247
111. Jaramillo-Botero A, Liu Y, Goddard WA (2006) The computational materials design facility: CMDF. Available from: <http://www.wag.caltech.edu/multiscale>
112. Featherstone R (1983) The calculation of robot dynamics using articulated-body inertias. *Int J Rob Res* 2(1):13–30
113. Fijany A et al (1998) Novel algorithms for massively parallel, long-term, simulation of molecular dynamics systems. *Adv Eng Softw* 29(3–6):441–450
114. Shelley J et al (2001) A coarse grain model for phospholipid simulations. *J Phys Chem B* 105(19):4464–4470
115. Groot RD (2000) Mesoscopic simulation of polymer-surfactant aggregation. *Langmuir* 16(19):7493–7502
116. Groot RD, Rabone KL (2001) Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. *Biophys J* 81(2):725–736
117. Jedlovsky P (1998) Investigation of the orientational correlation of the molecules in liquid H₂S with reverse Monte Carlo simulation. *Mol Phys* 93(6):939–946
118. Jedlovsky P et al (1996) Investigation of the uniqueness of the reverse Monte Carlo method: studies on liquid water. *J Chem Phys* 105(1):245–254

119. Lopez CF et al (2002) Computer simulation studies of biomembranes using a coarse grain model. *Comput Phys Commun* 147(1–2):1–6
120. Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. *J Phys Chem B* 108(2):750–760
121. Marrink SJ, Mark AE (2003) Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J Am Chem Soc* 125(49):15233–15242
122. Marrink SJ, Mark AE (2004) Molecular view of hexagonal phase formation in phospholipid membranes. *Biophys J* 87(6):3894–3900
123. Marrink SJ et al (2007) The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B* 111(27):7812–7824
124. Molinero V, Goddard WA (2004) M3B: a coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. *J Phys Chem B* 108(4):1414–1427
125. Vaidehi N, Goddard WA (2000) Domain motions in phosphoglycerate kinase using hierarchical NEIMO molecular dynamics simulations. *J Phys Chem A* 104(11):2375–2383
126. Cagin T et al (2001) Multiscale modeling and simulation methods with applications to dendritic polymers. *Comput Theor Polym Sci* 11(5):345–356
127. Elezgaray J, Laguerre M (2006) A systematic method to derive force fields for coarse-grained simulations of phospholipids. *Comput Phys Commun* 175(4):264–268
128. Hunger J, Huttner G (1999) Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks. *J Comput Chem* 20(4):455–471
129. Hunger J et al (1998) How to derive force field parameters by genetic algorithms: modelling tripod-Mo(CO)(3) compounds as an example. *Eur J Inorg Chem* 6:693–702
130. Jaramillo-Botero A et al (2010) First-principles based approaches to nano-mechanical and biomimetic characterization of polymer-based hydrogel networks for cartilage scaffold-supported therapies. *J Comput Theor Nanosci* 7(7):1238–1256
131. Varghese S, Elisseeff JH (2006) Hydrogels for musculoskeletal tissue engineering. In: *Polymers for regenerative medicine*. Springer-Verlag, Berlin, pp 95–144
132. Butler DL, Goldstein SA, Guilak F (2000) Functional tissue engineering: the role of biomechanics. *J Biomech Eng Trans Asme* 122(6):570–575
133. Guilak F (2000) The deformation behavior and viscoelastic properties of chondrocytes in articular cartilage. *Biorheology* 37(1–2):27–44
134. Guilak F, Butler DL, Goldstein SA (2001) Functional tissue engineering – the role of biomechanics in articular cartilage repair. *Clin Orthop Relat Res* 391:S295–S305
135. Gong JP et al (2003) Double-network hydrogels with extremely high mechanical strength. *Adv Mater* 15(14):1155–1158
136. Lorenz CD, Ziff RM (2001) Precise determination of the critical percolation threshold for the three-dimensional “Swiss cheese” model using a growth algorithm. *J Chem Phys* 114(8):3659–3661
137. Kremer K, Grest GS (1990) Dynamics of entangled linear polymer melts – a molecular-dynamics simulation. *J Chem Phys* 92(8):5057–5086
138. Blanco M, Jaramillo-Botero A, Goddard W III (2010) The percolation limit near the Flory-Stockmayer transition in polymer hydrogel networks. California Institute of Technology, Pasadena

Dynamic QM/MM: A Hybrid Approach to Simulating Gas–Liquid Interactions

Scott Yockel and George C. Schatz

Abstract In this chapter we describe molecular dynamics simulation methods in which the system being studied is divided into a region where quantum mechanics (QM) is used to determine forces for doing Born-Oppenheimer direct dynamics calculations (i.e., doing electronic structure calculations on the fly to determine energies and forces) and another region where empirical potentials that are commonly used in molecular mechanics (MM) calculations are used to determine forces. The two regions are linked through an embedding process that may or may not involve the possibility that atoms can be passed back and forth between regions at each time step. The idea with this dynamic QM/MM methodology is that one uses QM calculations to define the potential surface in portions of the system where reaction occurs, and MM to determine forces in what is typically a much larger region where no reaction occurs. This approach thereby enables the description of chemical reactions in the QM region, which is a technology that can be used in many different applications. We illustrate its use by describing work that we have done with gas–liquid reactions in which a reactive atom (such as an oxygen or fluorine atom) reacts with the surface of a liquid and the products can either remain in the liquid or emerge into the gas phase. Applications to hydrocarbon and ionic liquids are described, including the characterization of reaction mechanisms at hyperthermal energies, and the determination of product branching and product energy distributions.

Keywords Gas-liquid scattering · Hybrid QM/MM molecular dynamics · Interfacial chemistry · Room temperature ionic liquids · Squalane

S. Yockel
University of North Texas, 1155 Union Circle #305398, Denton, TX 76203-5017, USA

G.C. Schatz (✉)
Northwestern University, 2145 Sheridan Rd, Evanston, IL 60208-3113, USA
e-mail: schatz@chem.northwestern.edu

Contents

1	Introduction	44
2	Description of the Method	46
2.1	Basic QM/MM	46
2.2	Dynamic Partitioning via the “Seed Atom” Method	47
2.3	Issues with Dynamic Partitioning	49
3	Simulating Gas–Liquid Interactions	51
3.1	Building a Model Liquid Surface	51
3.2	Gas–Liquid Scattering Model	51
4	Application of Dynamic QM/MM to Gas–Liquid Scattering	53
4.1	Squalane	53
4.2	[emim][NO ₃]	58
5	Summary	63
	References	65

1 Introduction

There are many cases in simulating chemical processes in which the necessary system size to provide a realistic model is quite large (e.g., enzyme reactions, surface phenomena, condensed phase, ...). While the structure and electrostatic interactions in these systems can be well studied with typical molecular mechanics (MM) force fields using molecular dynamics (MD), no chemical reactions (bond making/breaking events) are allowed to take place. On the other hand, it is well established that standard quantum mechanical (QM) methods can compute bond making/breaking with great accuracy; however, the efficiency of these codes limits their practical use to only ~100 atoms on typical computational resources. Even if the computing power was available to handle a larger (~1,000 atoms) QM computation, then it is unlikely that a simulation could be performed for long enough (~10–100 ps) to represent appreciable chemical change in the molecular system. Therefore, dividing up a large-scale system into reactive (bond making/breaking) and nonreactive (electrostatic, steric, interacting vdW, ...) regions is essential.

Research combining QM and MM has an extensive history going back to the early 1970s, and there are numerous review articles [1–12], mostly concerned with enzymatic reactions. A key component of many of these studies concerns the technique used to describe covalent bonds that cross between the QM and MM regions. A common procedure is to add an atom in the QM region where the covalent bond is cut so as to avoid the creation of dangling bonds. This link-atom approach [13] is relatively easy to implement, but there are many variations on what the properties of the link-atom should be [14]. Typically it is taken to be a hydrogen atom, which means that standard electronic structure packages can be used without modification; however in some cases this is not appropriate, or modified hydrogen atom properties should be used. There are also questions about how the QM calculations should be embedded into the overall system. Popular schemes are based on the Morokuma ONIOM scheme or related schemes [15–17], but there

are issues concerning the partitioning of charges between QM or MM regions that keep being refined.

One major difference in the charges in the QM and MM regions is that the MM charges typically do not change during the simulation as they are predefined in the potential force field. The development of QM/MM methods in which the charges from one region dynamically affect those in the other, as would be the case in a true system, is an active aspect of research [18–20]. This approach is referred to as electrical embedding, as opposed to the simple mechanical embedding where the charges are fixed. Evidently, mechanical embedding should be used when charge transfer or long range polarization effects are minimal for the chemical system being studied.

Only a small fraction of the QM/MM studies have explicitly considered molecular dynamics calculations involving both QM and MM regions, as the computational time associated with a QM molecular dynamics calculation is quite demanding. Some examples where this has been done are covered in [21–26]. Recently, our group has developed a simulation model to specifically study reactions at the gas–liquid interface that merges classical MM for the entire system with a QM treatment of the atoms in the reaction region(s) at each time step to form a hybrid QM/MM–MD approach. This new computational approach extends the direct dynamics models our group has used to study gas phase collision chemistry previously [27], and now includes MM calculations for the liquid that is not involved in reactive events. Our primary target liquids have been squalane (for which there are many experiments) and ionic liquids (an unconventional fuel of current interest), and in most cases we have been interested in studying the collision of reactive atoms such as atomic oxygen (O) and fluorine (F) with the liquid surfaces in order to simulate gas–surface molecular beam experiments. In addition, since most of these experiments refer to reactive atoms that have several eV of energy (hyperthermal energies), our applications have been concerned with highly nonthermal processes in which reaction mechanisms that do not normally contribute to thermal kinetics are described.

Projects concerning the O + squalane [28] and F + squalane [29] reaction dynamics have been completed, and they provide detailed information about the spatial distribution of reactive sites and the correlation of the reaction mechanism with the angular and translational distribution of the scattered products. These results have been used to give a detailed mechanistic understanding of beam–liquid surface experiments in the Minton lab [30, 31]. In some cases we were able to identify products not initially considered in the experimental studies, providing stimulus for subsequent molecular beam measurements where they were seen [32]. Our F + squalane studies [29] have been used to interpret experiments done by Nesbitt and coworkers at thermal collision energies [33–35], showing that there is a component of the HF product vibrational and rotational distributions which involves escape of HF from the surface with little relaxation. In more recent work we have considered the collisions of atomic oxygen (and also hyperthermal argon) with the surfaces of ionic liquids [36, 37], providing detailed information about a class of chemical reactions that has not been considered previously, and enabling the interpretation of experiments by Minton, McKendrick, and collaborators.

Our motivation for doing these studies is that the structure and dynamics of chemical interfaces at phase boundaries (e.g., gas–liquid interface) is of significant importance to chemistry and only recently has the capability been established to describe these processes (both with modeling and with experiments) at an atomistic level. Newly developed experimental methods provide the capability to study the reactivity of a liquid surface with beam–surface scattering, which provides detailed mechanistic information, but it is often difficult to interpret these experiments without theoretical modeling. Among the experiments being developed is work by Nesbitt et al. [33–35] who have resolved quantum-state reaction dynamics via direct absorption detection of the rovibrational states of the nascent gaseous products coming from the surface using high resolution infrared spectroscopy. Related technology has been developed by Nathanson [38, 39] and McKendrick [40, 41], who have used laser-induced fluorescence methods to detect the nascent products in gas–liquid surface reaction dynamics studies. In addition, Minton and coworkers [31, 32] have performed molecular beam experiments involving hyperthermal oxygen scattering from liquid surfaces, in which a mass spectrometer is used to determine the angular and translational energy distributions of the nascent gaseous products. These experiments, as well as others, require theoretical modeling that includes the flexibility to compute the dynamics of the reactive events (and sometimes several sequential reactive events), while at the same time the simulation needs to include a considerable portion of the liquid (thousands of atoms) so that energy transfer, diffusion, and electrostatic interactions can be described. Also, the hydrocarbon and ionic liquids are often nanostructured, which means that a simulation region of sufficient size is needed in order to capture all of the dynamical complexity that can occur.

2 Description of the Method

2.1 *Basic QM/MM*

Since most large-scale chemical processes can generally be partitioned into reactive and nonreactive regions, dividing up the system to be treated with different levels of sophistication in the theoretical model is logical. As stated above, the atoms in the reactive parts of the model are treated by direct dynamics QM (electronic structure) computations, while the remaining nonreactive part is treated with an empirical MM potential. Our dynamic QM/MM code utilizes various subroutines from TINKER 4.2 [42] to compute the MM part of the simulation (forces and potential of the MM atoms). For the QM part of the code, subroutines from the MSINDO 2.1 electronic structure code were used [43], which includes the MSINDO (modified symmetrically orthogonalized intermediate neglect of differential overlap) semi-empirical Hamiltonian to compute the energy [44, 45]. In general, semiempirical methods are a middle ground between a fully ab initio QM calculation and fully

empirical MM potentials and are suitable for efficiently computing the electronic structures of hundreds of atoms in a reasonable time (seconds) to be used in dynamics. In comparison to other standard semiempirical methods, MSINDO provides reaction energies and barriers for gas-phase reactions of atoms with small hydrocarbons that are nearer to experimental values than AM1 and PM3 [27]. Therefore, it has been the QM method of choice for our studies on $\text{O}(^3\text{P})$ and $\text{F}(^2\text{P})$ with squalane. Additionally, we have used MSINDO when studying $\text{O}(^3\text{P}) + [\text{emim}][\text{NO}_3]$, a room temperature ionic liquid.

Many of the features of the nondynamic QM/MM model, which has been widely used as a multilevel computation, exist in our dynamic QM/MM algorithm. When bonds straddle the QM/MM boundary, the link atom method is employed to truncate the QM region without leaving dangling bonds. For example, if a C–C bond spans this boundary, then the C(MM) atom is replaced with an H(QM) atom at 0.9 Å (the equilibrium C–H bond length) from the C(QM) during the QM part of the computation. During a full QM/MM computation, three separate calculations are performed, as prescribed by the ONIOM mechanical embedding approach. In the following notation, the atomic system is denoted in parenthesis. First the full atomic system is computed with MM, $[E_{\text{MM}}(\text{QM} + \text{MM})]$. Then the QM atomic system is computed with MM $[E_{\text{MM}}(\text{QM})]$, and the QM system with QM $[E_{\text{QM}}(\text{QM})]$. Therefore, the energies, potentials, forces and other properties are computed as follows: $E_{\text{QM/MM}} = E_{\text{MM}}(\text{QM} + \text{MM}) - E_{\text{MM}}(\text{QM}) + E_{\text{QM}}(\text{QM})$.

Subtracting $E_{\text{MM}}(\text{QM})$ from $E_{\text{MM}}(\text{QM} + \text{MM})$ is an approximation that describes electrostatic interactions between charges in the QM and MM regions at the MM level. The importance of this depends on the system being studied and the choice of partitioning at the QM/MM boundary, but it is consistent with the assumptions in the mechanical embedding model. In the case of our studies with squalane, the system is a neutral nonpolar liquid and charge transfer between interacting atoms from inside to outside the reactive region is thought to be negligible; therefore a simple mechanical embedding approach was chosen. Additionally, we rely on this assumption, as discussed later, when we allow the QM region to change over time.

One of the most common numerical methods used in molecular dynamics to solve Newton's equations of motions is the Velocity Verlet integrator. This is typically implemented as a second order method, and we find that it can become numerically unstable during the course of hyperthermal collision events, where the atom velocities are often far from equilibrium. As an alternative, we have implemented a fifth/sixth order predictor–corrector scheme for our calculations. Specifically, the driver we chose utilizes the Adams–Bashforth predictor method together with the Adams–Moulton corrector method for approximating the solution to the equations of motion.

2.2 Dynamic Partitioning via the “Seed Atom” Method

In some simulations, like that of large biological systems that involve reactive sites and proton transfer, the choice of a reactive region does not change while the

reaction occurs and the QM/MM scheme is easily implemented. However, there are many chemical systems in which a priori knowledge about where the reactive region will be during the course of the simulation is not predictable by chemical intuition. Within our group we have had success employing a simplistic approach to a QM/MM–MD algorithm that is applicable to chemical systems where a priori knowledge is largely unavailable as to which atoms should be considered as part of the reactive region [28, 29]. This is specifically the case when gaseous atoms collide with an amorphous surface that can readily diffuse over time. For example, in our studies with squalane we even found that the incident atom was not always confined within a ~ 20 Å (radius) $\times$ 20 Å (height) cylindrical box that contains over 2,000 atoms comprising the liquid. This is not the situation that is observed with crystalline solids or self-assembled monolayers. When modeling those surfaces, one can predefine atoms into a QM region because the incident atoms would rarely interact with more than the first or second monolayer.

In order to circumvent the limitation of predefining which atoms are to be treated with QM, we allow atoms to be redefined dynamically as “in” or “out” of the reactive region depending on their location relative to radical species that are capable of undergoing a reaction. Specifically, the reactive regions are centered around “seed atoms”, which are defined as all of the open-shell/radical atoms. In the case of our studies involving atomic radicals reacting with squalane, the seed atoms were defined as the initial gaseous incident atom plus, over the course of the simulation, any atom that loses one of its originally bonded atoms. For example (see Fig. 1), if the $F(^2P)$ atom abstracts a H from squalane, then the carbon radical

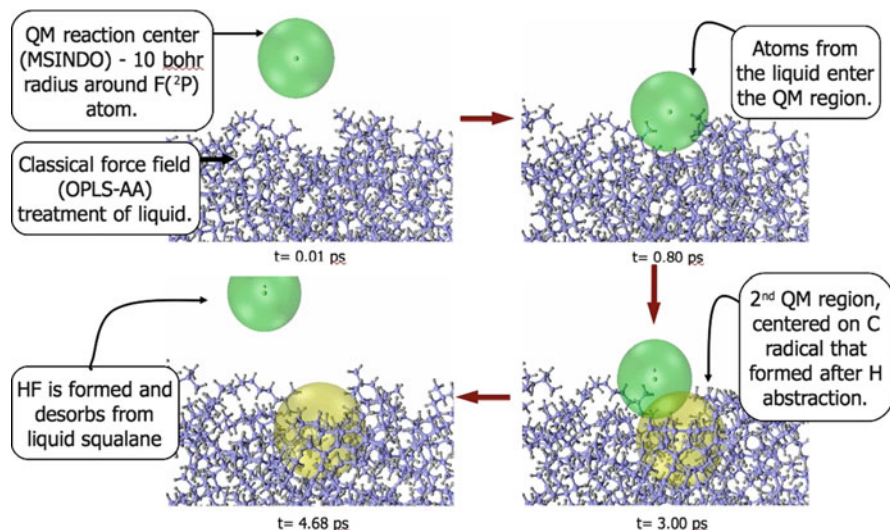


Fig. 1 Pictorial representation of the dynamic partitioning using the “seed atom” method for the $F(^2P)$ + squalane reaction

site also becomes a seed atom, since this radical site is considered to be a reactive atom. The radius of the QM region around a seed atom (or the distance from the seed atom to atoms treated as MM) was defined as the distance where the forces between the atoms are less than 1×10^{-7} hartree/bohr, at which point the differences between MM and QM forces are thought to be negligible. This was empirically determined to be 10.0, 10.0, 12.0, 12.5 bohr for F, O, H, and C, respectively, when interacting with methane. We also found it convenient to do an assessment on which atoms should or should not be in the reactive region only every ten time steps, since this involves a distance search and, for 10-a.u. time steps, the atoms have not moved very far. Throughout much of the simulation it is common to have between 75 and 150 atoms within the reactive region(s). Given the effort associated with QM calculations at each time step, the use of a computationally efficient semiempirical QM method such as MSINDO is clearly desirable.

2.3 *Issues with Dynamic Partitioning*

It must be noted that the instantaneous switching of an atom in and out of the reactive region has raised concern in the past [21]. This is because from one time step to the next there will inevitably be an abrupt change in the forces on that atom. In dynamics studies this discontinuity can influence the integration with respect to time and conceivably lead to nonphysical behavior. To circumvent this we use a more robust integration method (a sixth order predictor–corrector algorithm) that is not available in TINKER. This integration technique should help dampen out any fluctuations in the force/position changes that would lead to discontinuities when atoms are switched back and forth between the MM and QM regions.

Recently other groups have designed algorithms that include dynamically moving QM regions to model explicit solvent interactions. Morokuma et al. have extended their multilevel ONIOM technique to include the exchange of solvent (ONION-XS) molecules dynamically throughout the simulation [23]. They implemented a fifth order polynomial switching function to accommodate the instantaneous change in forces and potentials on atoms changing from one level to the next in the ONIOM regime. This technique extends the “Hot Spot” method that Rode et al. originally used [46] by not only smoothing out the differences in forces of atoms that are in a defined “buffer” region between the various levels, but also smoothing out the forces for all atoms in the QM region when any atom crosses in or out of the buffer region. In addition, the ONIOM-XS method uses the same switching function algorithm to smooth the potential energy for the atoms in the QM and buffer regions. However, knowing how large the buffer region needs to be for each chemical system is uncertain. Furthermore, it has been demonstrated that when multiple molecular groups are in the buffer region the ONIOM-XS method is no longer able to remove all discontinuities in the potential [23].

Truhlar and coworkers have developed an elaborate algorithm for smoothing the forces and the potential of the atoms switching regions in order to conserve both energy and momentum [21]. Their adaptive partitioning (AP) method is able to fix the conservation problem in a long time regime (400–1,000 ps), and they suggest using their permuted AP method when dealing with equilibrium conditions, a situation when energy or momentum drift could noticeably influence the physical or chemical behavior of the system. The permuted AP method does however include a number of multilevel computations at each time step (as opposed to just one) in order to assess how the potential smoothing functions are applied. This increases the number of computation by 2^N for each N groups in the buffer zone. By our estimates, in the squalane system ~ 4.4 carbon atoms would be in the buffer zone on average, thus increasing the QM calculations by >21 times and making the AP method unattractive when thousands of trajectories are needed.

In both of the algorithms just described, only small solvent molecules (e.g., H_2O , NH_3 , ...) are present in the simulations, and these molecules never straddle a boundary. The use of switching functions when an atom or small molecule is in a buffer region is convenient when the solvent molecule does not span from inside the QM region, through the buffer, and into the MM region. In this situation, the buffer region atoms are being given some MM character while still being bonded to atoms in the QM region, which could lead to extraneous forces on all the QM atoms in that molecule. This makes the inclusion of such algorithms to our current model nontrivial. Additionally, it must be noted that in our simulations we are not concerned with equilibrium conditions since the incident atom collides at the surface with 1–5 eV of translational energy. Furthermore, our simulation times are <10 ps, at which appreciable energy or momentum drift should be very minimal.

Though we had success with dynamic partitioning with squalane, when we began to study gas–liquid collision chemistry of room temperature ionic liquids, dynamic partitioning as we had done in the past became problematic. This is because the liquid, 1-ethyl-3-methyl-imidazolium nitrate or [emim][NO_3], is composed of cations and anions, and each of these cannot meaningfully be divided into parts in the QM and MM region. This would require some sort of charge partitioning algorithm, likely requiring charges to switch back and forth between regions as a function of time, resulting in extraneous forces. Because the liquid is ionic, it has slower diffusion and much higher density than the hydrocarbon liquids, preventing the surface from changing much over time. After testing numerous trajectory paths, we found that it was rare that the incident atom with an $E_T \leq 5.0$ eV had a direct interaction with more than the first two ion pair layers. Therefore we decided to fix the QM region to include five ion pairs (115 atoms) at or near the surface and the incident $\text{O}(^3\text{P})$ atom. In comparison to squalane, where there were often several well separated radical species, $\text{O}(^3\text{P}) + [\text{emim}][\text{NO}_3]$ can be treated with a more localized QM description.

3 Simulating Gas–Liquid Interactions

3.1 *Building a Model Liquid Surface*

Prior to the scattering of a gaseous atom off of a liquid surface, a properly equilibrated surface must be created. Initially, we have created a bulk model of the desired liquid with the TINKER program by using the OPLS all-atom force field [47] and the isobaric–isothermal (NPT) ensemble for ~ 1.0 ns at 400 K with periodic boundary conditions applied. This ensures that the fluctuations in bulk density have reached a minimum. Afterward, the final structure of this simulation is used as the initial structure for an additional NPT simulation which is cooled to 298 K (or some arbitrary experimental temperature) for ~ 0.5 ns at which the average bulk density is measured in comparison with physical property values. In our work with squalane ($\text{C}_{30}\text{H}_{62}$) the density was found to be 0.769 g/cm^3 , compared to the observed value of 0.815 g/cm^3 . To create a liquid surface, at this point the periodic boundaries in one direction are extended by a factor of three so that two empty vacuum regions are present on top and bottom of a slab of the equilibrated bulk liquid. This system is then run for ~ 1.0 ns with the NVT ensemble to allow for surface relaxation of the bulk liquid. After the surface density has equilibrated, random snapshots of the surface are used as the starting surface for the reactive scattering simulations. It is these surfaces that are used to compute the average surface density profiles, which enables us to analyze the types, locations, and orientations of different molecules (or parts of the molecule, like functional groups) within the liquid.

In our previous work, we were able to show that, even though the bulk squalane contains twice as many secondary carbons as primary or tertiary, at the surface the majority of the carbons are primary and stick up out of the surface (see Fig. 1). It is important to understand what types of atoms are on the surface as their reactivity is what drives the surface chemistry. Additionally, with our [emim][NO_3] surface we found that the anion is more abundant at the surface than the cation and that the ethyl-chains that lie on the surface tend to stick up towards the vacuum [37]. However, we found that there exists a slightly different surface topology for [emim][NTf_2], which includes a larger anion that keeps the ethyl chain from protruding out of the surface [36]. In contrast, the surface of [C_{12}mim][NTf_2] has long hydrocarbon chains sticking up into the vacuum, causing a noticeable difference in the reaction profile that occurs upon gaseous atom collision.

3.2 *Gas–Liquid Scattering Model*

In order to simulate gaseous atoms scattering from a liquid surface that represents a realistic sampling of the various chemical and surface features, various input conditions should be considered to enable a connection to the experimental conditions. The following conditions are what we have chosen for our models. After an

equilibrated surface has been generated, multiple snapshots that are well separated in time are used as the input for the scattering surface. Each of these snapshots represents a unique surface, and several such surfaces are sampled in the trajectory calculations. Because of the nature of the QM/MM calculation, periodic boundaries are not used, and therefore, to keep the surface density consistent over the time of the simulation, we fix the coordinates of atoms in the outer walls and base of the liquid as shown in Fig. 2 for O + [emim][NO₃]. The incident atom is directed to collide with the surface at locations that are chosen from a series of grid points spaced ~ 2.5 Å apart to ensure sampling of the various functional groups or atom types that are present at the surface. In order to understand the angular dependence of the scattering and not bias the incident atom to a particular angle of incidence, several azimuthal angles are sampled and results are considered from different incident polar angles (θ_i) relative to the surface normal. Figure 3 provides a schematic representation of this basic setup, which was utilized in our various gas-liquid scatter experiments. In the case of our O/Ar + [emim][NO₃] studies, ten unique surfaces, nine points on the surface grid, four azimuthal angles (0° , 90° , 180° , 270°), and four incident polar angles (0° , 30° , 45° , 60°) are considered for

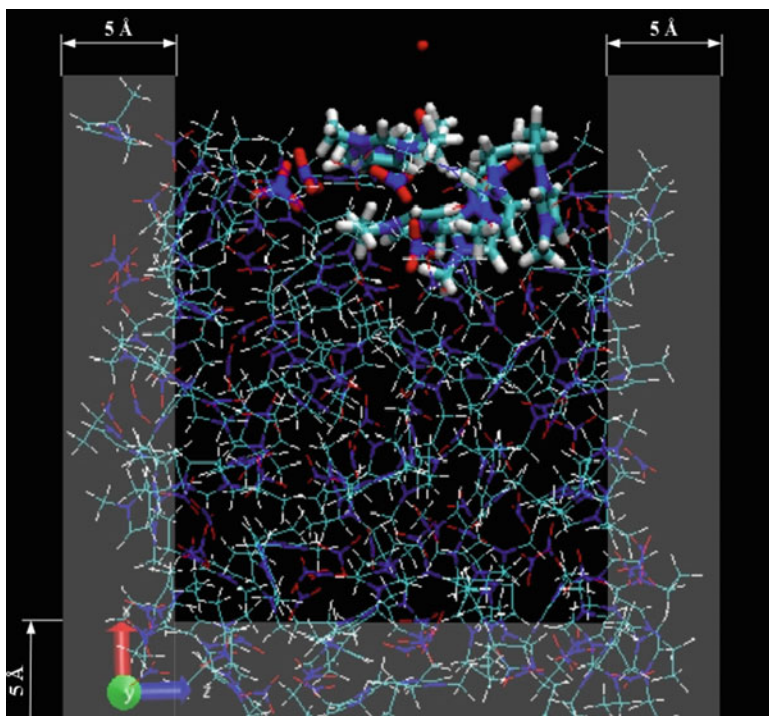
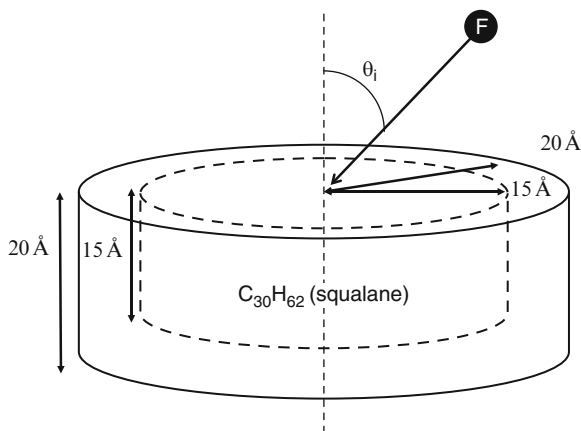


Fig. 2 Side view depiction of the $32.3 \times 34.5 \times 31.6$ Å³ QM/MM simulation box of O + [emim][NO₃]. The *sticks* represent the 2,093 atoms computed with MM and the *tubes* are the 116 atoms computed with QM. Atoms within the 5 Å thick shaded regions on the sides and bottom are the MM atoms kept fixed during the simulations to keep the liquid density from changing over time

Fig. 3 Schematic picture of the simulation region. The *outer region* is composed of fixed coordinate atoms while atoms in the *inner region* are allowed to move freely.

Arrows indicate the trajectory of the incident fluorine atom with angle of incidence θ_i



a total of 1,170 trajectories. For this chemical system there exist ~40–60% scattered nascent products, which provide a large enough sample set to compute statistical averages.

4 Application of Dynamic QM/MM to Gas–Liquid Scattering

4.1 Squalane

Squalane ($\text{C}_{30}\text{H}_{62}$) is a branched hydrocarbon with very low vapor pressure. It is a popular liquid for molecular beam experiments as the chance that atoms in the beam will collide with molecules evaporating from the liquid is sufficiently low; thus the beam atoms can move ballistically, and reactions will only occur at the interface. Experiments from Nesbitt et al. [34, 35] and Minton et al. [31, 32] have demonstrated that, with incident $\text{F}(^2\text{P})$ or $\text{O}(^3\text{P})$ atoms, there exist two distinct reactive processes with squalane. When the incident atom and nascent products have very few collision events and spend very little time (on the order of ps) in the liquid, the products leave the liquid with velocities and internal states that are largely reflective of the primary reactive event. For hyperthermal initial conditions, this mechanism is referred to as the hyperthermal desorption/scattering (HDS) component. On the other hand, the incident atom and nascent products may have many collisions, spending significant time in the liquid (ns– μs). In this case, the products leave the surface with comparable energy to the surface temperature, and the mechanism is referred to as the trapped desorption (TD) component. Due to current computing resource limitations, our current dynamic QM/MM can only simulate <10 ps of these reactions. Thus, any TD products coming from the liquid are missed; these products are still trapped in the liquid at the termination of the simulation.

We have completed two complementary gas/surface collision studies with squalane, using incident $O(^3P)$ with 5 and 1 eV and $F(^2P)$ with 1 and 0.5 eV initial translational energy [28, 29]. To understand the viability of using MSINDO for the QM method, we computed the reaction enthalpies and barriers for the H abstraction and elimination reactions of O/F with methane and ethane, and compared the results to those from the advanced ab initio method CCSD(T). These comparisons show that for the reaction barriers, which are important for the bond making/breaking events in our simulations, the MSINDO computations are comparable (~ 0.15 eV) to the CCSD(T) results. For the methane reaction, the H abstraction reaction ($O + CH_4 \rightarrow CH_3 + OH$) has an $\Delta E_{rxn} = -0.342$ eV whereas with F the reaction is more exothermic (as expected) with an $\Delta E_{rxn} = -1.418$ eV, based on MSINDO. Also, the reaction barriers are completely distinctive for these two reactions, with the barrier for $O + CH_4$ being 0.564 eV and that for $F + CH_4$ only 0.167 eV. The H elimination reaction ($O + CH_4 \rightarrow CH_3O + H$) has a quite different thermochemistry with a $\Delta E_{rxn} = 0.081$ eV while that for $F + CH_4$ is $\Delta E_{rxn} = -0.015$ eV; the reaction barriers are 1.869 and 1.769 eV, respectively. Generally, the MSINDO computed thermochemistry with ethane has slightly lower reaction barriers as compared to methane for both the H abstraction and elimination channels.

Considering the differences in thermochemistry between F and O reactions with methane and ethane, it is not surprising that in our F + squalane studies, with incident energies of 1.0 and 0.5 eV, no H elimination reactions are found; in contrast, in O + squalane with the incident energy of 5 eV the probability of an H elimination reaction is as high as 0.14. Additionally, in the O collision studies double H abstraction and C–C bond cleavage occur with probabilities as high as 0.22 and 0.06, respectively. Another interesting difference in these two reactions is in the probability that the product does not desorb from the surface by the end of the simulation time (~ 10 ps). Upon H absorption, HF has a desorption probability of 0.90–0.95 (dependent on incidence angle), while OH has a bit lower probability of 0.71–0.83. This may be in part due to the average depth of penetration of the incident atom before the reaction occurs, whereas the O average depth for H abstraction is ~ 0.5 Å deeper than for F. Also, HF is formed with a higher probability than OH, with reaction probabilities of 0.78 and 0.41, respectively, which is understandable given that the MSINDO reaction barriers for H abstraction (of methane or ethane) are ~ 3 – 4 times higher for O than F. However, the copious formation of HF does not necessarily translate into the efficient transfer of the incident energy into the product’s vibrational, rotational, and translational modes.

To examine these results in more detail, we now look at how the vibrational states of the HDS component for the H abstraction product are affected by the choice of F or O as the reactant. We compute these vibrational states based on the classical histogram method [48], wherein the vibrational quantum numbers (calculated from the vibrational action [49]) is rounded to the nearest integer to determine vibrational state populations. A similar method is used to define rotational quantum number starting from the classical rotational angular momentum.

The dynamic QM/MM calculations show that the nascent gaseous product HF leaves the surface with nearly all its vibrational distribution in either the

$\nu = 1$ or 2 state, and is split fairly equally between them. This is similar to what is seen for gas phase reactions of atomic fluorine with hydrocarbons, indicating that there is only modest vibrational relaxation as the HF exits the liquid. For the OH product, most of the vibrational distribution is in $\nu = 0$, with only a fraction (0.13) being in the $\nu = 1$ state. This reflects the smaller energy release in this reaction, together with inefficient conversion of reagent kinetic energy into product vibrational energy (as is also known from gas phase analogs). When the incident energy of O is lowered from 5 to 1 eV there is even less population of OH in the $\nu = 1$ state, with a probability of only 0.05. On the other hand, when changing the incident energy of F from 1.0 to 0.5 eV, there is a decrease in the $\nu = 0$ population and an increase in the $\nu = 2$ population. The vibrational distribution of HF with the lower incident energy of 0.5 eV is more closely aligned with the experimental results from Nesbitt et al. where the incident energy is ~ 0.08 eV. This indicates that the population of the product vibrational modes is not governed by the incident energy of the colliding gaseous atom, but is mostly due to the release of energy in the bond breaking event.

Some of the energy released during the reaction is also transferred into the rotational states of the nascent products. Figure 4 presents the distribution of rotational states of HF for the major vibrational states. For the $\nu = 1$ state of HF, the distribution of rotational states is largely unchanged by the incident energy, while for $\nu = 2$, there are slightly hotter rotational states from the 0.5 eV initial energy. A more drastic change in rotational distribution can be seen when changing

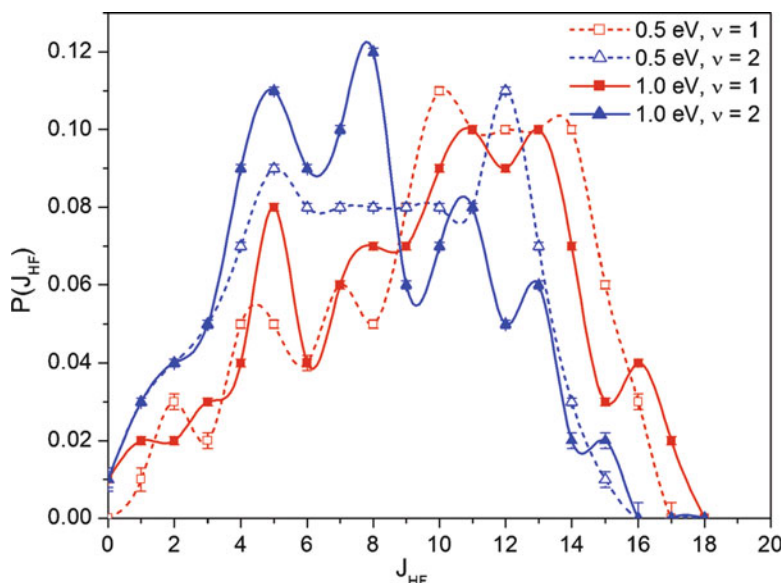


Fig. 4 Distribution of rotational states for the $\nu = 1$ (red) and $\nu = 2$ (blue) vibrational states of HF, as obtained from F + squalane simulations. Solid lines are for 1.0 eV and dashed lines are for 0.5 eV input translational energy

the incident energy of O from 5 to 1 eV. With 5.0 eV incident energy, the OH rotational distribution peaks around $J = 5$ –6 and has a bit of a Boltzmann-like distribution. On the other hand, with 1 eV incident energy the OH rotational distribution is hotter and peaks around $J = 10$ –12. This is further evidence that the transfer of energy into the product states is not governed by the incident energy.

Comparing translational energy distributions provides another contrast of the results from F vs O. Here we compare the H abstraction products (OH and HF) as well as the inelastic scattered atoms (O and F). For both inelastically scattered products there is a similar variation of final translation energy distribution with respect to angle of incidence. At 60° angle of incidence there is a broad distribution of translational energies, covering the range from almost zero energy to an energy equal to the incident energy (i.e., 1.0 eV for F and 5.0 eV for O) as shown in Figs. 5 and 6. By contrast, at 30° angle of incidence, most of the input translational energy is transferred into the liquid.

There is a notable difference in the final translation energy of the H abstraction products, OH and HF. The OH final translational energies are very similar to the O energies and still have a distinctive variation with respect to the incidence angle. In contrast, the HF product final translational energies are largely independent of

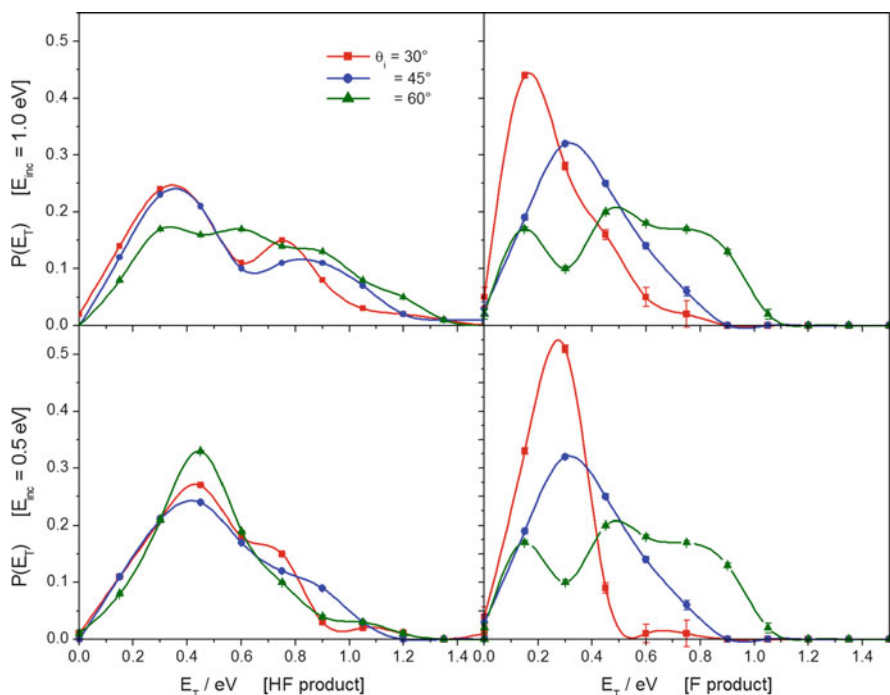


Fig. 5 Distribution of product translational energies in eV of HF (*left*) and scattered F (*right*) from incident translational energies 1.0 eV (*top*) and 0.5 eV (*bottom*) and for each angle of incidence 30° , 45° , and 60°

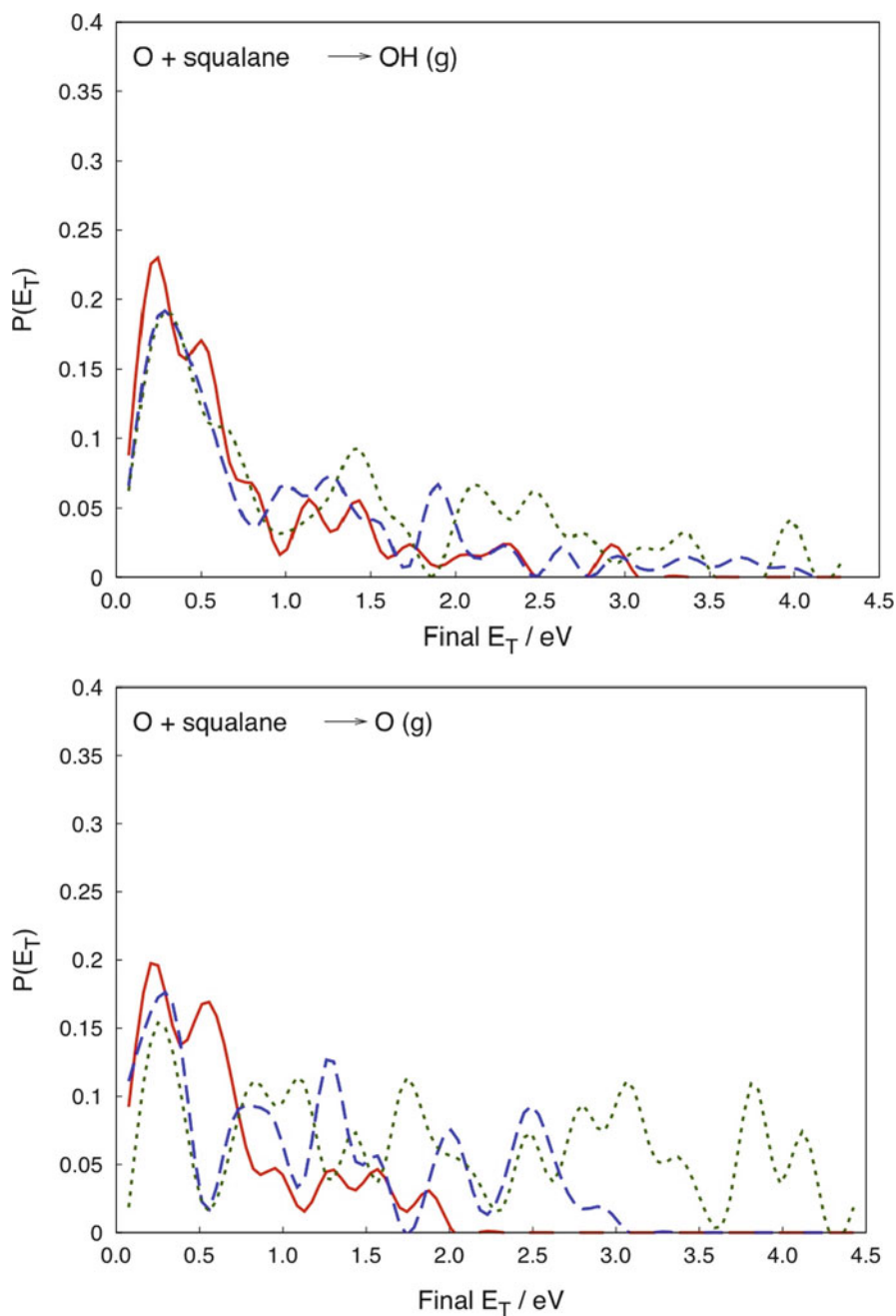


Fig. 6 Distribution of product translational energies in eV for OH (*top*) and scattered O (*bottom*) for O + squalane collisions with an incident translational energy of 5 eV and for each angle of incidence 30° (solid red), 45° (dashed blue), and 60° (dotted green)

incidence angle in comparison to the scattered F energies. Also, there is very little difference in the final translational energy of the HF product with regard to input translational energy. Furthermore, for the 0.5 eV incident energy it is even common that the HF product has more final translational energy than F had initially. This brings clarity to the notion that the molecular energy transferred to the product states is mainly from the chemical reaction that takes place and not from the collision energy.

4.2 [emim][NO₃]

We have also applied our dynamic QM/MM model to study the surface reactivity of room temperature ionic liquids (RTIL). These RTIL are quite unique liquids in that they are comprised solely of cation and anion pairs that can be interchanged with other types of cations or anions; therefore, their physical and chemical properties are a bit tunable. Generally these liquids are nonvolatile and nonflammable, chemically and thermally stable, and possess high ionic conductivity. Our interest in RTIL is in their use as environmentally conscious hypergolic bipropellants. Although these liquids are becoming more readily utilized as solvents, little is known about the surface reactivity of these liquids. Since we have had some success in analyzing the reactivity of squalane via gaseous atom/surface scattering, we extended our work to include a simple RTIL, 1-ethyl-3-methyl-imidazolium nitrate, referred to as [emim][NO₃]. This liquid can be viewed as three distinctive components – the NO₃ anion, the imidazolium ring cation, and the nonpolar hydrocarbon tail (ethyl-chain) – all with their own unique local chemistry.

The [emim][NO₃] RTIL was chosen because an OPLS-AA force field had already been developed for it [50], and it limits the chemical interactions to only C, H, O, and N, atoms for which we have confidence in using MSINDO to predict reasonably accurate thermochemistry of nitrogen-containing ring systems [51, 52]. In order to grasp the difference in the reactivity of the different surface components with each other. As opposed to the incident atom, we chose to compare the nonreactive scattering of Ar with the surface to the chemistry that results in reactive scattering with O(³P), both with an initial 5.0 eV of translational energy.

The gas–liquid scattering setup is very similar to that described with the squalane experiments in the previous section. However, as mentioned in Sect. 2.3, the QM region in these simulations was fixed to a localized region near the surface. Our surface analysis of [emim][NO₃] indicates that the ethyl-C sticks up out of the surface, followed by an even distribution of cations and anions. This means that there are readily accessible H atoms sticking up out of the surface for H abstraction or H elimination reactions to occur before the incident O crosses the surface threshold (the point where the density is equal to half the bulk density). Our study was done in tandem with Minton and coworker’s similar experimental study [36] that used the RTILs [emim][NTf₂] and [C₁₂mim][NTf₂], which they are able to obtain with high purity, unlike [emim][NO₃] at the time of their study.

In both of these liquids the computed surface topology demonstrated that the large anion and the ethyl-chain are equally present at the surface, while the longer C₁₂-chain extends well out of the surface, thus exposing a significant amount of accessible H to the incoming O atoms.

Upon collision with Ar, some translational energy is transferred into the [emim][NO₃] surface causing minor changes in the local chemistry as shown in Table 1. Nearly all of the incident Ar atoms scatter from the surface (a probability of 0.91) with a small fraction trapped at the end of the simulation time of 7.3 ps. The momentum transfer causes on average two to three proton transfers to occur between the imidazolium ring and the anion (see Table 2), thus neutralizing the

Table 1 Probabilities for scattering Ar and O(³P) from an [emim][NO₃] surface, including the distribution of each nascent product

	From Ar	From O(³ P)
Total ^a	0.17 ± 0.01	0.57 ± 0.02
Incident atom ^b	0.91 ± 0.03	0.23 ± 0.01
HNO ₃	0.06 ± 0.01	0.04 ± 0.006
emim ^c	0.10 ± 0.01	0.12 ± 0.01
<i>emim</i> + [NO ₃] ^d	0.01 ± 0.003	0.05 ± 0.006
O	–	0.13 ± 0.01
OH	–	0.04 ± 0.005
O ₂	–	0.02 ± 0.005
HNO ₂	–	0.05 ± 0.007
Me/Et fragment ^e	–	0.02 ± 0.005
IM-chain ^f	–	0.01 ± 0.003
IM-chain fragment ^g	–	0.04 ± 0.006

^aFor Ar this value includes the sum of all non-Ar products; for O, this includes all desorption products

^bIncludes only desorption products containing the incident atom

^cRefers to the neutral species from the cation [emim] after loss of H⁺

^dRefers to any pairs or groups derived from the imidazolium ring (cation or neutral) with the anion

^eMethyl or ethyl fragment from elimination of hydrocarbon from ring

^fThe unfolded imidazolium ring from ring scission – O addition

^gA piece of the unfolded ring from ring fragmentation – O addition

Table 2 Distribution of number of proton transfers and cation/anion pair recombinations

Proton transfer	Incident O	Incident Ar
0	0.12 ± 0.01	0.13 ± 0.01
1	0.29 ± 0.02	0.21 ± 0.01
2	0.35 ± 0.02	0.31 ± 0.02
3	0.20 ± 0.01	0.30 ± 0.02
4	0.04 ± 0.01	0.05 ± 0.01
5	0.00	0.00
[emim] + [NO ₃] → emim-ONO ₂ ^a	0.12 ± 0.01	0.12 ± 0.01

^aAn oxygen from NO₃ attaches to the imidazolium ring

cation/anion pair. Once neutralization occurs, it is inevitable that the neutral species will eventually desorb from the charged liquid. Instead of proton transfer, the surface collision can also supply enough energy to enable the anion to react with the cation. This bond usually occurs between an O in $[\text{NO}_3]$ and one of the C–H carbons on the imidazolium rings. It is likely that these recombinations with the cation and anion are only short lived in the actual system. However, within our simulation time, we found that the probability of these events was 0.12. When the proton transfer event occurs to a molecular species that is located at the surface edge, there is sometimes enough momentum to “push” the neutral species off of the surface, which is listed as “total” in Table 1. The probability for this event to happen to [emim] is 0.10 and for $[\text{NO}_3]$ is 0.06. There is only a very slight probability of 0.01 that more than one molecular species departs the surface together. Overall, for nonreactive scattering of Ar from the surface, little change occurs chemically to the ionic liquid molecules. This adds to knowledge concerning the chemical and thermal stability of RTIL.

The collision of a reactive atom, $\text{O}(^3\text{P})$, with [emim] $[\text{NO}_3]$ causes a multitude of chemical reactions to ensue at the surface. This is in contrast to the scenario that we just presented with the bombardment with Ar, a nonreactive atom, although there are some common features as well. The types of reactions that occur have been grouped into reaction types in Table 3. One of the most common reactions, with a probability of 0.43, is the elimination of NO_3 , which occurs when the incident O collides with $[\text{NO}_3]$ forming NO_2 or HNO_2 and O_2 or occasionally OOH . The second most prevalent reaction type, occurring with a probability of 0.21, is ring scission – O addition; this occurs when O addition causes a bond scission in the imidazolium ring, opening it up into an intact chain. About a quarter of the time there is enough vibrational energy in the chain that it splits apart creating the channel labeled ring fragmentation – O addition. Commonly, in both the ring scission reaction types, an aldehyde is formed. As with the O + squalane reactions, H abstraction also occurs in the O + [emim] $[\text{NO}_3]$ reaction yielding OH with a probability of 0.09. Note that this is much lower than the 0.41 that is found with

Table 3 Reaction probability from $\text{O}(^3\text{P}) + [\text{emim}][\text{NO}_3]$ by reaction type

Distribution of reactions by type	
O scattering	0.13 ± 0.01
O addition to NO_3	0.02 ± 0.004
Elimination of NO_3	0.43 ± 0.02
O addition to HC^{a}	0.01 ± 0.002
Elimination of HC from ring	0.04 ± 0.01
Ring scission – O addition ^b	0.21 ± 0.01
Ring fragmentation – O addition ^c	0.05 ± 0.01
H abstraction	0.09 ± 0.01
NO_3 substitution	0.01 ± 0.003
HC substitution – H elimination	0.02 ± 0.004

^aHC refers to the ethyl or methyl hydrocarbon on the imidazolium ring

^bAfter ring scission the [emim] becomes an intact chain

^cAfter becoming a chain, fragmentation occurs

squalane, which makes sense since the ionic liquid has a smaller fraction of abstractable hydrogens. There is a small chance, 0.13, that the O does not react and bounces back from the [emim][NO₃] surface; inelastic collisions of this nature with O + squalane occur more often, with an overall average probability of 0.23.

In the O + [emim][NO₃] collisions, more than half of the products desorb from the surface during our simulation time, as shown in Table 1. This is in contrast to the Ar collision study where the probability of non-Ar containing products was 0.17. As with the nonreactive study, the O collisions also cause one to three proton transfers (see Table 2). Interestingly, a similar portion of HNO₃, emim, and *emim* + [NO₃] desorb from the [emim][NO₃] surface as compared to the nonreactive study; HNO₂ is an additional product found in the O scatter data. It is likely that the proton transfer event is responsible for these species desorbing from the [emim][NO₃] surface. The H abstraction channel, which occurs predominantly at the ethyl or methyl sites on the imidazolium ring, forms OH and has a desorption probability of 0.04. Nearly half of all the OH created desorbs from the surface by the end of the simulation (7.3 ps). In squalane, around 80% of the OH exits the surface. This is reflective of the differences in the surface density and the amount of diffusion that is possible within these two liquids. It must be noted that no charged species were found to desorb, which is not surprising because they would inevitably be attracted back to either the cations or anions at the surface.

In connection with the experimental observables that Minton et al. have measured for the similar reaction (O + [emim][NTf₂]), we also report the translational energy distribution of the inelastically scattered atoms as well as of the major molecular products. Figure 7 provides a basis for understanding how the [emim][NO₃] liquid is

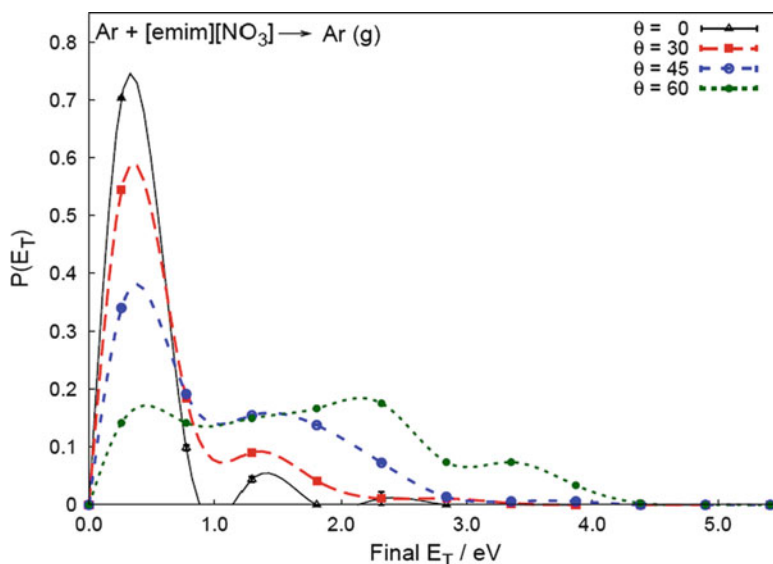


Fig. 7 Distribution of the final translational energy (in eV) of Ar for each angle of incidence ($\theta_i = 0^\circ, 30^\circ, 45^\circ$, and 60°)

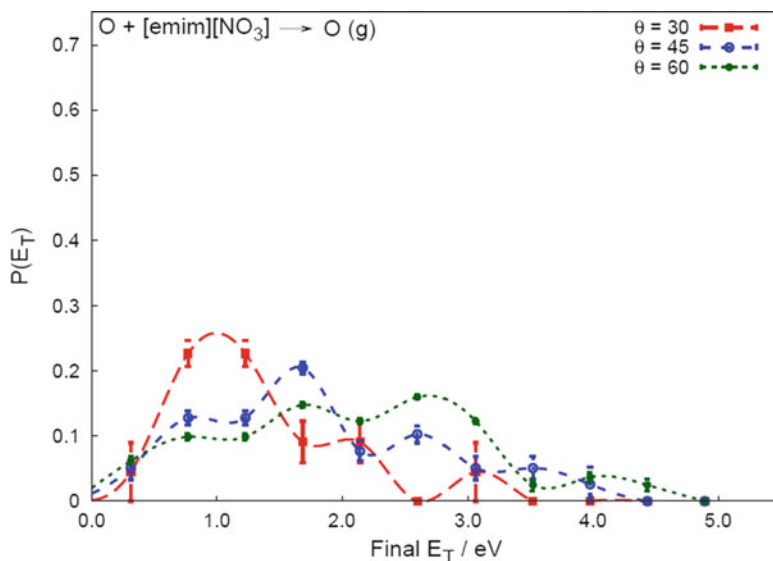


Fig. 8 Distribution of the final translational energy (in eV) of $\text{O}(^3\text{P})$ for each angle of incidence ($\theta_i = 0^\circ, 30^\circ, 45^\circ$, and 60°)

able to absorb energy upon bombardment by a nonreactive atom. Overall there is a broad distribution of final translational energies from the more glancing angle of 60° , and on average about half the 5.0 eV initial energy is transferred to the liquid. For more direct angles of incidence, nearly all the final translational energy is below 1.0 eV, indicating that multiple collisions play an important role. Figure 8 shows the corresponding results for inelastically scattered O atoms. (No 0° data are presented because almost all the incident Os undergo reaction for this angle.) For an incident angle of 60° , there is a very similar distribution of final translational energies for O as was found with Ar. This is likely due to the fact that at this glancing angle there is very little penetration of the surface ($0.94 \text{ \AA}$ for Ar and $-0.68 \text{ \AA}$ for O on average), thus minimizing the number of collisions and leaving the incident atom mostly unchanged. For the more direct collisions at 30° and 45° , the scattered O has more translational energy than Ar, as if the larger, heavier Ar atom is a “softer” atom losing more energy to the liquid. The distribution of final translational energy from the OH desorption product is depicted in Fig. 9. There is a slight decrease ($\sim 0.5 \text{ eV}$) in the translational energies of the OH product in comparison to the inelastically scattered O, with similar distributions between the various angles of incidence. Most of the OH products are from H abstraction at the C–H groups on the ethyl and methyl substituents on the imidazolium ring, which are sticking up out of the surface. On average, at 60° incidence angle, both O and OH scatter from [emim][NO_3], as we found with our QM/MM–MD model, with about half the initial translational energy; this coincides with the experimentally determined final translational energy of O and OH scattering from [emim][NTf_2]. Other nascent products, which are not shown

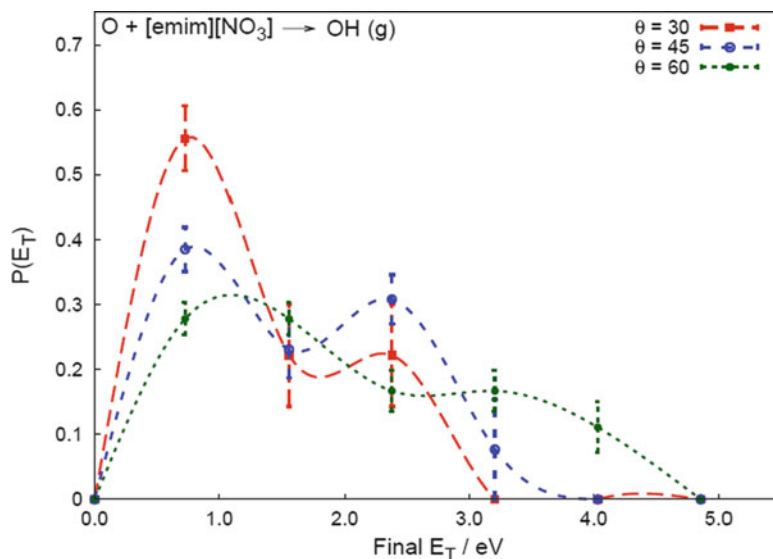


Fig. 9 Distribution of the final translational energy (in eV) of OH for each angle of incidence ($\theta_i = 0^\circ, 30^\circ, 45^\circ$, and 60°)

here, are HNO_2 and HNO_3 , both of which have smaller and much narrower distributions of translational energies and are largely independent of incidence angle. It is thought that this is because the reactive channels that govern these products involve a proton transfer event, and are not dependent directly upon a primary collision event.

5 Summary

In this chapter we reviewed the details of our theoretical methods that are used to describe gaseous atoms scattering from liquid surfaces. Scattering experiments of this type allow for the direct study of the surface reactivity of liquids, including studies of the partitioning of molecular energy transfer into various vibrational, rotational, and translational modes of the scattered products. One of the complexities in modeling such a dynamic system at hyperthermal energies is that many chemical changes can occur (i.e., the making/breaking of several bonds), and a priori knowledge of when and where this will occur is not easy to estimate in the absence of a simulation. Additionally, the spatial length scale needed to describe gas–surface reactions is larger than what can normally be computed purely with *ab initio* QM. Thus we use a hybrid approach that partitions the system into reactive and nonreactive regions to be treated by a dynamics QM/MM approach.

The nonreactive regions (thousands of atoms) were described with predetermined force fields (OPLS-AA) using MM, and the reactive regions (hundreds of

atoms) were computed with semiempirical QM (MSINDO), an efficient approach to describe bond-breaking for the reactions we are interested in. The coordinates and forces in our simulations are propagated in time using a rather robust fifth/sixth order predictor–corrector scheme. One unique aspect of our QM/MM algorithm is that we allow for a dynamic partitioning of atoms to be treated either with MM or QM, which we refer to as the “seed atom” method. Though we took a simplistic approach to the dynamic partition of atoms, we found this to be very useful in studying a surface like squalane that is very porous, allowing the incident atoms lots of freedom to move around within the liquid surface. We further discussed in Sect. 2.3 some of the difficulties involved in reassigning atoms in and out of the reactive region, as this changes the forces computed from one step to the next with a different level of theory.

Thus far, we have applied our QM/MM–MD model to study reactions at the surface of a well known hydrocarbon liquid, squalane, and we have also provided perhaps the first theoretical study of gaseous atom scattering from an RTIL surface. With squalane we observed the reactivity of both $\text{O}(^3\text{P})$ and $\text{F}(^2\text{P})$ scattering from the surface at a few different initial translational energies. The chemistry that ensued upon bombardment with $\text{F}(^2\text{P})$ with 1.0 and 0.5 eV incident energy was limited to H abstraction. However, when $\text{O}(^3\text{P})$ hits the surface of squalane at 5 eV, it is capable not only of H abstraction but also of H elimination and C–C bond scission. Even though understanding the barriers to these reactions helps determine whether there is enough initial translation energy to overcome these barriers, it is only through the dynamics simulations that we gain an insight into the mechanism of these reactions and the likelihood of these reactions occurring as a function of position relative to the liquid interface. One major theme that is noted throughout the discussion in Sect. 4.1 is that the molecular energy (be it vibrational, rotational, or translational) of the nonthermal nascent products emerging from the surface is predominantly due to the bond making/breaking event and is largely independent of the incident atom’s energy and collision angle. This indicates that nonthermal products are incompletely relaxed as they emerge from the liquid surface, even when they are produced a few angstroms beneath the surface.

Probing the reactivity of our RTIL surface of [emim][NO_3] was done with both Ar and $\text{O}(^3\text{P})$ at 5 eV to study both the nonreactive and reactive scattering. The Ar scattering shows that proton transfers can occur between the cations and anions, and there exists a small probability that the neutral species emim or HNO_3 can subsequently desorb during our simulation time of 7.3 ps. Desorption of neutral emim and HNO_3 can also occur in the $\text{O}(^3\text{P})$ scattering simulations and other neutrals such as HNO_2 and O_2 can also be produced. Bombardment of the RTIL with $\text{O}(^3\text{P})$ can also produce many other chemical species. Consistent with the Minton et al. experimental studies, we find that the abstraction product OH desorbs from the RTIL, resulting in an average translational energy that is about half of the input energy at an angle of incidence of 60° , and even less when the angle of incidence is more direct.

The QM/MM applications that we have considered have focused on hyperthermal chemistry for the most part, and this is a natural direction for this research due

to the short time (few ps) and small spatial regions (100–200 quantum atoms) required for describing the hyperthermal dynamics. Even with this limitation, there are a growing number of applications accessible to this method, as high energy atomic sources provide useful information about the structure and reactivity of liquid surfaces. The same technology can also be used to describe gas–solid reactions, as reviewed elsewhere [27], and in this case it is relevant to a number of important industrial etching processes as well as space materials research. Thermal reactions at gas–liquid interfaces can also be described, but the computational effort can easily get out of hand unless one implements additional components in the calculations (such as umbrella sampling so that the reacting species are initiated in close contact). Ultimately one needs to separate the reactive event for such problems from adsorption and diffusion, and there are serious questions as to whether a dynamic QM/MM approach is going to be the most useful way to describe the dynamics.

Acknowledgments This research was supported by AFSOR Grant FA9550-10-1-0205 and by the CENECI CCI NSF grant CHE-0943639.

References

1. Acevedo O, Jorgensen WL (2010) Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. *Acc Chem Res* 43:142–151
2. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W (2010) P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem Rev* 110:949–1017
3. Higashi M, Truhlar DG (2008) Electrostatically embedded multiconfiguration molecular mechanics based on the combined density functional and molecular mechanical method. *J Chem Theory Comput* 4:790–803
4. Kamerlin SCL, Haranczyk M, Warshel A (2009) Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pK(a), redox reactions and solvation free energies. *J Phys Chem B* 113:1253–1272
5. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem* 48:1198–1229
6. Cui Q, Karplus M (2000) Molecular properties from combined QM/MM methods. 2. Chemical shifts in large molecules. *J Phys Chem B* 104:3721–3743
7. Cui Q, Karplus M (2000) Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations. *J Chem Phys* 112:1133–1149
8. Gao JL, Truhlar DG (2002) Quantum mechanical methods for enzyme kinetics. *Annu Rev Phys Chem* 53:467–505
9. 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* 109:9082–9095
10. Lin H, Truhlar DG (2005) Redistributed charge and dipole schemes for combined quantum mechanical and molecular mechanical calculations. *J Phys Chem A* 109:3991–4004
11. 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* 109:17715–17733
12. Shurki A, Warshel A (2003) Structure/function correlations of proteins using MM, QM/MM, and related approaches: methods, concepts, pitfalls, and current progress. In: *Adv Protein Chem* 66:249–313

13. Bakowies D, Thiel W (1996) Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J Phys Chem* 100:10580–10594
14. Slavicek P, Martinez TJ (2006) Multicentered valence electron effective potentials: a solution to the link atom problem for ground and excited electronic states. *J Chem Phys* 124:084107
15. 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 $\text{Pt}(\text{P}(\text{t-Bu})(3))_2 + \text{H}_2$ oxidative addition. *J Phys Chem* 100:19357–19363
16. Vreven T, Morokuma K (2000) On the application of the IMOMO (integrated molecular orbital plus molecular orbital) method. *J Comp Chem* 21:1419–1432
17. Shoemaker JR, Burggraf LW, Gordon MS (1999) SIMOMM: an integrated molecular orbital/molecular mechanics optimization scheme for surfaces. *J Phys Chem A* 103:3245–3251
18. Rode B, Hofer TS, Randold BR, Schwenk CF, Xenides D, Vchirawongkwin V (2006) Ab initio quantum mechanical charge field (qmcf) molecular dynamics: a QM/MM-MD procedure for accurate simulations of ions and complexes. *Theor Chem Acc* 115:77–85
19. Lin H, Truhlar DG (2007) QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc* 117:185–199
20. Rappe AK, Goddard WA III (1991) Charge equilibration for molecular dynamics simulations. *J Phys Chem* 95:3358–3363
21. Heyden A, Truhlar DG (2007) Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *J Phys Chem A* 111:2231–2241
22. Rohrig UF, Frank I, Hutter J, Laio A, VandeVondele J, Rothlisberger U (2003) QM/MM car-parrinello molecular dynamics study of the solvent effects on the ground state and on the first excited singlet state of acetone in water. *ChemPhysChem* 4:1177–1182
23. Kerdcharoen T, Morokuma K (2002) ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem Phys Lett* 355:257–263
24. Bolten K, Hase WL, Doubleday C Jr (1999) A QM/MM direct dynamics trajectory investigation of trimethylene decomposition in an argon bath. *J Phys Chem B* 103:3691–3698
25. Stanton CL, Kuo I-FW, Munday CJ, Laino T, Houk KN (2007) QM/MM metadynamics study of the direct decarboxylation mechanism for orotidine-5'-monophosphate decarboxylase using two different QM regions: acceleration too small to explain rate of enzyme catalysis. *J Phys Chem B* 111:12573–12581
26. Kerdcharoen T, Liedl KR, Rode BM (1996) A QM/MM simulation method applied to the solution of Li^+ in liquid ammonia. *Chem Phys* 211:313–323
27. Troya D, Schatz GC (2004) Hyperthermal chemistry in the gas phase and on surfaces: theoretical studies. *Int Rev Phys Chem* 23:341–373
28. Kim D, Schatz GC (2007) Theoretical investigation of hyperthermal reactions at the gas-liquid interface: O (3P) and squalane. *J Phys Chem A* 111:5019–5031
29. Radak B, Yockel S, Kim D, Schatz GC (2008) Modeling reactive scattering of F(2P) at a liquid squalane interface: a hybrid QM/MM molecular dynamics study. *J Phys Chem A* 113:7218–7226
30. Garton DJ, Minton TK, Alagia M, Balucani N, Casavecchia P, Gualberto Volpi G (2000) Comparative dynamics of Cl(2P) and O(3P) interactions with a hydrocarbon surface. *J Chem Phys* 112:5975–5984
31. Zhang J, Garton DJ, Minton TK (2002) Reactive and inelastic scattering dynamics of hyperthermal oxygen atoms on a saturated hydrocarbon surface. *J Chem Phys* 117:6239–6251
32. Zhang J, Upadhyaya HP, Brunsvold AL, Minton TK (2006) Hyperthermal reactions of O and O_2 with a hydrocarbon surface: direct C-C bond breakage by O and H-atom abstraction by O_2 . *J Phys Chem B* 110:12500–12511
33. Perkins BG Jr, Nesbitt DJ (2006) Quantum-state-resolved CO_2 scattering dynamics at the gas-liquid interface: incident collision energy and liquid dependence. *J Phys Chem B* 110:17126–17137

34. Zolot AM, Harper WW, Perkins BG, Dagdigian PJ, Nesbitt DJ (2006) Quantum-state resolved reaction dynamics at the gas-liquid interface: direct absorption detection of HF(v, J) product from F(2P) + squalane. *J Chem Phys* 125:021101/1–021101/4
35. Zolot AM, Dagdigian PJ, Nesbitt DJ (2008) Quantum-state resolved reactive scattering at the gas-liquid interface: F + squalane (C₃₀H₆₂) dynamics via high-resolution infrared absorption of nascent HF(v, J). *J Chem Phys* 129:194705/1–194705/11
36. Wu BH, Zhang JM, Minton TK et al (2010) Scattering dynamics of hyperthermal oxygen atoms on ionic liquid surfaces: emim NTf₂ and C(12)mim NTf₂. *J Phys Chem C* 114: 4015–4027
37. Yockel S, Schatz GC (2010) Modeling O(3P) and Ar scattering from the ionic liquid [emim][NO₃] at 5 eV with hybrid QM/MM molecular dynamics. *J Phys Chem B* 114:14241–14248
38. Krebs T, Nathanson GM (2010) Reactive collisions of sulfur dioxide with molten carbonates. *PNAS* 107:6622–6627
39. Muentner AH, DeZwaan JL, Nathanson GM (2007) Interfacial interactions of DO with salty glycerol solutions of KI, NaI, LiI, and NaBr. *J Phys Chem C* 111:15043–15052
40. Waring C, Bagot PAJ, Slattery JM, Costen ML, McKendrick KG (2010) O(P-3) atoms as a probe of surface ordering in 1-alkyl-3-methylimidazolium-based ionic liquids. *J Phys Chem Lett* 1:429–433
41. Waring C, Bagot PAJ, Slattery JM, Costen ML, McKendrick KG (2010) O(P-3) atoms as a chemical probe of surface ordering in ionic liquids. *J Phys Chem A* 114:4896–4904
42. Jay William Ponder (2010) TINKER – software tools for molecular design. Accessed at <http://dasher.wustl.edu/tinker/>
43. Bredow T, Geudtner G, Jug K (2006) MSINDO (version 3.2)
44. Ahlswede B, Jug K (1999) Consistent modifications of SINDO1: I. approximations and parameters. *J Comp Chem* 20:563–571
45. Ahlswede B, Jug K (1999) Consistent modifications of SINDO1: II. Applications to first and second row elements. *J Comp Chem* 20:572–578
46. Kerdcharoen T, Rodes BM (2000) What is the solvation number of Na + in ammonia? An ab initio QM/MM molecular dynamics study. *J Phys Chem A* 104:7073–7078
47. 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* 118:11225–11236
48. Bowman JM, Kuppermann A (1971) Classical and quantum reaction probabilities and thermal rate constants for the collinear H + H₂ exchange reaction with vibrational excitation. *Chem Phys Lett* 12:1–4
49. Porter RN, Raff LM, Miller WH (1975) Quasiclassical selection of initial coordinates and momenta for a rotating morse oscillator. *J Chem Phys* 63:2214–2218
50. Canongia Lopes JN, Deschamps J, Padua AAH (2004) Modelling ionic liquids using a systematic all-atom force field. *J Phys Chem B* 108:2038–2047
51. Jug K, Chiodo S, Calaminici P, Avramopoulos A, Papadopoulos MG (2003) Electronic and vibrational polarizabilities and hyperpolarizabilities of azoles: a comparative study of the structure – polarization relationship. *J Phys Chem A* 107:4172–4183
52. Raabe G, Wang Y, Fleischhauer J (2000) Calculation of the proton affinities of primary, secondary, and tertiary amines using semiempirical and ab initio methods. *Z Naturforsch A* 55:687–694

Multiscale Modelling in Computational Heterogeneous Catalysis

F.J. Keil

Abstract The goal of multiscale modelling of heterogeneous catalytic reactors is the prediction of all steps, starting from the reaction mechanism at the active centre, the rates of reaction, adsorption and diffusion processes inside the porous system of the catalyst support, based on first principles, quantum chemistry, force field simulations and macroscopic differential equations. The progress in these fields of research will be presented, including linking models between the various levels of description. Alkylation of benzene will be used as an example to demonstrate the various approaches from the active centre to the reactor.

Keywords Computational heterogeneous catalysis · Molecular dynamics · Monte Carlo · Multiscale modelling · Quantum chemistry

Contents

1	Introduction	69
2	Computational Approaches	73
2.1	Quantum Chemistry	73
2.2	Monte Carlo, Molecular Dynamics	81
2.3	QM/MM, Transition State Theory	89
3	Applications: From the Active Centre to the Chemical Reactor	93
4	Outlook	98
	References	99

1 Introduction

Catalysis is the study of materials that can accelerate reactions and control reaction mechanisms. After a catalytic cycle, namely adsorption of reactants, reactions and desorption of products, the catalyst is restored to its initial state. An ideal catalyst

F.J. Keil
Chemical Reaction Engineering, TU Hamburg-Harburg, Eissendorfer Str. 38, 21073 Hamburg,
Germany
e-mail: keil@tu-harburg.de

for a given reaction should convert the reactants to the desired products with no production of undesired by-products, which means it should have a high selectivity. Furthermore, the catalyst should accelerate the rates of reactions by orders of magnitude. In comparison with the reaction without using a catalyst, the catalyst, in general, changes the reaction mechanisms. In reality these ideal requirements cannot be fully met, but one has to accept compromise on these requirements. The challenge of catalyst development is to find materials that will come as close to optimal performance as possible. At present, catalyst development is in essence an experimental discipline, but computational approaches play an ever increasing role in catalysis research. The final goal is the design of a catalyst on a rational basis. Currently, computational catalysis can already give deep insight into the details of catalytic reaction mechanisms, diffusion inside catalyst supports, and adsorption processes which cannot be obtained by experiments only. The most useful approach is a combination of experiments, like various spectroscopic tools, gas chromatography, mass spectrometry and computational methods. Massively parallel computers, quantum chemical software and classical molecular simulation methods have considerably expanded the classes of catalytic problems that can be treated on the basis of first principles approaches.

In the present article multiscale computation in heterogeneous catalysis in porous catalyst supports will be described. Multiscale modelling is the field for solving physical problems which have important features at multiple spatial and/or temporal scales. An important problem is the scale linking. Catalysis spans a significant range of length and time scales. The active centre can be composed from a single metal atom or acid site to a few dozen metal atoms. The active sites are inside porous materials which may be crystalline [e.g. zeolites, metal-organic frameworks (MOFs)] or amorphous, like silicas or aluminas. The reactants have to adsorb onto the porous materials, followed by a diffusion process inside the pores to the active centres. There, the catalysed reactions occur, followed by diffusion of the products to the surface of the porous material. The reactions have to be computed by quantum mechanical methods as chemical bonds are broken and formed. This is a genuine quantum mechanical phenomenon on the nanoscale with respect to the length and the femtosecond time scale. Most existing codes employ a form of the Kohn–Sham density function theory [1–3]. For the case that van der Waals (vdW) forces play an important role, more advanced approaches like Møller–Plesset-2 or even coupled cluster (CC) methods have to be used [4–6].

The reaction pathways, and in particular the transition states, have to be calculated followed by the determination of all the rate constants, employing the transition state theory (TST) [7–9]. Here, chemical knowledge and intuition is needed to include all significant reaction steps. Finding transition state structures (first-order saddle points) in high-dimensional systems can be a difficult problem. According to the experience of the present author's group, a combination of a modified nudged-elastic-string approach [10] and the so-called dimer method [11] turned out to be most useful for this purpose. These approaches are installed in various commercial programs like Vienna Ab-initio Simulation Package (VASP) [12] and QChem [13]. Another method which might be employed, either as a stand-alone program or

in combination with the two routines just mentioned, is the partitioned rational function optimisation (P-RFO) [14]. One has to make sure that the reactants and products are connected via the transition structure by following pathways downward along the potential energy surface (PES). One should keep in mind that the kinetics of reacting systems are mostly governed by the free energy of populations of molecules, and not the potential energy of single molecules. The individual molecules within the population may be exchanging energy with one another to rise and fall relative to the PES, but the net distribution remains determined by temperature. Many molecules of the population may pass to the left or right of the transition state structure. The transition state structure, a stationary point, should not be confused with the transition state which has $3N-7$ degrees of freedom for an N -atom system. The transition state is also called “activated complex”. A relationship between the properties of the activated complex and reaction kinetics is obtained by means of TST [7–9]. Here, statistical thermodynamics comes into play as one has to calculate partition functions of the reactants and transition state structure [15–18]. In the simplest version of TST the rigid-rotor-harmonic-oscillator approximation is employed. The kinetic constants have to be inserted into proper kinetic expressions to calculate the rates of reaction. Comparisons of experimental kinetic data and their computed values have to be made under consistent conditions. The canonical TST may be improved by variationally moving the reference position along the minimum energy path (MEP) that is used for the computation of the activated complex free energy, either backwards or forwards from the TS structure, until the rate is minimised (variational transition-state theory, VTST) [19]. Quantum effects, like tunnelling, can also be included. Non-adiabatic effects can, at present, only be calculated for small systems. Hopping probabilities between different PES can sometimes approximately be calculated by the Landau–Zener model [20]. The kinetic constants have to be inserted into proper kinetic expressions to calculate the rates of reaction.

The adsorption of the reactants and their diffusion into the pores of the catalyst support is calculated by Monte Carlo (MC) methods [21–25] and Molecular Dynamics (MD) [21, 23, 26–28]. The fluid dynamics around the porous particles inside a channel reactor may be described by various computational fluid dynamics (CFD) approaches [29–31]. In Fig. 1 the hierarchy of modelling approaches is presented.

As mentioned above, the limited time and length scales accessible with quantum chemistry calculations require the use of a modelling hierarchy as given in Fig. 1. This hierarchy is referred to as “multiscale approach”. This term is not restricted to a specific set of methods employed for calculating different time and length scales, but it is also used in various disciplines for a variety of modelling approaches [32–39]. In the present chapter the term is used for a modelling approach starting with quantum chemical calculations at the active centre of the catalyst, followed by MC and MD simulations of the multi-component adsorption and diffusion of the reactants and products. The concentration profiles inside the pores of the pellet can be calculated by partial differential equations whereby the data from the previous molecular simulations are included. A protruding problem in multiscale modelling

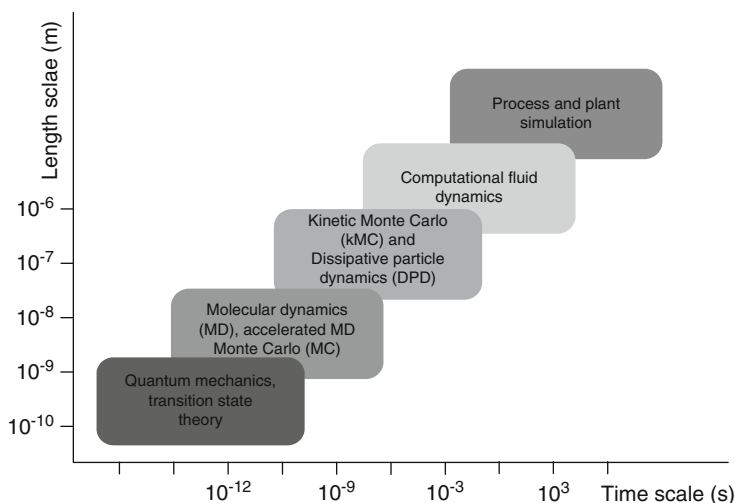


Fig. 1 Modelling hierarchy

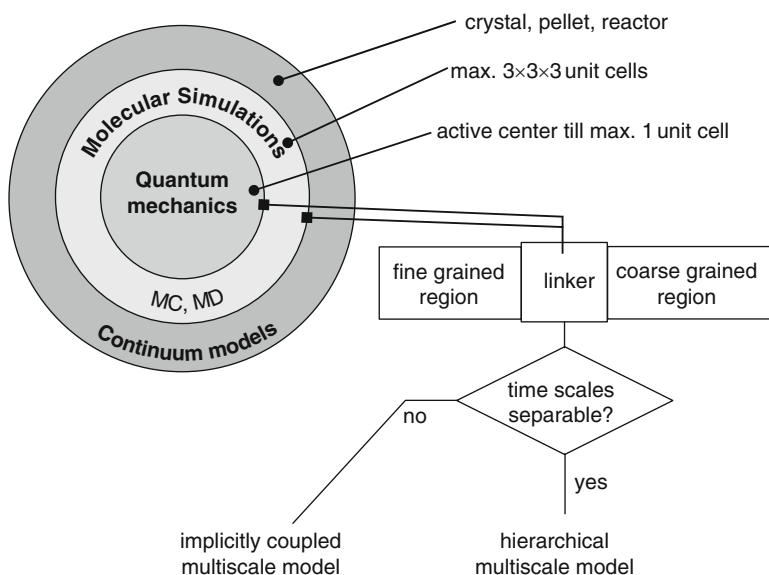


Fig. 2 Multiscale modelling

is the linking of the various levels of the model hierarchy. One needs a linking model which can include the data from a higher level of simulation in a well-founded way (see Fig. 2).

In the following paragraphs a short presentation of the approaches employed is given accompanied by a demonstration of a multiscale modelling in heterogeneous

catalysis. Compilations of density functional theory (DFT) applications to catalysis may be found in the books by van Santen and Neurock [40] and van Santen et al. (eds.) [41].

2 Computational Approaches

Some of the most widely used computational approaches will be briefly described below, namely some quantum chemical methods, classical simulations by Monte Carlo and Molecular Dynamics techniques and a few mesoscale methods.

2.1 Quantum Chemistry

In this section a recapitulatory description of the most common ab initio and density functional approaches will be presented. Ab initio methods calculate the electron properties of atoms and molecules at the absolute temperature ($T = 0$). The starting point is in most cases the non-relativistic Schrödinger equation

$$H\psi = E\psi, \quad (1)$$

where H is the Hamiltonian operator which is given by

$$H = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i < j} \frac{e^2}{r_{ij}} + \sum_{k < l} \frac{e^2 Z_k Z_l}{r_{kl}}, \quad (2)$$

where i and j run over electrons, k and l run over nuclei, $\hbar$ is Planck's constant divided by 2π , m_e is the mass of the electron, m_k is the mass of the nucleus k , ∇^2 is the Laplacian operator, e is the charge on the electron, Z is an atomic number and r_{ab} is the distance between particles (electrons or nuclei) a and b . The wave function ψ is thus a function of $3n$ coordinates where n is the total number of particles (electrons and nuclei).

The Born–Oppenheimer approximation leads to the electronic Schrödinger equation

$$H_{el}\psi_{el} = E_{el}\psi_{el}. \quad (3)$$

PES can only be computed by employing the Born–Oppenheimer approximation. The variational principle leads to an upper bound of the electronic energy

$$\frac{\int \psi_{el} H \psi_{el} d\mathbf{r}}{\int \psi_{el}^2 d\mathbf{r}} \geq E_0. \quad (4)$$

If we consider only the nuclei of a molecule in fixed positions and just one electron in the field of the nuclei, the eigenfunctions of (3) are called molecular orbitals (MOs). In quantum chemistry an MO is often expanded in a sum of atomic orbitals, for example, the familiar hydrogenic atomic orbitals, ϕ_{A0} $1s$, $2s$, $2p$, $3s$, $3p$, $3d$, etc.:

$$\psi_{el,k} = \sum_{i=1}^N a_{i,k} \phi_{A0}. \quad (5)$$

The Hartree–Fock (HF) method is widely used in electronic structure calculations, which is based on the following assumptions: (1) Born–Oppenheimer approximation, (2) the many-electron Hamiltonian is replaced with an effective one-electron Hamiltonian which acts on orbitals (one-electron wave functions), (3) the Coulomb repulsion between electrons is represented in an averaged way.

In practice the HF method is applied in the form of so-called Roothaan equations (see [4]):

$$\mathbf{F}\mathbf{C}_k = \varepsilon_k \mathbf{S}\mathbf{C}_k, \quad (6)$$

where $\mathbf{F}$ is the Fock matrix and $\mathbf{S}$ contains the overlap elements between the basis set functions. The term $\mathbf{C}$ is a square matrix of the basis set expansion coefficients, and ε represents a diagonal matrix of the orbital energies.

The HF iteration scheme is given in Fig. 3a. The HF wave function $|\text{HF}\rangle$ is thus the variationally best one – determinant approximation to the electronic ground

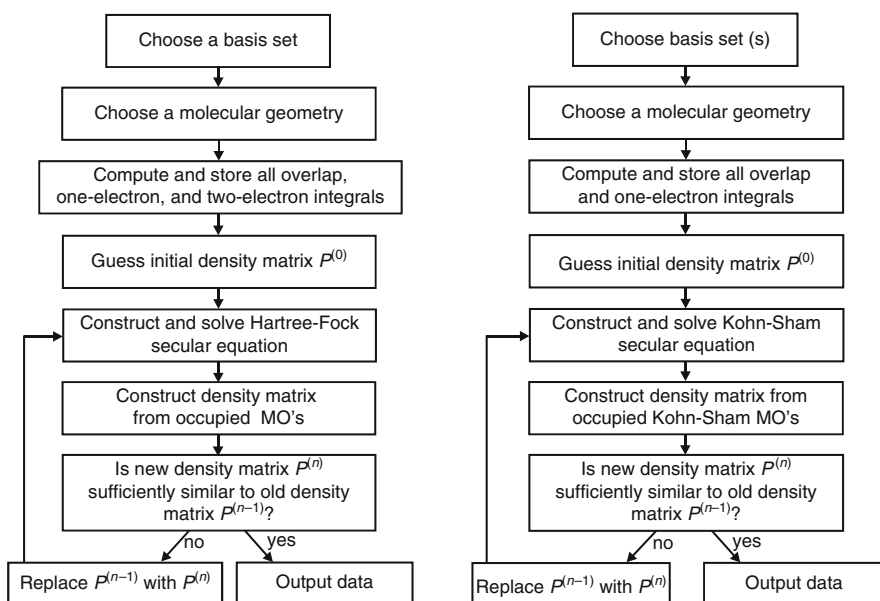


Fig. 3 Hartree–Fock and Kohn–Sham SCF scheme

state of the system. The HF method predicts molecular geometries (distances, angles) within a few percent of experiment. Even vibrational frequencies, derived from the curvature of the total energy as a function of nuclear separations, were found to be within about 10% of experiment. The binding energies of molecules are less satisfactory. Some systems like F_2 caused serious problems. The F_2 molecule was predicted to be less stable than two isolated F atoms. The calculation of the geometry and vibrational properties of O_3 also turned out to be quite difficult.

The HF model has also been used for solid state applications. To obtain more exact results, one has to calculate the correlation energy which is defined as the difference between the HF energy and the exact energy:

$$E_{corr} = E_{exact} - E_{HF}, \quad (7)$$

where E_{exact} is the energy of the system obtained from solving exactly the non-relativistic Schrödinger equation. In the HF approach the instantaneous position of an electron is not influenced by the presence of other electrons. In fact, the motions of electrons are correlated and they tend to “avoid” each other more than HF predicts, giving rise to a lower energy. There are a number of techniques to improve the HF approach, like the configuration interaction (CI) method, Møller–Plesset perturbation theory (MP) or CC approach, and many others [5]. The CI approach describes the total wave function as a linear combination of the ground- and excited-state wave functions. A CI calculation is variational and, therefore, gives an upper bound of the true energy. Of course, a full CI calculation is expensive, such that only a few excitations are taken into account. But those calculations are not size consistent. That means that the energy of a number N of non-interacting atoms or molecules is not equal to N times the energy of a single atom or molecule. To overcome this problem the quadratic configuration interaction method (QCISD) was introduced to try to deal with this. It can be considered a size consistent CISD theory. The procedure involves the addition of higher excitation terms which are quadratic in their expansion coefficients (see [5]).

The MP perturbation theory, basically the Rayleigh–Schrödinger perturbation theory, is size-independent. The idea is a partitioning of the Hamiltonian into a HF part and a perturbation V :

$$H = H_{HF} + \lambda V, \quad (8)$$

where λ is a parameter that can vary between zero and one.

Further evaluation reveals that at least second-order perturbations have to be included. This level of theory is referred to as MP2 and involves an evaluation of $E_0^{(2)}$. Higher order approximations are also possible. The MP perturbation theory is not variational and can sometimes give energies that are lower than the “true” energy. As any perturbation theory, MP perturbation theory depends on how close the starting wave function is to the exact wave function. When this is the case, convergence of the MP series is rapid. However, when bonds are stretched the MP

series sometimes become oscillatory. Moreover, higher orders of MP perturbation theory can even diverge [5].

Another popular post-HF technique is the CC method [4–6]. The CC method solves the size consistency problem of CI by forming a wave function where the excitation operators are exponentiated:

$$|\text{CC}\rangle = \exp(\hat{T})|\text{HF}\rangle, \quad (9)$$

where

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \quad (10)$$

and $\hat{T}_n$ is a linear combination of n -type excitations, for example,

$$|\text{CCS}\rangle = \exp(\hat{T}_1)|\text{HF}\rangle, \quad (11)$$

$$|\text{CCSD}\rangle = \exp(\hat{T}_1 + \hat{T}_2)|\text{HF}\rangle, \quad (12)$$

$$|\text{CCSDT}\rangle = \exp(\hat{T}_1 + \hat{T}_2 + \hat{T}_3)|\text{HF}\rangle \quad (13)$$

and so on. Thus, at the $|\text{CCS}\rangle$ level, all possible single excitations are included in the cluster operator, and at the $|\text{CCSD}\rangle$ the double excitations are also taken into consideration, etc. At each level of CC theory one includes through the exponential parametrisation of (9) all possible determinants that can be generated within a given orbital basis. These are all the determinants that enter the full CI wave function in the same orbital basis. Thus, the improvement in the sequence $|\text{CCSD}\rangle$, $|\text{CCSDT}\rangle$, and so on does not occur because more determinants are included but because of an improved representation of their expansion coefficients. Owing to the presence of the disconnected clusters, CC wave functions truncated at a given excitation level also contain contributions from determinants corresponding to higher-order excitations. The terms that are missing relative to full CI represent higher-order connected clusters and the associated disconnected clusters. By contrast, CI wave functions truncated at the same level contain contributions from determinants only up to this level.

The CC model is not variational. With a large enough basis set CCSD typically recovers 95% of the correlation energy for a molecule at equilibrium geometry, while the inclusion of triple excitations give rise to a further five- to tenfold reduction in error. The CCSD(T) method has become the method of choice for accurate small-molecule calculations.

In recent years there has been a growing interest in numerical techniques which can speed up quantum chemical computations. Various methods are available to approximate the four-index electron repulsion integrals as products of three-index intermediates. These methods are called density fitting (DF) or resolution of the identity (RI), and Cholesky decomposition (CD) techniques. A general comparison

of RI and CD methods has been published by Weigend et al. [42]. That article also gives the citations of the methods mentioned. It is shown that RI methods lead to insignificant errors only, which are partly comparable to or even better than that of CD treatments, but RI procedures are superior in speed. Other problems are electronic near-degeneracies, which occur for bond-breaking processes, diradicals, first-row transition metals, etc. Potential energy curves for bond-breaking reactions can be qualitatively incorrect even for CC calculations. For systems with few atoms, multi-reference configuration interaction (MRCI) is the method of choice [43]. For slightly larger molecules, complete-active-space second-order perturbation (CASPT2) is the preferred approach [44]. MRCI and CASPT2 are computationally demanding. Partially this problem can be overcome by parallel algorithms [45]. Work has also been done on linear scaling MRCI using local correlation and integral screening techniques [46]. Avoiding wave functions and solving directly for the two-electron reduced density matrix (2-RDM) using N-representability constraints on the possible 2-RDMs considered is now allowing accurate computations of molecular energies directly from linear functionals of the 2-RDM [47]. In a recent PhD thesis, Chaykin [48] has developed approaches for bounding optimal values of semi-definite programming applied to the electronic structure problem in the reduced density matrix (RDM) formulation. This approach introduced a verification algorithm adapted to utilise the specific problem structure for computing a rigorous lower bound of the ground state energy.

Fast computers led the interest of many researchers to general many-electron systems like CI expansions based on an orbital description and Slater determinants. The main advantage of these methods is the reduction of n -electron Hamiltonian matrix elements to one- and two-electron integrals, as stated in the Slater–Condon rules, but also showing a slow convergence. There are two sources of the slow convergence of the CI expansion. (1) The “combinatorial problem”. For an n -electron system and a basis of m spin-free one-electron functions the number of Slater determinants in a “full CI” goes as $\binom{2m}{n}$, which is a very large number unless n and m are quite small. (2) The slow convergence of “the partial wave expansion” even for a two-electron atom. If one truncates the one-electron basis at the same angular quantum number l , then the error is of the order l^{-3} . Dissatisfaction with the level of accuracy obtainable with orbital-based approaches has led to renewed interest in methods that include an explicit r_{12} -dependence into the wave function. Explicitly correlated methods have to overcome the integration over n coupled electronic degrees of freedom, required for the evaluation of Hamiltonian expectation values. Even if the n -particle basis functions are restricted to correlate explicitly only one electron pair at the time, numerous three- and four-electron integrals must be calculated, which is a formidable task. In 1985 Kutzelnigg [49] published a new explicitly correlated approach. He proposed to use an orbital basis set and to augment it with geminals that depend on a single, universal, linear r_{12} -term. This term was to describe the electronic cusp that is poorly represented in conventional orbital expansions. Kutzelnigg avoided direct evaluation of the

numerous three- and four-electron integrals, and got very good results nevertheless. The first MP2-R12 calculations on molecules were reported by Klopper and Kutzelnigg [50]. Developments of explicitly correlated wave functions can be found in a book by Rychlewski [51]. There are many extensions of the R_{12} method. Examples are calculations of intermolecular interaction energies [52] and an MP2-R12 method in which each electron pair uses a set of contracted Gaussian-type Geminals (GTG) with fixed exponents, whose coefficients are optimised linearly [53].

At present no single method can give quantitative results (accuracy of 1 kJ/mol or less) for practical problems. Therefore, so-called composite methods have been developed in which the contributions are computed at different levels of theory in accordance with the accuracy desired. For example, the Gaussian-4 theory (G4) has been developed for the calculation of energies of molecules containing first- and third-row main-group atoms [54]. Starting points are Becke–Lee–Yang–Parr B3LYP/6-31G(2df,p) geometries and harmonic vibrational frequencies, scaled by a factor of 0.9854, for the zero-point vibrational energies (ZPVE). The HF energies have been extrapolated to the limit of a complete basis, and Møller–Plesset-4/6-31G(d) single-point calculations are executed. When the MP4/6-31(d) energy E_{MP4} is considered as the base line of G4 theory, then the following energy corrections are added in the G4 composite scheme:

$$E_{G4} = E_{MP4} + \Delta E_{HF} + \Delta E_{CCSD(T)} + \Delta E_{diffuse} + \Delta E_{polarisation} + \Delta E_{G3LargeXP} + \Delta E_{HLC} + \Delta E_{SO} + \Delta E_{ZPVE}, \quad (14)$$

where

$\Delta E_{HF} \equiv$ correction for the HF basis set limit

$\Delta E_{CCSD(T)} \equiv E(CCSD(T))/6-31G(d) - E(MP4/6-31G(d))$

$\Delta E_{diffuse} \equiv$ corrections for diffuse functions

$\Delta E_{polarization} \equiv$ corrections for polarisation functions

$\Delta E_{G3LargeXP} \equiv$ correction for larger basis set effects

$\Delta E_{HLC} \equiv$ empirical energy correction that depends on the number of occupied α and β orbitals

$\Delta E_{SO} \equiv$ spin-orbit term, taken from experiment or accurate calculations

$\Delta E_{ZPVE} \equiv$ zero-point vibrational energy

The Gaussian-4 theory was tested on the G3/05 test set [55] including 454 energies. The overall average absolute deviation for these energies was found to be about 3.5 kJ/mol.

There are several other composite approaches, for example, complete basis set (CBS) models [56], focal-point analysis [57], multi-coefficient correlation methods [58], “high-accuracy extrapolated ab initio thermo-chemistry” (HEAT) [59] and the Weizmann-4 theory [60]. Hansen et al. [61] have employed the so-called MP2:DFT [193] scheme for analysing benzene ethylation over H-ZSM-5. Density functional calculations applying periodic boundary conditions [Perdew–Burke–Ernzerhof (PBE) functional] were combined with MP2 energy calculations on a series of

cluster models of increasing size which allows extrapolation to the periodic MP2 limit. Basis set truncation errors (BSSE) are estimated by extrapolation of the MP2 energy to the CBS limit. Contributions from higher-order correlation effects are accounted for by CCSD(T) coupled cluster calculations. The sum of all contributions provides the “final estimates” for adsorption energies and energy barriers. Dispersion contributes significantly to the PES. As a result, the MP2:DFT potential energy profile is shifted downward compared to the PBE profile. More importantly this shift is not the same for reactants and transition structures due to different self-interaction errors.

Other recent developments in electronic structure theory were reviewed by Sherrill [62] and Huang et al. [195].

Only a very brief description of some quantum chemical correlation approaches could be given. For details see [5, 6].

Over the last 10 years the DFT became more and more popular because of its high computational efficiency and good accuracy [2, 3]. The basis for DFT is the proof of Hohenberg and Kohn [63] that the ground state electronic energy is determined completely by the electron density ρ . In other words, there exists a one-to-one correspondence between the electron density of a system and the energy. Within DFT all aspects of the electronic structure of the system of interacting electrons in an “external” potential $V_{\text{ext}}(\mathbf{r})$ generated by atom cores are completely determined by the electronic charge density ρ . In DFT, the total energy is decomposed into three contributions, a kinetic energy, a Coulomb energy due to classical electrostatic interactions among all charged particles in the system and an exchange-correlation energy term that captures all many-body interactions. Unfortunately the exact expressions that should be used for the many-body exchange and correlation interactions are unknown. The local density approximation (LDA) turned out to be computationally convenient and very accurate. The LDA assumes that the density locally can be treated as a uniform electron gas, or equivalently that the density is a slowly varying function. An improvement is the local spin density approximation (LSDA) which is useful in cases where the α - and β -spin densities are not equal. Kohn and Sham [1] developed a self-consistent system including exchange-correlation effects. For some problems further improvements have to be made. The most common approach is the generalised gradient approximation (GGA). These approximations depend upon the gradient of the electron density at each point in space and not just on its value. These gradient corrections are typically divided into separate exchange and correlation contributions. A variety of gradient corrections have been proposed in the literature. A very popular one is the B3LYP functional [64]. In most DFT programs for calculating the electronic structure of molecules the Kohn–Sham orbitals (KS) are expressed as a linear combination of atomic-centred basis functions:

$$\psi_i(r) = \sum_{v=1}^K c_{vi} \phi_v. \quad (15)$$

This expansion is substituted into the KS equations which lead to matrix form like the HF Roothaan equations

$$\mathbf{HC} = \mathbf{SCE}. \quad (16)$$

The Kohn–Sham iteration procedure is presented in Fig. 3b.

Obviously the procedure is very similar to the HF scheme. Details of the DFT approach are given in [2, 3].

A variety of DFT-based codes is commercially available, for example, VASP [12, 65], Cambridge Serial Total Energy Package (CASTEP) [66], TURBOMOLE [67], CRYSTAL [68], Jaguar [69], GAUSSIAN [70], General Atomic and Molecular Electronic Structure System (GAMESS) [71] and QChem [13], amongst others. A rather complete list of DFT codes is available from an internet address¹ which also contains programs available for free. Some of these programs also comprise many other QM approaches.

There is a problem with DFT: the conventional functionals cannot treat dispersive forces. GGA, meta-GGA and hybrid functionals are unreliable for systems where vdW interactions are important. The most rigorous description of dispersion interactions is provided by explicitly non-local correlation functionals. However, these methods are computationally demanding and far more complicated than standard DFT. Although vdW interactions are often considered to be weak, they dominate the behaviour of all neutral physical systems at separations of order 0.5 nm or larger. vdW interactions are crucial for the chemistry and physics of weakly bound systems, for example bio-molecules, layered materials, organic crystals or adsorbing of neutral molecules on surfaces. To overcome this deficiency of DFT, two strategies have been adopted: semi-empirical approaches have been developed where an approximately derived R^{-6} term, multiplied by a suitable short-range damping function, is explicitly introduced. The R^{-6} term describes the correlated instantaneous dipole fluctuations together with higher order terms. The second strategy is the introduction of new density functionals and/or complex schemes that allow for a first-principles treatment of the vdW interactions. For example, a seamless vdW density functional (vdW-DF), valid for all interatomic distances, has been developed by Langreth and co-workers [72, 73]. An example of this approach is given by Rudenko et al. [74], which describes adsorption of halogen molecules on graphene. Implementation of vdW-DF is non-trivial. An implementation of vdW-DF with Gaussian basis functions has been presented by Vydrov et al. [75]. The semi-empirical so-called DFT plus dispersion approaches (DFT-D) employ addition of empirical, pair-wise atomic dispersion corrections of the form $-C_6R^{-6}$, which are used in the force field methods. To avoid double-counting electron correlation effects at short range, these contributions are damped for small inter-nuclear distances. This approach has been refined by Grimme et al. [76] by introducing atom-pair-wise specific dispersion coefficients and cut-off radii that are both

¹<http://dft.sandia.gov/Quest/DFT-codes.html>.

computed from first principles. The coefficients for new eighth-order dispersion terms are computed, and system (geometry) dependent information is used for a DFT-D type approach by employing the new concept of fractional coordination numbers. They are used to interpolate between dispersion coefficients of atoms in different chemical environments. The method only requires adjustment of two global parameters for each density functional. Three-body non-additivity terms are considered. Benchmark calculations showed an improvement by 15–40% compared to previous DFT-D functionals. An approach which combines the simplicity of the semi-empirical formalism with the accuracy of the first principles methods has been developed by Silvestri [77]. This approach is based on the use of the maximally localised Wannier functions.

A general scheme for systematically modelling long-range corrected (LC) hybrid density functionals has been proposed by Chai and Head-Gordon [78, 79]. Adapted to B3LYP functionals the LC hybrid functionals are quite accurate in thermo-chemistry, kinetics and non-covalent interactions, when compared to common hybrid density functionals.

There are also problems if one deals with the chemistry of an oxide surface and its role in the reactivity with adsorbed species [80].

Here, only a few papers on recent developments of including vdW forces into the DFT scheme could be mentioned.

As soon as one knows the molecular wave function, one can calculate any property of the molecule like dipole moment, chemical shift, force constants, etc. Of particular importance is the PES, which will be used in the statistical mechanics procedures discussed below.

2.2 *Monte Carlo, Molecular Dynamics*

Quantum chemical approaches give results for the absolute zero temperature ($T = 0$). In fact, systems in chemical engineering are ensembles at higher temperatures. In order to compute properties of systems of many particles at any temperature and pressure, one has to refer to statistical mechanics [15–18]. As most systems occurring in practice are very complicated, computational methods have to be employed. The most prominent ones are MC and MD approaches.

The first MC scheme was developed by Metropolis et al. [81]. The MC methods follow a Markov process to evolve a system towards equilibrium, regardless of pathway. In principle one starts with an initial configuration of molecules. The total potential energy of the initial configuration is calculated. Then several thousand random moves of the particles are executed. After each move it is checked whether the energy goes down. If it does, the move is accepted. If not, the energy is compared with a random number. If the energy is lower or higher than the random number, the movement is accepted or rejected. MC methods do generate states that are correlated owing to the sequential Markov process. In detail the so-called Metropolis algorithm works as in Fig. 4a.

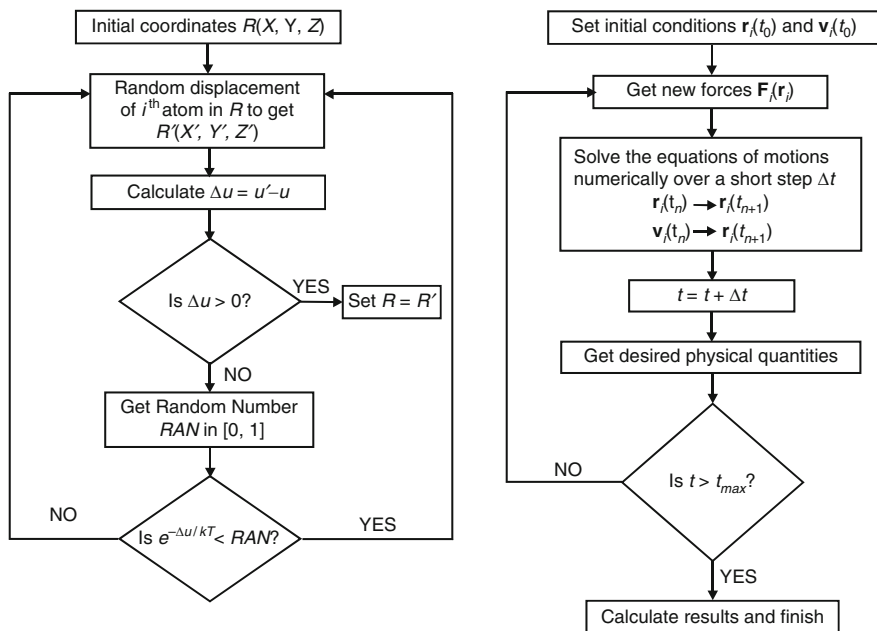


Fig. 4 Simplified Metropolis Monte Carlo and molecular dynamics scheme

After equilibrium is achieved, equilibrium properties can be calculated by an appropriate sampling of the ensemble. As only about a few hundred or thousand particles are used in a simulation box, many particles are close to the wall. This creates special wall effects which have to be avoided. This can be overcome by implementing periodic boundary conditions.

MC techniques can also be used for solving the Schrödinger equation (quantum MC) [82, 83]. QMC differs from other post-HF approaches in that it is a fully correlated method from the outset, rather than building on a mean-field approach. All QMC methods are presently computationally very demanding. Details of the MC approach can be found in the books [23–25].

The accuracy of MC and MD calculations depends primarily on proper force fields. General force fields for molecular systems are not of sufficient quality for universal use. As they were adapted to specific interactions, the terms are not transferable to any other application. Nevertheless, one can formulate some requirements force fields should satisfy. They should be robust, that means they should be valid in varying environments. For example, phase behaviour can only be described correctly if the same model represents all phases of the same quality. Other requirements are accuracy, simplicity and transferability. The accuracy of a force field should be such that experimental data can be reproduced with the required precision. Transferability means that the underlying principles of constructing the model and the parameters obtained should be applicable for various molecular problems. It is often the non-additivity of constituent terms, and the omission of

important contributions, that makes the terms non-transferable. Since most force fields contain parameters adjusted to measurements, one obtains effective interaction parameters, that means a deviation or omission in one term is compensated by other terms, which are then not precise enough in different molecular environments or various physical conditions. In principle, quantum chemical *ab initio* calculations can provide proper PES for atoms and molecules, but the calculations are extremely time consuming. The simplicity requirement means that force fields should include as few terms as possible. The following elements can be found in most simple force fields. The interactions between atoms are divided into bonded and non-bonded interactions. Bonded interactions are the covalent bonds between two atoms in a molecule, mostly described by harmonic potential. Other bonded interactions are bond angles, dihedral angles or improper dihedral angles (to keep planar molecules planar). Non-bonded interactions are in general described as pair-additive, and a function of the distance between the two particles of each pair. If all pair-wise interactions are included, the algorithm has an N^2 complexity which has to be simplified for large systems. Dispersion interactions and short-range repulsions are often described by a Lennard-Jones potential:

$$U_{LJ} = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (17)$$

The empirical r^{-12} repulsion term may be replaced by more realistic expressions in an exponential form. Coulomb interactions can be added by a Coulomb term

$$U_{\text{coul}} = f \cdot \frac{q_i q_j}{r}. \quad (18)$$

Special care is needed for the treatment of the long-range Coulomb interactions. The atomic charges in (18) should not be taken from a simple Mulliken analysis but from a fit of the charges to multi-poles [84–86] derived from accurate quantum calculations. Reducing the N^2 complexity by introducing a proper cut-off radius creates only a small error for the r^{-6} dispersion potential, but for the r^{-1} Coulomb potential simple cut-offs result in gross errors. When an abrupt potential cut-off is used, the force is no longer a derivative of the potential, and therefore, the potential is no longer conservative. The derivative of a truncated potential contains a delta distribution at the cut-off radius. This creates artefacts which cannot even be removed by shifted potentials. When particles diffuse through the limit of the interaction range they encounter a sudden force change leading to extra noise, to heating artefacts and to artefacts in the density distributions. Also so-called switching functions are of limited use, as inadequacies of the effective potentials, which deviate from the true Coulomb form, might occur. Fluids containing ions show an accumulation of like ions and a depletion of oppositely charged ions near the cut-off radius. Like ions repel each other until they reach the cut-off distance, after which they will try to diffuse back in. The effect of dielectric response of the medium beyond the cut-off radius can be incorporated by the introduction of a reaction field [87, 88].

The reaction field addition does not account for the polarisation effects in the medium beyond the cut-off due to charges rather than dipoles, and they are not satisfactory in inhomogeneous systems and systems with a long-range correlation. Ways out are, for example, the smooth-particle mesh-Ewald (SPME) approach of Essmann et al. [89] and the continuum correction methods by Wood [90]. The dominant and most relevant omission in the usual force fields is the incorporation of electronic polarizability, which is a non-additive electrical interaction. The electron distribution around a given nuclear configuration depends on the presence of external electric field. A comprehensive review of approaches including polarizabilities is given for the simulation of water by Guillot [91]. A strategy of a systematic development of force fields based on quantum calculations is described by Saint-Martin et al. [92, 93]. After obtaining the average values of certain molecular properties from short simulations with a fully flexible model, those average values can be kept fixed and the simulations continued. This is a self-consistent alternative to construct potentials based on ab initio calculations and single molecule properties, without reparametrising for each different set of thermophysical conditions, and with the advantage that even the simpler models will reflect the improvement in quality of the ab initio data used as the learning set. This is particularly important for a variety of molecular systems for which the experimental data of condensed phases are rather scarce. Different types of force fields were developed over the last few years, among them being MM3 [94], MM4 [95], Dreiding [96], SHAPES [97], VALBON [98], UFF [99], CFF95 [100], AMBER [101], CHARMM [102], OPLS [103], MMFF [104], GROMOS [105] and MARTINI [106]. Many other force fields and their respective citations are given by Jalaie and Lipkowitz [107]. Coarse-grained force fields like, e.g. MARTINI are typically parametrised based on comparison to detailed atomistic simulations, using inverted MC schemes [108, 109] or force matching approaches [110].

Fitting complex PES is a highly non-trivial task. Many optimisation algorithms for this purpose have been described by Schlick [111] and Leach [112]. The functional form has to be sufficiently flexible to adapt to the reference points with high accuracy. The obtained PES should have continuous derivatives for applications in MD simulations. The choice of the functional form requires great care, because otherwise unphysical artefacts may be introduced. In connection with the description of molecule-surface interactions, the modified Shepard method [113, 114] turned out to be useful. It is based on a Taylor expansion of the energy around the reference points. In recent years artificial neural networks (NN) [115] have become a promising new tool for the representation of PES. Due to their flexibility they are able to reproduce accurately a given set of electronic structure data, while the resulting continuous NN-PES can be evaluated several orders of magnitude faster than the underlying electronic structure calculations. Examples of NN-PES are given by Lorenz and Scheffler [116] for the dissociation of H_2 on Pd(100) surfaces, and by Behler et al. [117] for the dissociation of oxygen molecules on Al(111).

Enabling dynamical simulations on large reacting systems ($>>1,000$ atoms), so-called reactive force fields have been introduced. An example of such force

fields is the ReaxFF [118]. The harmonic approximation employed in non-reactive force fields to describe bond stretching needs to be replaced by a description that converges to the bond dissociation energy at infinite atom separation, rather than to infinite energy as in the harmonic description. Furthermore, this bond distance/energy relation has to be continuous and should, preferably, be based on physical theory. The ReaxFF uses the bond order/bond energy concept of Pauling [119] which effectively describes the number of shared electrons in the bond between two atoms, and is linked by means of a continuous function to the bond length. At short distances, this bond order/bond distance relation approaches a maximum while at infinite distance the bond order goes to zero. ReaxFF is able to describe atom hybridisation and the related multi-body interactions. Non-bonded vdW and Coulomb interactions are also included in this model. The ReaxFF has been applied to the study of the activity of Cu, Co and Ni metal atoms for initiating nanotubes growth [120], conversion of methanol to formaldehyde on $V_2O_5(001)$ [121], dissociation of H_2 , CH_4 , C_2H_5 and C_3H_8 on Ni(111) [122], amongst others.

In the acceptance steps of the above-mentioned MC algorithm (see Fig. 4a), trial steps are sometimes discarded, in particular if the temperature is low compared to typical interaction energies. Bortz et al. [123] suggested an N-fold way algorithm that avoids discarded attempts. This algorithm has been the first step towards kinetic Monte Carlo (kMC) simulations. In kMC the possible configurations of the system, i.e. the micro-states contributing to the macro-state of a statistical ensemble, need to be numerable in order to set up a list of all possible process types. In an MC simulation, on the other hand, there is no limit on the number of micro-states. For this reason, the MC algorithm can be applied to problems with a huge configuration space. In advantage over MC, a kMC simulation allows one to assign a physical time to the simulation steps [124]. In order to make use of this advantage, one has to provide as input the rates of all relevant individual processes. This can be a difficult task. The best way for getting values for the individual rates is by performing MD simulations, possibly with first-principles electronic structure methods such as DFT. In order to be able to interpret the results of kMC simulations in terms of time, one has to refer to some assumptions about the separation of time scales. The shortest time scale in the problem is given by the time it takes for an elementary process. This time scale should be clearly separated from the time interval between two events taking place at the same process (e.g. adsorption-desorption of a molecule at the same site). This second time is called the waiting time between two subsequent events. The waiting time follows a Poissonian distribution. It is assumed that any particular transition which is possible at t can again be possible at some later time $t + \Delta t$ with a uniform probability based on its rate and is independent of previous events. This is a Poisson process [125]. An important attribute of Poisson process is that the probability density of time t between successive events is

$$p(t) = Re^{-Rt}, \quad (19)$$

where R is the total rate of events.

From this probability density, the mean time between successive events is $\langle t \rangle = R^{-1}$. The time to an event is

$$T(\tau) = \int_0^\tau dt' R e^{-Rt'} = 1 - e^{-R\tau} \quad (20)$$

which lies between $[1, 0]$. So a random variable $U = e^{-R\tau}$ is uniformly distributed between $[0, 1]$, and this relation allows one to obtain the real time τ (in units of MC steps) between successive events as

$$\tau = -\frac{\ln U}{R}. \quad (21)$$

This random sampling of the Poisson time distribution for each chosen event ensures that a direct and unambiguous relationship between a real time step and an MC step is established. There are various algorithms for kMC simulations possible. In connection with catalytic reactions on surfaces, Lukkien et al. [126] have developed an efficient time–order list algorithm. Kinetic MC simulations allow one to bridge time scales over several orders of magnitude.

MD calculates the “real” dynamics of the system, from which time averages of properties can be calculated. In contrast to MC, with MD non-equilibrium properties like, for example, transport diffusivities can also be calculated. This is important for calculating diffusion of reactants and products in porous catalyst supports. The position of the molecules as a function of time are obtained by integrating Newton’s equation of motion over several thousand or even million time steps, typically up to a few femtoseconds (10^{-15}) per step. At each step, the forces on the molecules are computed and combined with the current positions and velocities to generate new positions and velocities a short time ahead. The force acting on each molecule is assumed to be constant during the time interval. The molecules are then moved to new positions, the forces are updated, and so on. By this approach, trajectories of all molecules are generated.

A simplified MD algorithm is presented in Fig. 4b.

The MD approach therefore provides information about the time dependence of the properties of the system whereas there is no such information within the MC scheme. In an MC simulation the outcome of each trial move depends only upon its immediate predecessor, whereas in MD it is possible to predict the configuration of the system at any time in the future. MD has a kinetic energy contribution to the total energy whereas in an MC simulation the total energy is determined from the potential energy function only. In general, MD simulations are executed for a micro-canonical ($N, V, E \equiv \text{const.}$) ensemble whereas MC simulations are executed for a canonical (N, V, T) ensemble. MD and MC schemes can, however, be modified for other ensembles.

Three aspects have to be considered with MD simulations: (1) the model describing the intra- and inter-particle interactions, (2) the calculation of energies and forces from the model and (3) the algorithm employed to integrate the

equations of motion. The interaction potentials have been explicated in connection with the MC approach. The most time-consuming step is the evaluation of energies and, when needed, forces [21, 23]. The Newtonian equations of motion are solved in proper time steps Δt . The following considerations influence the choice of algorithm: (1) the time reversibility of Newton's equations of motion should be conserved and (2) the generated trajectories should conserve volume in phase space; that means the integrator should be symplectic (see p. 381 in [127]). This is important to conserve equilibrium distributions in phase space, because deviation from symplectic behaviour will produce time-dependent weight factors in phase space [128]. The computing time for MD simulations is dominated by force calculations. Therefore, approaches that use only one force evaluation per time step are preferred.

A more rigorous derivation of integration schemes, which leads to the possibility of splitting the propagator of the phase space trajectory into several time scales, is based on the phase space description of a classical system. The time evolution of a point in the 6N dimensional phase space is given by the Liouville equation

$$\mathbf{\Gamma}(t) = e^{iLt}\mathbf{\Gamma}(0), \quad (22)$$

where $\mathbf{\Gamma} = (\mathbf{q}, \mathbf{p})$ is the 6N dimensional vector of generalised coordinates, $\mathbf{q} = \mathbf{q}_1, \dots, \mathbf{q}_N$, and momenta $\mathbf{p} = \mathbf{p}_1, \dots, \mathbf{p}_N$. L is the Liouville operator, defined as

$$iL = \{..., H\} = \sum_{j=1}^N \left(\frac{\partial \mathbf{q}_j}{\partial t} \frac{\partial}{\partial \mathbf{q}_j} + \frac{\partial \mathbf{p}_j}{\partial t} \frac{\partial}{\partial \mathbf{p}_j} \right). \quad (23)$$

Equation (22) is the starting point for the derivation of numerical integration schemes. In order to construct a discrete time-step integrator, the Liouville operator is split into two parts, $L_1 + L_2$, and a Trotter expansion [129, 130] is performed:

$$e^{iL\delta t} = e^{i(L_1+L_2)\delta t} = e^{iL\delta t} = e^{iL_1\delta t/2} \cdot e^{iL_2\delta t} e^{iL_1\delta t/2}. \quad (24)$$

The exact formula is the so-called Baker–Campbell–Hausdorff formula [131].

Any partial operators can be chosen to act only on positions or momenta. Assuming Cartesian coordinates for a system of N free particles, this can be written as

$$iL_1 = \sum_{j=1}^N \mathbf{F}_j \frac{\partial}{\partial \mathbf{p}_j} ; \quad iL_2 = \sum_{j=1}^N \mathbf{v}_j \frac{\partial}{\partial \mathbf{r}_j}. \quad (25)$$

Applying (24) to the phase space vector $\mathbf{\Gamma}$ and using the property $\exp(a\partial/\partial x)f(x) = f(x + a)$ for any function f , where a is independent of x , gives

$$\mathbf{v}_i(t + \delta t/2) = \mathbf{v}(t) + (\mathbf{F}_i(t)/m)\delta t/2, \quad (26)$$

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \delta t/2)\delta t, \quad (27)$$

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t + \delta t/2) + (\mathbf{F}_i(t + \delta t)/m_i)\delta t/2. \quad (28)$$

This is the velocity Verlet algorithm. Interchanging $L_1 \rightarrow L_2$ and $L_2 \rightarrow L_1$ leads to the position Verlet algorithm. Both Verlet algorithms are examples of symplectic integrators, which are volume conserving of the phase space. Any method which is based on the splitting of the Hamiltonian is symplectic. The time reversibility is guaranteed by symmetric methods [132]. The decomposition of the Liouville operator offers the opportunity for a decomposition of time scales in an MD simulation. A somewhat more general factorization of (24) leads to the reversible reference system propagator algorithm (r-RESPA) [133]. Multiple time scale (MTS) methods partition the MD simulation into “slow” and “fast” portions, assigning appropriate time steps to each segment. The r-RESPA provides an integration scheme which is reversible in time and evolves in a symplectic way. This algorithm has been implemented, for example, in simulation packages such as NAMD2 [134], AMBER [101] and DESMOND [135]. Although symplectic integrators like r-RESPA can generate long stable trajectories, the performance of the algorithms depends on the details of the implementation. This is particularly important in the way in which non-bonded interactions are partitioned. The algorithm r-RESPA and related integrators are known to suffer from resonance instabilities [136], which bound the size of the time step of the slowest motions relative to the size of the faster modes. Resonance phenomena engender the building up of energy in the system, thereby giving rise to drifts in average properties and inaccurate sampling. Morrone et al. [137] show how this can be corrected. Skeel [138] describes various integration schemes for MD simulations and related applications.

The micro-canonical ensemble (NVE) may be considered as the natural ensemble for MD simulations, as it is the canonical ensemble (NVT) for MC simulations. By control of certain thermodynamic quantities, it is possible to realise different types of thermodynamic ensembles by MD simulations. This can be achieved by external control methods like stochastic approaches, strong or weak coupling methods and extended system dynamics. The application of stochastic disturbances to control temperature was introduced by Schneider and Stoll [139] and corresponds to a Langevin thermostat. The idea is to apply a frictional force and a random force to the momenta. Velocity rescaling, as is executed by an Andersen thermostat [140], distributes the external disturbance less smoothly. Strong coupling methods artificially constrain a property to the desired value, e.g. the total kinetic energy to a given temperature. This iso-kinetic or Gauss thermostat was introduced by Hoover et al. [141] and Evans et al. [142]. Strong coupling can also be applied to constrain the pressure to a preset value [143, 144]. It produces a canonical distribution as has been shown by Tuckerman et al. [145]. Weak coupling methods [146] are not stochastic, and can be employed both for temperature and pressure control. For temperature control they do have the same effect as a Langevin thermostat on the variance of velocities. The velocities are rescaled per

step in such a way that the total temperature of the system will decay with a first-order process to the desired temperature T .

Nosé [147, 148] introduced a so-called extended system dynamics by giving an extra degree of freedom that can be used to control a variable in the system. Hoover [149] modified the time-scaling somewhat into a scheme known as Nosé–Hoover thermostat.

Jakobtorweihen et al. [150–152] have developed a so-called “Lowe–Andersen interface-fluid collision thermostat” (LA-IFC) that mimics the thermal effects of a flexible porous framework stochastically. This thermostat avoids simulations employing a flexible carbon nanotube, for example. The results of diffusivity calculations coincide with MD simulations using fully flexible pore walls, but LA-IFC simulations are orders of magnitude faster than MD simulations.

One of the key quantities in thermodynamics is the free-energy associated with changes in molecular conformations or chemical reactions. Unfortunately, it is generally much more difficult to calculate free energy differences from simulations than to obtain energy differences. The reason for this is that the free-energy expression contains an entropic term $-TS$, whereby entropy is given by an integral over phase space, while energy is an ensemble average. Free energies can be evaluated from completely equilibrated trajectories or ensembles that contain all accessible regions of configurational space. In practice it is hard to generate such complete ensembles, in particular for “rough” energy landscapes. For this purpose various new techniques have been developed, like parallel tempering [153, 154], hyper-dynamics [155], meta-dynamics [156], blue-moon ensemble approach [157, 158], umbrella sampling [159] and adiabatic free-energy dynamics [160], amongst others. A comprehensive review of free-energy techniques was presented by Chipot and Pohorille (eds.) [161].

Since the 1990s, massively parallel computers have become available, where thousands of processors may work on a single task. An MD program consists essentially of the force routine, which usually takes up more than 90% of the execution time. The integrator may be naturally parallelised, since the loop over N particles may be subdivided and performed on different processors. The new message passing interface (MPI) II standard offers parallel read/write operations. Other libraries, like Pnet CDF and net CDF-4, offer even more efficient parallel I/O. Some examples of parallel MD simulations are given by [27]. Other details of MD simulations may be found in the books [21, 23, 26, 27].

2.3 QM/MM, Transition State Theory

The paper written by Warshel and Levitt [162] introduced the first QM/MM concept. Since about 1990, QM/MM approaches have been used by many groups, in particular for modelling enzyme reactions, but also for other catalytic reactions, for example, in zeolites [163]. QM/MM methods are established as a powerful computational technique to treat reactive processes in large systems. They are

useful whenever one has to model a localised chemical reaction at an active site that is influenced by an interacting larger environment. Recent advances in QM/MM have been reviewed in various papers, for example [164–167]. In QM/MM simulations one wants to retain as much as possible the formalism of the methods that are being combined and to introduce well-defined coupling terms. The entire system is divided into the inner QM region that is treated quantum-mechanically and the outer MM region that is described by a force field. There is not one single QM/MM method, and the multitude of different implementations can be characterised by several main distinctions. For example, the so-called subtractive methods apply the QM approach to the active site and the MM method to the entire system, including the active site. The MM contribution for the active site has to be subtracted:

$$E = E_{QM}^{acts} + E_{MM}^{tot} - E_{MM}^{acts}. \quad (29)$$

The advantage of this approach is that it allows one, in a simple way, to combine different QM schemes and MM schemes. The disadvantage is that the active site has to be calculated by MM which might be difficult for complex electronic structures. The additive scheme applies the MM only to the environment of the active site, and a coupling term has to be introduced for the two regions:

$$E = E_{QM}^{acts} + E_{MM}^{envir} + E_{QM/MM}^{coupl}. \quad (30)$$

The coupling terms normally include bonded terms across the QM/MM boundary, non-bonded vdW terms and electrostatic terms. A further problem is the treatment of the QM/MM boundary. The choice of the QM region is usually made by chemical intuition. This region can be enlarged step-wise, and its sensitivity to the QM/MM results can be checked. Standard QM/MM applications employ a fixed QM/MM partitioning where the boundary between the QM and MM regions is defined once and for all at the outset. Park and Heyden [168] have derived a mixed-resolution Hamiltonian and an explicit symplectic integrator for conservative mixed-resolution systems that allow for a dynamic change in resolution of selected groups of atoms during MD simulation. The so-called adaptive partitioning of the Lagrangian (APL) method permits a simulation with accuracy comparable to an atomistic one at the computational cost of a coarse-grained one.

DFT is the workhorse for the QM part. For extensive QM/MM MD simulations one has to refer to semi-empirical methods. Linear scaling local correlation methods have also been used [169]. For the MM part proper force fields have to be employed, e.g. GROMOS [105], AMBER [101], OPLS [103]. The electrostatic coupling between the QM charge density and the MM charge model can be done in various ways [170]. Mostly, electrostatic embedding is employed which allows for the polarisation of the QM region since the QM calculation is performed in the presence of the MM charge model, whereby the MM point charges are included as one-electron terms in the QM Hamiltonian. The treatment of the QM/MM boundary can be executed in various ways. Most schemes give nearly the same results as long

as the charges at the QM/MM boundary are carefully treated [171]. Introduction of dangling hydrogen bonds or treating the frontier functional group as a pseudo-atom with an effective one-electron potential are the most common approaches.

Nowadays many QM and MM software packages offer QM/MM capabilities. ChemShell (www.chemshell.org) software is an example of a modular QM/MM implementation.

Finding the transition states in high-dimensional spaces is a challenging problem. Transition states are first-order saddle points. The algorithms for finding first-order saddle points on one spin PES can be divided into two groups: (1) approaches based on interpolation between a reactant and a product minimum and (2) those using only local information. A combination of both algorithms is probably the most efficient way of finding first-order saddle points. Interpolation methods generate a sequence of approximate MEP by interpolating between a reactant and a product state. The highest energy configuration along an MEP is a first-order saddle point. Both reactant and product states must be known so that these methods cannot reveal unexpected chemical pathways with multiple intermediates. Furthermore, if multiple pathways exist, only that nearest to the interpolated guess will be found [10]. The interpolation algorithms convert a saddle point search in configuration space to a minimisation problem in discretised path space. Minimisation problems in path space can easily handle large numbers of low-frequency modes, a significant challenge for most local surface walking algorithms. Interpolation algorithms include, for example, nudged elastic band (NEB) [172] and the string method [173]. These methods initiate the search for a transition state by assuming that the MEP is a straight line in multidimensional space connecting the reactant and product states. Peters et al. [10] have shown that the growing string method, an interpolation method that does not require an initial guess for the initial pathway, needs significantly fewer gradient calculations to find the saddle point than the NEB and the string method.

Local surface-walking algorithms explore the PES using local gradient and usually second derivative information. These methods can be initiated anywhere on the PES. These algorithms perform poorly for systems with several low-frequency vibrational modes or for searches started far from a transition state. Furthermore, even if a transition state is found it is possible that it does not connect reactant and product states. Therefore, it is recommendable to employ an interpolation algorithm like the growing string method to generate a starting point for the local surface walking algorithm. Two of the most used algorithms of this type are the P-RFO method by Baker [174] and the dimer method by Henkelman and Jónsson [175] or its improved version by Heyden et al. [11]. The latter method is available in some commercial program packages like VASP [65] or QChem [13].

The reaction rate constants are mostly calculated based on the harmonic TST. Comprehensive review of this subject was presented by Hänggi et al. [176]. The rate coefficients for elementary reactions on a catalyst surface are obtained by conventional TST in the following way:

$$k_{TST}(T) = \frac{k_B T Q_{TS}(T)}{h Q_R(T)} \exp\left(-\frac{E^+}{RT}\right), \quad (31)$$

where k_B is Boltzmann's constant, h Planck's constant, T is the absolute temperature and E^+ is the difference in electronic energies between the transition state and the reactant state, respectively. The partition functions of the transition state, $Q_{TS}(T)$, and the reactant state may be calculated, for example, like this:

$$Q = \left(\frac{2\pi M k_B T}{h^2}\right) \frac{\sqrt{\pi}}{\sigma} \left(\frac{T^3}{\theta_A \theta_B \theta_C}\right)^{0.5} \left(\prod_j^{3n-6} \frac{1}{1 - e^{-\theta_{vj}/T}}\right) \omega_{el}, \quad (32)$$

where M is the molecular mass, θ_i are the moments of inertia, the θ_{vj} the normal modes and ω_{el} the electronic energy. A transmission coefficient can also be introduced which has the general form like this [19]:

$$\gamma(T) = \Gamma(T) \kappa(T) g(T), \quad (33)$$

where $\Gamma(T)$ arises from dynamical recrossing. It takes into account that some trajectories that cross the dividing surface in the direction of products recross and return to the reactant region. $\Gamma(T)$ is smaller than one. $\kappa(T)$ arises from quantum mechanical tunnelling. $\kappa(T)$ is greater or equal to one. $g(T)$ takes deviations of the equilibrium distribution in phase space into account. $g(T)$ can be either less than or greater than one. In conventional TST $\gamma(T)$ is set equal to one. Further developments of TST may be found in papers by Truhlar's group [7–9].

There are important examples where the harmonic/rigid-rotor approximation to TST fails in describing the reaction kinetics. Even worse, simulations based on the static approach can sometimes lead to completely incorrect prediction of the reaction mechanism. For example, in catalytic transformations of short alkanes, entropy plays an important role. During the reaction the mobility of the reactants varies according to the strength of their interactions with the zeolite, leading to a substantial entropy contribution to the free-energy reaction barrier. Entropy can even stabilise some otherwise unstable reaction intermediates, opening unexpected alternative reaction channels competing with the mechanism deduced from a static TST search. Therefore, one has to explore the free-energy surface and not just of the PES in configuration space. Bucko and Hafner [177] have shown that the static approach, corrected for dynamical effects within harmonic TST, is insufficient for describing reactions including weakly bound adsorption complexes such as hydrocarbon conversion reactions. The most important reasons for this failure were found to be as follows. (1) An adsorption complex identified by static total-energy minimisation is not a proper representation of the reactant state. Hence the work needed to create an adsorption complex represents in some cases an important contribution to the free-energy barrier. This contribution is not taken into account in harmonic TST, which is based on the analysis of the energy surface in the vicinity of stationary points only. (2) The static approach does not account for reaction

intermediates which are not potential-energy minima, hence it does not allow for changes in the reaction mechanism induced by thermal fluctuations.

Bucko and Hafner [177] employed transition path sampling techniques [178, 179] to overcome the static TST problems (see also [196]).

Other general aspects for computing catalytic reactions are discussed by Raimondeau and Vlachos [38, 184], Berendsen [180] and Broadbelt and Snurr [181].

3 Applications: From the Active Centre to the Chemical Reactor

An example of a heterogeneous catalytic reaction will be presented, which, to the best of my knowledge, for the first time describes the complete picture consistently from the active centre to the chemical reactor [182, 183]. Previous simulations of heterogeneous catalytic reactions were summarised by Raimondeau and Vlachos [184], Broadbelt and Snurr [181], Vlachos [185] and Santiso and Gubbins [36]. Furthermore, results on simulations of heterogeneous catalytic reactions may be found in books by van Santen and Sautet (eds.) [40] and van Santen and Neurock [41]. Christensen and Norskov [186] describe some computational investigations on trends of reactivity of catalytic surfaces.

Molecular simulation of heterogeneous catalytic reactors, which is a multiscale problem (Fig. 5) initiated by quantum chemical calculations, may be used in combination with TST to obtain intrinsic kinetic data and to elucidate the reaction mechanism.

In [182, 183] the alkylation of benzene with ethene over H-ZSM-5 was investigated. The electronic energies and vibrational frequencies of each stationary point along the reaction coordinate were calculated for the two mechanisms presented in Fig. 6.

Adsorption from the gas phase into the zeolite leads to a decrease of the potential of each molecule (Fig. 7).

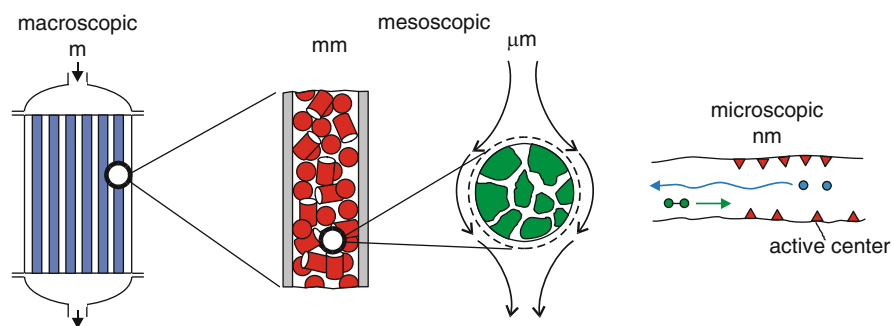


Fig. 5 Orders of magnitude in heterogeneous catalytic reactors

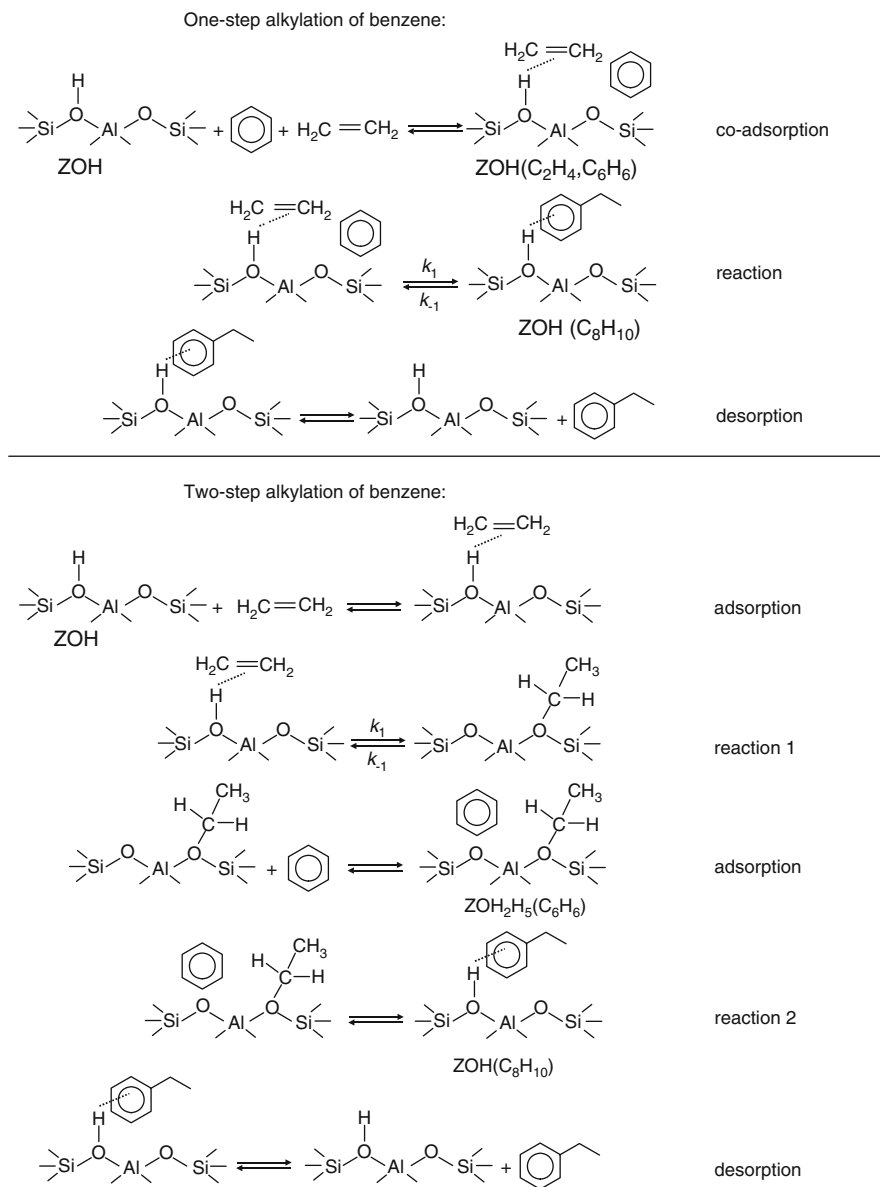


Fig. 6 Elementary steps involved in the alkylation of benzene

This is caused by dispersion interaction between the carbon atoms of the reactant and the oxygen atoms of the zeolite framework. Interaction of the reactants with the Brønsted acid site further reduces the potential. The intrinsic energy barrier for the forward reaction is the difference between the bottom of the well for the co-adsorption of A and B at the active site and the top of the transition state. First, the

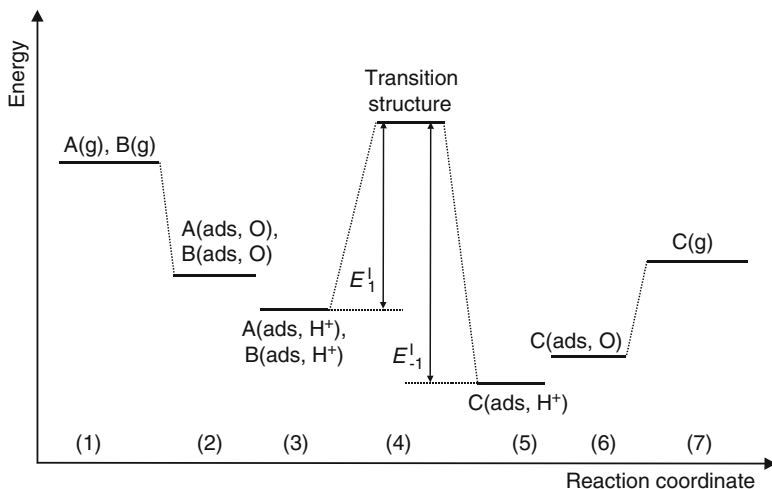


Fig. 7 Schematic energy diagram for a zeolite-catalysed one-step reaction $A + B \rightarrow C$. (1) Reactants in the gas phase. (2) Adsorption of A and B in the zeolite channels. (3) Co-adsorption of A and B on the Brønsted acid site. (4) Formation of the transition state. (5) Product adsorption on the Brønsted site. (6) Product adsorption in the channel system. (7) Product in the gas phase

reaction mechanism for the alkylation was elucidated using DFT applied to cluster representations of the active site. Second, the MP2:DFT hybrid approach suggested by Tuma and Sauer [193] was employed. Structure optimisation of all stationary points within the full ZSM-5 unit cell using DFT with periodic boundary conditions has been the first step in this approach. Stationary points were characterised by harmonic frequencies obtained by diagonalisation of the full dynamical matrices (absence of imaginary frequencies for minima and presence of exactly one imaginary frequency for all transition structures). To confirm that the transition states were connected to the correct energy minima, each transition state was perturbed slightly along the reaction coordinate in the reactant and product direction. The perturbed geometries were used as starting geometries for energy minimisation. The second step was the calculation of the high-level correction; that is the difference between MP2 and DFT single-point adsorption energies and energy barriers, respectively, for clusters of progressively larger size. These clusters were cut out from the periodic DFT-optimised structure. The size-dependent high-level correction [187]

$$\Delta E(C)_{high} = \Delta E(C)_{MP2} - \Delta E(C)_{DFT} \quad (34)$$

was then extrapolated to the periodic structure (S). This periodic model limit, $\Delta \tilde{E}(S)_{high}$, is added to the plane-wave DFT energy for the periodic structure, $\Delta E(S)_{DFT}$, to get an estimate of the MP2 energy for the full periodic system,

$$\Delta\tilde{E}(\mathbf{S})_{MP2} = \Delta\tilde{E}(\mathbf{S})_{DFT} + \Delta\tilde{E}(\mathbf{S})_{high}, \quad (35)$$

where the tilde is used to discriminate energies which were obtained through fitting and/or extrapolation, respectively, from those obtained directly from quantum chemical calculations. The third step was the extrapolation of the MP2 energy to the CBS limit. Additionally, single-point coupled cluster calculations [CCSD(T)] were conducted to account for higher order correlation effects. Thermodynamic contributions arising from finite temperatures were taken into account by calculation of partition functions within the harmonic approximation assuming separability of electronic translational, rotational and vibrational terms. The first step in linking the outcome of a quantum chemical calculation (i.e. the PES) to a rate of reaction is the calculation of rate coefficients for elementary steps. For the alkylation of benzene the barrier crossing is an infrequent event because the condition $E_{barr}/RT \gg 1$ holds. This allows the use of classical TST. Therefore, the intrinsic rate constants were calculated according to (31)/(32).

The next step is the modelling of multi-component adsorption and diffusion inside the pores (Fig. 8).

This sequential strategy is possible if the length and time scales of the problem are well separated, and if suitable models for linking the levels of modelling are available. The adsorption isotherms for ethene, ethane, benzene, ethylbenzene and hydrogen in MFI (all silica form) were determined by “configurational-bias Monte Carlo” (CBMC) simulations in the grand canonical ensemble at a variety of temperatures. The outcome of these simulations depends on the quality of the force fields. A united atom force field was employed that was parametrised against adsorption isotherms in zeolites. The simulated isotherms were fitted to a three-site Langmuir expression:

$$q_i(f_i) = \frac{q_{i,sat,A} b_{i,A} f_i}{1 + b_{i,A} f_i} + \frac{q_{i,sat,B} b_{i,B} f_i}{1 + b_{i,B} f_i} + \frac{q_{i,sat,C} b_{i,C} f_i}{1 + b_{i,C} f_i}, \quad (36)$$

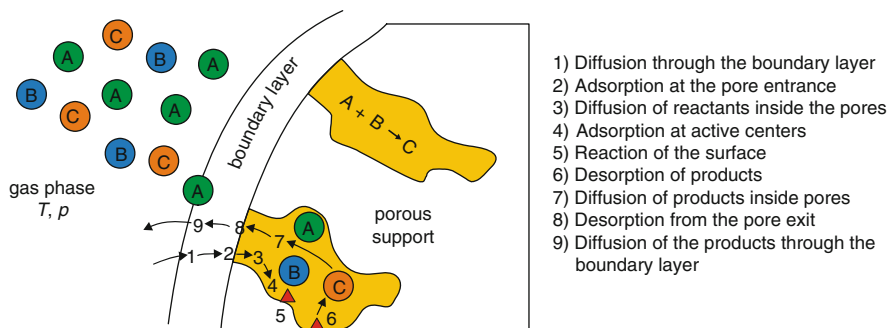


Fig. 8 Adsorption and diffusion inside the pores

where $q_{i,\text{sat},X}$ denotes the saturation capacity of species i on site X , $b_{i,X}$ is the affinity constant and f_i is the gas phase fugacity of species i . The continuum level of an entire pellet is calculated according to the partial differential equation

$$\frac{\partial q_i}{\partial t} = -\frac{1}{\rho} \frac{1}{\zeta^2} \frac{\partial}{\partial \zeta} (\zeta^2 N_i) + \frac{1}{\rho} v_i r; \quad i = 1, 2, \dots, n \quad (37)$$

where q_i is the loading of species i , ρ is the zeolite framework density, ζ the diffusion path, N_i the molar flux of species i , v_i the stoichiometric coefficient and r is the rate of reaction. As the composition of reactants and products along the pores changes continuously due to reaction, the adsorption equilibria for arbitrary compositions inside the pores need to be calculated. This was achieved by means of ideal adsorbed solution theory (IAST) [194] which requires only the pure component isotherm data as input. These isotherms are taken from CBMC simulations. The suitability of the IAST has to be checked by a limited number of MC simulations of the multi-component adsorption equilibria. Therefore, IAST is a linking model for MC adsorption results.

MD simulations were carried out in a rigid zeolite framework for a variety of loadings and temperatures, employing the same force field as was used in the CBMC simulations. From these data, self-diffusivities and Maxwell–Stefan diffusivities were extracted. These diffusivities were used in the N_i terms of (37). Again one has to check the validity of the Maxwell–Stefan approach by means of some multi-component MD simulations. As the composition of the molecular mixture inside the pores changes along the pore and with time, one has also to refer to a linking model for diffusivities, as a complete calculation by MD would be by far too time consuming. The Maxwell–Stefan approach serves as a linker. For details see [182].

The rate of reaction of the one-step mechanism (see Fig. 6) is given by

$$\frac{r}{\rho} = \tilde{r} = k_f q_{E+B,H^+} - k_r q_{EB,H^+}, \quad (38)$$

where k_f and k_r are the rate coefficients for the forward and reverse reaction of ethene and benzene to form ethylbenzene, respectively, q_{E+B,H^+} is the amount of co-adsorbed “ethene + benzene” at the active sites and q_{EB,H^+} is the amount of adsorbed ethylbenzene at the active sites. As benzene and ethylbenzene are mostly sited in the zeolite cages whilst the ethene molecules can move into the channels and the cages, the assumptions made for Langmuir–Hinshelwood kinetics are not fulfilled. For q_{E+B,H^+} and q_{EB,H^+} analytical expressions are needed in order to calculate them from species loadings q_E , q_B , and q_{EB} . As was outlined in [182], such terms may be obtained from MC simulations of the multi-component adsorption isotherms. As benzene and ethylbenzene are mostly sited inside the cages, fitting to Langmuir–Hinshelwood kinetics is not possible.

To conclude, by employing suitable linking models based solely on molecular data, one can obtain results on a macro level, i.e. composition profiles inside

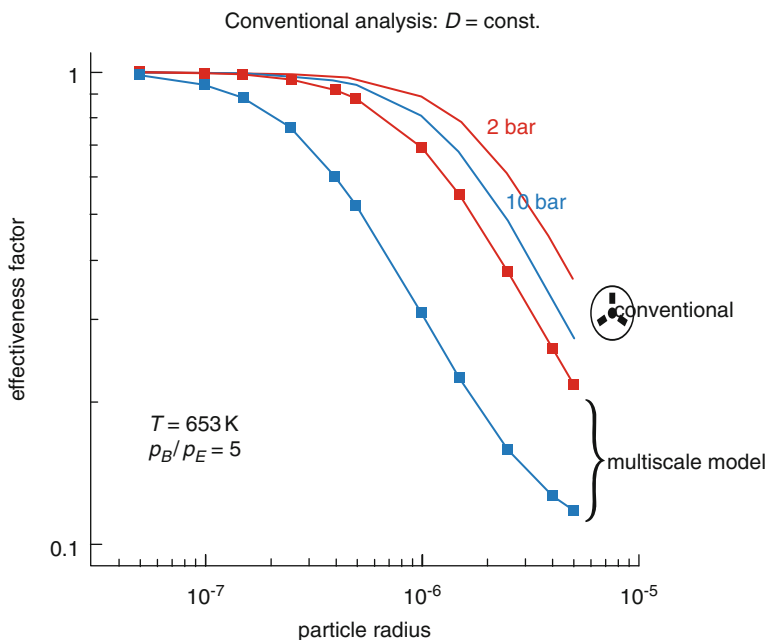


Fig. 9 Effectiveness factor

catalyst pellets and reactors. The pellet (37) is solved, for example, by a finite difference approach, and a reactor plug flow model was used for the entire reactor. The results are presented in [182, 183]. One result is discussed here. Figure 9 shows plots of the effectiveness factor vs particle radius. The bulk phase composition has a benzene/ethylene molar ratio of five. It can be seen that the conventional results obtained for a fixed value of an effective diffusivity deviates considerably from detailed simulations based on real local compositions and rates along the pores, in particular for higher pressures.

4 Outlook

Over the last few years there has been considerable progress in multiscale modelling of catalytic processes, although there is still considerable room for improvement on all levels. One of the weakest points is the harmonic TST. There is increasing evidence that this fails for reactions involving loosely bound reactant and/or transition states where entropy makes a significant contribution to the free-energy reaction barrier. The free energy of activation may be derived by free-energy integration schemes such as the Blue-Moon ensemble technique in combination with constrained ab initio MD simulations. There is a need for accurate experimental kinetic data to assess the reliability of modelling approaches for predicting effective

reaction rates for zeolite-catalysed reactions. The problem of vdW forces with DFT calculations is in part solved. In heterogeneous catalysis, only for very few reactions is the mechanism known in detail. Well understood are ammonia synthesis and the hydrogenation of ethylene. The industrial Envinox process has also been investigated in great detail [188–191]. On the force field level there is a need for force fields that can describe interactions between hydrocarbons and Brønsted acid sites. Such force fields have to be parametrised against accurate quantum mechanical calculations because experimental data are hard to obtain. Reliable approaches are required to calculate diffusion coefficients for slowly diffusing molecules such as aromatics which are difficult to obtain directly from classical MD simulations. On the continuum level, the description of multi-component diffusion can be improved by accounting for strong adsorption sites by using, for example, the effective medium approximation by Coppens and co-workers [192]. Therefore, modelling of catalysed heterogeneous reactions will be an active field of research in the future.

References

1. Kohn W, Sham LS (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev A* 140:1133–1138
2. Kohanoff J (2006) *Electronic structure calculations for solids and molecules: theory and computational methods*. Cambridge University Press, Cambridge
3. Martin RM (2008) *Electronic structure: basic theory and practice methods*. Cambridge University Press, Cambridge
4. Szabo A, Ostlund NS (1996) *Modern quantum chemistry: an introduction to advanced electronic structure theory*. Dover Publications, Mineola, NY
5. Helgaker T, Jørgensen P, Olsen J (2002) *Molecular electronic theory*. John Wiley & Sons, New York
6. Shavitt I, Bartlett RJ (2009) *Many-body methods in chemistry and physics: MBPT and coupled cluster theory*. Cambridge University Press, Cambridge
7. Truhlar DG, Steckler R, Gordon MS (1987) Potential energy surfaces for polyatomic reaction dynamics. *Chem Rev* 87:217–236
8. Pu J, Gao J, Truhlar DG (2006) Multidimensional tunnelling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem Rev* 106:3140–3169
9. Gao J, Truhlar DG (2002) Quantum mechanical methods for enzyme kinetics. *Ann Rev Phys Chem* 53:467–505
10. Peters B, Heyden A, Bell AT, Chakraborty A (2004) A growing string method for determining transition states: comparison to the nudged elastic band and string methods. *J Chem Phys* 120:7877–7886
11. Heyden A, Bell AT, Keil FJ (2005) Efficient methods for finding transition states in chemical reactions: comparison of improved dimer method and partitioned rational function optimization method. *J Chem Phys* 123:224101-1/14
12. Kresse G, Hafner J (1993) Ab initio molecular dynamics for liquid metals. *Phys Rev B* 47:558–561
13. Shao Y et al (2006) Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* 8:3172–3191

14. Banerjee A, Adams N, Simons J, Shepard R (1985) Search for stationary points on surfaces. *J Phys Chem* 89:52–57
15. McQuarrie DA (2000) Statistical mechanics. University Science Books, Sausalito
16. Hill TL (1987) An introduction to statistical mechanics. Dover Publications, Mineola, NY
17. Chandler D (1987) Introduction to modern statistical mechanics. Oxford University Press, Oxford
18. Tuckerman ME (2010) Statistical mechanics: theory and molecular simulation. Oxford University Press, Oxford
19. Truhlar DG, Garrett BC, Klippenstein SJ (1996) Current status of transition-state theory. *J Phys Chem* 100:12771–12800
20. Zener C (1932) Non-adiabatic crossing of energy levels. *Proc Roy Soc London Ser A* 132:696–702
21. Frenkel D, Smit B (2001) Understanding molecular simulation, 2nd edn. Academic, San Diego
22. Landau DP, Binder K (2009) A guide to Monte Carlo simulations in statistical physics, 3rd edn. Cambridge University Press, Cambridge
23. Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Oxford University Press, Oxford
24. Newman MEJ, Barkema GT (1999) Monte Carlo methods in statistical physics. Oxford University Press, Oxford
25. Binder K, Heermann DW (2010) Monte Carlo simulation in statistical physics: an introduction. Springer, Heidelberg
26. Rapaport DC (2004) The art of molecular dynamics simulation, 2nd edn. Cambridge University Press, Cambridge
27. Griebel M, Knappek S, Zumbusch G (2009) Numerical simulation in molecular dynamics: numerics, algorithms, parallelization, applications. Springer, Heidelberg
28. Tuckerman ME, Martyna GJ (2000) Understanding modern molecular dynamics: techniques and applications. *J Phys Chem B* 104:159–178
29. Versteeg HK, Malalasekera W (1995) An introduction to computational fluid dynamics. Prentice Hall, Upper Saddle River, NJ
30. Lomax H, Pulliam TH, Zingg DW (2001) Fundamentals of computational fluid dynamics. Springer, Heidelberg
31. Laney CB (1998) Computational gasdynamics. Cambridge University Press, Cambridge
32. Grotendorst J, Attig N, Blügel S, Marx D (2009) Multiscale simulation methods in molecular sciences. Jülich Supercomputing Centre, Jülich
33. Sherwood P, Brooks BR, Sansom MSP (2008) Multiscale methods for macromolecular simulations. *Curr Opin Struct Biol* 18:630–640
34. Koci P, Novak V, Stepanek F, Marek M, Kubicek M (2010) Multi-scale modelling of reaction and transport in porous catalysts. *Chem Eng Sci* 65:412–419
35. Lynbarsev A, Tu YQ, Laaksonen A (2009) Hierarchical multiscale modelling scheme from first principles to mesoscale. *J Comput Theor Nanosci* 6:951–959
36. Santiso EE, Gubbins KE (2004) Multi-scale molecular modeling of chemical reactivity. *Mol Simul* 30:699–748
37. Starrost F, Carter EA (2002) Modelling the full monty: baring the nature of surfaces across time and space. *Surf Sci* 500:323–346
38. Vlachos DG (2005) A review of multiscale analysis: examples from systems biology materials engineering, and other fluid-surface interacting systems. *Adv Chem Eng* 30:1–61
39. Baeurle SA (2009) Multiscale modeling of polymer materials using field-theoretic methodologies: a survey about recent developments. *J Math Chem* 46:363–426
40. van Santen RA, Neurock M (2006) Molecular heterogeneous catalysis. Wiley-VCH, Weinheim
41. van Santen RA, Sautet P (2009) Computational methods in catalysis and materials science. Wiley-VCH, Weinheim

42. Weigend F, Kattannek M, Ahlrichs R (2009) Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods. *J Chem Phys* 130:164106
43. Werner H-J, Knowles PJ (1988) An efficient internally contracted multiconfiguration - reference configuration interaction method. *J Chem Phys* 89:5803–5814
44. Andersson K, Roos BO (1995) In: Yarkony DR (ed) *Modern electronic structure theory*, Advanced Series in Physical Chemistry, Vol 2. World Scientific, Singapore
45. Schüller M, Kovar T, Lischka H, Shepard R, Harrison RJ (1993) A parallel implementation of the COLUMBUS multireference configuration interaction program. *Theor Chim Acta* 84:489–509
46. Chwee TS, Szilva AB, Lindh R, Carter EA (2008) Linear scaling multireference singles and doubles configuration interaction. *J Chem Phys* 128:224106-1/9
47. Mazzotti DA (2006) Quantum chemistry without wave functions: two-electron reduced density matrices. *Acc Chem Res* 39:207–215
48. Chaykin D (2009) Verification of semidefinite optimization problems with application to variational electronic structure calculation. PhD thesis, TU Hamburg-Harburg
49. Kutzelnigg W (1985) r_{12} -dependent terms in the wave function as closed sums of partial wave amplitudes for large I. *Theor Chim Acta* 68:445–469
50. Klopper W, Kutzelnigg W (1987) Møller-Plesset calculations taking care of the correlation cusp. *Chem Phys Lett* 134:17–22
51. Rychlewski J (2003) Explicitly correlated wave functions in chemistry and physics – theory and applications. Kluwer Academic, Dordrecht
52. Marchette O, Werner H-J (2008) Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys Chem Chem Phys* 10:3400–3409
53. Valeev EF (2006) Combining explicitly correlated R12 and Gaussian geminal electronic structure theories. *J Chem Phys* 125:244106-1/10
54. Curtiss LA, Redfern PC, Raghavachari K (2007) Gaussian-4 theory. *J Chem Phys* 126: 048108-1/12
55. Curtiss LA, Redfern PC, Raghavachari K (2005) Assessment of Gaussian-3 and density functional theories on the G3/05 test set of experimental energies. *J Chem Phys* 123:124107-1/12
56. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) A complete basis set model chemistry. *J Chem Phys* 110:2822–2827
57. Császár AG, Allen WA, Schaefer HF III (1998) In pursuit of the ab initio limit for conformational energy prototypes. *J Chem Phys* 108:9751–9765
58. Dahlke EE, Orthmeyer MA, Truhlar DG (2008) Assessment of multicoefficient correlation methods. *J Phys Chem B* 112:2372–2381
59. Bomble YJ, Vazquez J, Kallay M, Michank C, Szalay PG, Császár AG, Gauss J, Stanton JF (2006) High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J Chem Phys* 125:064108-1/8
60. Karton A, Taylor PR, Martin JML (2007) Basis set convergence of post-CCSD contributions to molecular atomization energies. *J Chem Phys* 127:064104-1/11
61. Hansen N, Kerber T, Sauer J, Bell AT, Keil FJ (2010) Quantum chemical modelling of benzene ethylation over H-ZSM-5 – approaching chemical accuracy: a Hybrid-MP2:DFT study. *J Am Chem Soc* 132:11525–11538
62. Sherrill CD (2010) Frontiers in electronic structure theory. *J Chem Phys* 132:110902
63. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev B* 136:864–871
64. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652
65. Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab-initio total-energy calculations using a plane-wave basis set. *Phys Rev B* 54:11169–11186
66. Segall MD, Lindan PLD, Probert MJ, Pickard CJ, Hasnip PJ, Clark SJ, Payne MC (2002) First principles simulation: ideas, illustrations, and the CASTEP code. *J Phys Condens Matter* 14:2717–2743

67. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Electronic structure calculations on workstation computers: the program system TURBOMOLE. *Chem Phys Lett* 162:165–169
68. Dovesi R, Saunders VR, Roetti R, Orlando R, Zicovick-Wilson CM, Pascale F, Civalieri B, Doll K, Harrison NM, Bush IJ, Llunell M (2009) CRYSTAL09 user's manual. University of Torino, Torino
69. Jaguar, Version 7.0 (2007) Schrödinger, LLC, New York, NY
70. Frisch MJ et al. (2009) Gaussian 09, Revision 1. Gaussian, Inc., Wallingford CT
71. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen SuS, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic structure system. *J Comput Chem* 14:1347–1363
72. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Van der Waals density functional for general geometries. *Phys Rev Lett* 92:246401
73. Thonhauser T, Cooper VR, Li S, Puzder A, Hyldgaard P, Langreth DC (2007) Van der Waals density functional: self-consistent potential and the nature of the van der Waals bond. *Phys Rev B* 76:125112
74. Rudenko A, Keil FJ, Katsnelson MI, Lichtenstein AI (2010) Adsorption of diatomic halogen molecules on graphene: a van der Waals density functional study. *Phys Rev B* 82:035427
75. Vydrov OA, Wu Q, Voorhis TV (2008) Self-consistent implementation of a nonlocal van der Waals density functional with a Gaussian basis set. *J Chem Phys* 129:014106
76. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate ab initio parametrisation of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 132:154104
77. Silvestri PL (2009) Van der Waals interactions in density functional theory using Wannier functions. *J Phys Chem A* 113:5224–5234
78. Chai J-D, Head-Gordon M (2008) Systematic optimization of long range corrected hybrid density functionals. *J Chem Phys* 128:084106-1/15
79. Chai J-D, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys Chem Chem Phys* 10:6615–6620
80. Pacchioni G (2008) Modeling doped and defective oxides in catalysis with density functional theory methods: room for improvement. *J Chem Phys* 128:182505
81. Metropolis NA, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
82. Anderson JB (2007) Quantum Monte Carlo. Oxford University Press, Oxford
83. Hammond BL, Lester WA, Reynolds PJ (1994) Monte Carlo in ab initio quantum chemistry. World Scientific, Singapore
84. Schwartz M (1987) Principles of electrodynamics. Dover Publications, New York
85. Swart M, van Duijnen PTh, Snijders JG (2001) A charge analysis derived from an atomic multisolet expansion. *J Comput Chem* 22:79–88
86. Sigfridsson E, Eyde U (1998) Comparison of methods for deriving atomic charges from electrostatic potential and moments. *J Comput Chem* 19:377–395
87. Tironi I, Sperber R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102:5451–5459
88. Hünenberger PH, van Gunsteren WF (1998) Alternative schemes for the inclusion of a reaction field correction into molecular dynamics simulations: influence on the simulated energetic, structural, and dielectric properties of liquid water. *J Chem Phys* 108:6117–6134
89. Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
90. Wood RH (1995) Continuum electrostatics in a computational universe with finite cutoff radii and periodic boundary conditions. Correction to computed free energies of ionic solvation. *J Chem Phys* 103:6177–6187
91. Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. *J Mol Liquids* 101:219–260

92. Saint-Martin H, Hernández-Cobos J, Ortega-Blake I (2005) Water models based on a single potential energy surface and different molecular degrees of freedom. *J Chem Phys* 122:224509-1/12
93. Hernández-Cobos J, Saint-Martin H, Mackie AD, Vega LF, Ortega-Blake I (2005) Water liquid-vapor equilibria predicted by refined ab initio derived potentials. *J Chem Phys* 123:044506-1/8
94. Allinger NL, Yuh YH, Lii J-H (1989) Molecular mechanics – the MM3 force-field for hydrocarbons. 1. *J Am Chem Soc* 111:8551–8566
95. Allinger NL, Chen KS, Lii J-H (1996) An improved force field (MM4) for saturated hydrocarbons. *J Comput Chem* 17:642–668
96. Mayo SL, Olafson BD, Goddard WA III (1990) DREIDING: a generic force field for molecular simulations. *J Phys Chem* 94:8897–8909
97. Allured VS, Kelly CM, Landis CR (1991) SHAPES empirical force field: a new treatment of angular potentials and its application to square-planar transition metal complexes. *J Am Chem Soc* 113:1–13
98. Root DM, Landis CR, Cleveland T (1993) Valence bond concepts applied to molecular mechanics description of molecular shapes. *J Am Chem Soc* 115:4201–4209
99. 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
100. Maple JR, Hwang M-J, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) Derivation of class II force fields. *J Comput Chem* 15:161–182
101. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profetta S, Weiner P (1984) A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106:765–784
102. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217
103. 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* 118:11225–11236
104. Halgren TA (1996) Merck molecular force field. 1. Basis form, scope, parametrization, and performance. *J Comput Chem* 17:490–519, 520–552, 553–586
105. Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CH_n united atoms of GROMOS96 force field. *J Comput Chem* 19:535–547
106. Massink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 111: 7812–7824
107. Jalaie M, Lipkowitz KB (2000) Published force field parameters for molecular mechanics, Molecular Dynamics and Monte Carlo simulations. In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 14. Wiley-VCH, Weinheim, p 441
108. Shelley JC, Shelley MY, Reeder RC, Bandyopadhyay S, Klein ML (2001) A coarse grain model for phospholipid simulations. *J Phys Chem B* 105:4464–4470
109. Elezgaray J, Laguerre M (2006) A systematic method to derive force fields for coarse-grained simulations of phospholipids. *Comput Phys Commun* 145:264–268
110. Izvekov S, Voth GA (2005) A multiscale coarse-graining method for biomolecular systems. *J Phys Chem B* 109:2469–2473
111. Schlick T (2002) *Molecular modeling and simulation*. Springer, Heidelberg
112. Leach AR (2001) *Molecular modeling – principles and applications*. Prentice Hall, New York
113. Crespos C, Collins MA, Pijper E, Kroes GJ (2003) Multi-dimensional energy surface determination by modified Shepherd interpolation for a molecule-surface reaction H₂ + Pt (111). *Chem Phys Lett* 376:566–575

114. Crespos C, Collins MA, Pijper E, Kroes GJ (2004) Application of the modified Shepherd interpolation method to the determination of the potential energy surface for a molecule-surface reaction: $\text{H}_2 + \text{Pt}(111)$. *J Chem Phys* 120:2392–2406
115. Zupan J, Gasteiger J (1993) Neural networks for chemists. VCH, Weinheim
116. Lorenz S, Scheffler M, Gross A (2006) Descriptions of surface chemical reactions using a neural network representation of the potential energy surface. *Phys Rev B* 73:115431-1/13
117. Behler J, Reuter K, Scheffler M (2008) Nonadiabatic effects in the dissociation of oxygen molecules at the $\text{Al}(111)$ surface. *Phys Rev B* 77:115421-1/16
118. van Duin ACT, Dasgupta S, Lorant F, Goddard WA III (2001) ReaxFF: a reactive force field for hydrocarbons. *J Phys Chem A* 105:9396–9409
119. Pauling L (1947) Atomic radii and interatomic distances in metals. *J Am Chem Soc* 69:542–553
120. Nielson KD, van Duin ACT, Oxgaard J, Deng W-Q, Goddard WA II (2005) Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes. *J Phys Chem A* 109:493–499
121. Chenoweth K, van Duin ACT, Persson P, Cheng M-J, Oxgaard J, Goddard WA III (2008) Development and application of a ReaxFF reactive force field for oxidative dehydrogenation on vanadium oxide catalysts. *J Phys Chem C* 112:14645–14654
122. Liu B, Lusk MT, Ely JF, van Duin ACT, Goddard WA III (2008) Reactive molecular dynamics force field for the dissociation of light hydrocarbons on $\text{Ni}(111)$. *Mol Simul* 34:967–972
123. Bortz AB, Kalos MH, Lebowitz JL (1975) New algorithm for Monte Carlo simulations of Ising spin systems. *J Comp Phys* 17:10–18
124. Fichtorn KA, Weinberg WH (1991) Theoretical foundations of dynamical Monte Carlo simulations. *J Chem Phys* 95:1090–1096
125. Van Kampen NG (2001) Stochastic processes in physics and chemistry. Elsevier, Amsterdam
126. Lekkien JJ, Segers JPL, Hilbers PAJ, Gelten RJ, Jansen APJ (1998) Efficient Monte Carlo methods for the simulation of catalytic surface reactions. *Phys Rev E* 58:2598–2610
127. Goldstein H, Poole C, Safko J (2002) Classical mechanics. Pearson, Addison Wesley, San Francisco
128. Leimkuhler BJ, Skeel RD (1994) Symplectic numerical integrators in constrained Hamiltonian systems. *J Comput Phys* 112:117–125
129. Trotter HF (1959) On the product of semi-groups of operators. *Proc Am Math Soc* 10:545–551
130. Reed M, Simon B (1980) Methods of modern mathematical physics. I Functional analysis, 2nd edn. Academic, New York
131. Hassani S (1999) Mathematical physics. Springer, Heidelberg
132. Buneman O (1967) Time-reversible difference procedures. *J Comp Phys* 1:517–535
133. Tuckerman ME, Berne BJ, Martyna GJ (1992) Reversible multiple time scales molecular dynamics. *J Chem Phys* 97:1990–2001
134. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable Molecular Dynamics with NAMD. *J Comput Chem* 26:1781–1802
135. Shaw DE (2005) A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J Comput Chem* 26:1318–1328
136. Ma Q, Izaguirre J, Skeel R (2003) Verlet-I/R-Respa/Impulse is limited by nonlinear instabilities. *J Sci Comput* 24:1951–1973
137. Morrone JA, Zhou R, Berne BJ (2010) Molecular dynamics with multiple time scales: how to avoid pitfalls. *J Chem Theor Comput* 6:1798–1804
138. Skeel RD (1999) Integration schemes for molecular dynamics and related applications. In: Ainsworth M, Levesley J, Marletta M (eds) The graduate student's guide to numerical analysis. Springer, New York, pp 119–176

139. Schneider T, Stoll E (1978) Molecular dynamics study of a three-dimensional n-component model for distortive phase transitions. *Phys Rev B* 17:1302–1322
140. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys* 72:2384–2393
141. Hoover WG, Ladd AJC, Moran B (1982) High strain rate plastic flow studied via non-equilibrium molecular dynamics. *Phys Rev Lett* 48:1818–1820
142. Evans DJ, Hoover WG, Failor BH, Moran B, Ladd AJC (1983) Nonequilibrium molecular dynamics via Gauss's principle of least constraint. *Phys Rev A* 28:1016–1021
143. Evans DJ, Morriss GP (1983) The isothermal isobaric molecular dynamics ensemble. *Phys Lett A* 98:433–436
144. Evans DJ, Morriss GP (1984) Non-Newtonian molecular dynamics. *Phys Rep* 1:297–344
145. Tuckerman ME, Mundy CJ, Martyna GJ (1999) On the classical statistical mechanics of non-Hamiltonian systems. *Europhys Lett* 45:149–155
146. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
147. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics method. *J Chem Phys* 81:511–519
148. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52:255–268
149. Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A* 31:1696–1697
150. Jakobtorweihen S, Verbeeck MG, Lowe CP, Keil FJ, Smit B (2005) Understanding the loading dependence of self-diffusion in carbon nanotubes. *Phys Rev Lett* 95:044501-1/4
151. Jakobtorweihen S, Lowe CP, Keil FJ, Smit B (2006) A novel algorithm to model the influence of host lattice flexibility in molecular dynamics simulations: loading dependence of self-diffusion in carbon nanotubes. *J Chem Phys* 124:154706-1/13
152. Jakobtorweihen S, Lowe CP, Keil FJ, Smit B (2007) Diffusion of chain molecules and mixtures in carbon nanotubes: the effect of host lattice flexibility and theory of diffusion in the Knudsen regime. *J Chem Phys* 127:024904-1/11
153. Swendsen RH, Wang JS (1986) Replica Monte Carlo simulation of spin-glasses. *Phys Rev Lett* 57:2607–2609
154. Woods CJ, Esser JW, King MA (2003) The development of replica-exchange based free-energy methods. *J Phys Chem B* 107:13703–13710
155. Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys Rev Lett* 78:3908–3912
156. Laio A, Parinello M (2002) Escaping the free energy minima. *Proc Natl Acad Sci* 99:12562–12566
157. Carter E, Cicotti G, Hynes JT, Kapral R (1989) Constrained reaction coordinate dynamics for the simulation of rare events. *Chem Phys Lett* 156:472–477
158. Sprik M, Cicotti G (1998) Free energy from constrained molecular dynamics. *J Phys Chem* 109:7737–7744
159. Torrie GM, Valleau JP (1974) Monte-Carlo free-energy estimates using non-Boltzmann sampling – application to subcritical Lennard-Jones fluid. *Chem Phys Lett* 28:578–581
160. Rosso L, Tuckerman ME (2002) An adiabatic molecular dynamics method for the calculation of free energy profiles. *Mol Simul* 28:91–112
161. Chipot C, Pohorille A (2007) Free energy calculations: theory and applications in chemistry and biology. Springer, Heidelberg
162. Washel A, Levitt M (1976) Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 103:227–249
163. Sierka M, Sauer J (2005) Hybrid quantum mechanics/molecular mechanics methods and their application. In: Yip S (ed) *The handbook of materials modeling, Part A. Methods*. Springer, Dordrecht, pp 241–258

164. Hu H, Yang W (2008) Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annu Rev Phys Chem* 59:573–601
165. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 48:1198–1229
166. York DM, Lee T-S (2009) Multi-scale quantum models for biocatalysis. Springer, Dordrecht
167. Lin H, Truhlar DG (2007) QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc* 117:185–199
168. Park JH, Heyden A (2009) Solving the equations of motion for mixed atomistic and coarse-grained systems. *Mol Simul* 35:962–973
169. Mata RA, Werner HJ, Thiel S, Thiel W (2008) Toward accurate barriers for enzymatic reactions: QM/MM case study on p-hydroxybenzoate hydroxylase. *J Chem Phys* 128: 025104-1/8
170. Bakowies D, Thiel W (1996) Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J Phys Chem* 100:10580–10594
171. König PH, Hoffmann M, Frauenheim T, Cui Q (2005) A critical evaluation of different QM/MM frontier treatments using SCC-DFTB as the QM method. *J Phys Chem B* 109: 9082–9095
172. 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* 113:9978–9985
173. E W, Ren W, Vanden Eijnden E (2002) String method for the study of rare events. *Phys Rev B* 66:052301-1/4
174. Baker J (1986) An algorithm for the location of transition states. *J Comp Chem* 7:385–395
175. Henkelman G, Jónsson H (1999) A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J Chem Phys* 111:7010–7022
176. Hänggi P, Talkner P, Borkovec M (1990) Reaction-rate theory: fifty years after Kramers. *Rev Mod Phys* 62:251–341
177. Bucko T, Hafner J (2010) Entropy effects in hydrocarbon conversion reactions: free energy integrations and transition path sampling. *J Phys Condens Matter* 22:384201
178. Pratt LR (1986) A statistical method for identifying transition states in high dimensional problems. *J Chem Phys* 85:5045–5048
179. Bolhuis PG, Chandler D, Dellago C, Gassler PL (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* 53:291–318
180. Berendsen HJC (2007) *Simulating the physical world*. Cambridge University Press, Cambridge
181. Broadbelt LJ, Snurr RQ (2000) Applications of molecular modelling in heterogeneous catalysis research. *Appl Catal A* 200:23–46
182. Hansen N, Krishna R, van Baten JM, Bell AT, Keil FJ (2009) Analysis of diffusion limitation in the alkylation of benzene over H-ZSM-5 by combining quantum chemical calculations, molecular simulations, and a continuum approach. *J Phys Chem C* 113:235–246
183. Hansen N, Krishna R, van Baten JM, Bell AT, Keil FJ (2010) Reactor simulation of benzene ethylation and ethane dehydrogenation catalyzed by ZSM-5: a multiscale approach. *Chem Eng Sci* 65:2472–2480
184. Raimondeau S, Vlachos DG (2002) Recent developments on multiscale, hierarchical modelling of chemical reactors. *Chem Eng J* 90:3–23
185. Vlachos D (2005) A review of multiscale analysis: examples from systems biology, materials engineering, and other fluid-surface interacting systems. *Adv Chem Eng* 36:1–61
186. Christensen CH, Norskov JK (2008) A molecular view of heterogeneous catalysis. *J Chem Phys* 128:182503-1/8
187. Hansen N, Kerber T, Sauer J, Bell AT, Keil FJ (2010) Quantum chemical modeling of benzene ethylation over H-ZSM-5 approaching chemical accuracy: a hybrid MP2:DFT study. *J Am Chem Soc* 132:11525–11538
188. Heyden A, Peters B, Bell AT, Keil FJ (2005) Comprehensive DFT study of nitrous oxide decomposition over Fe-ZSM-5. *J Phys Chem B* 109:1857–1873

189. Heyden A, Hansen N, Bell AT, Keil FJ (2006) Nitrous oxide decomposition over Fe-ZSM-5 in the presence of nitric oxide: a comprehensive DFT study. *J Phys Chem B* 110:17096–17114
190. Heyden A, Bell AT, Keil FJ (2005) Kinetic modelling of nitrous oxide decomposition on Fe-ZSM-5 based on parameters obtained from first-principles calculation. *J Catal* 233:26–35
191. Brüggemann TC, Przybylski M-D, Balaji SP, Keil FJ (2010) Theoretical investigation of the mechanism of the selective catalytic reduction of nitrogen dioxide with ammonia on H-form zeolites and the role of nitric and nitrous acids as intermediates. *J Phys Chem C* 114: 6567–6587
192. Liu X, Newsome D, Coppens M-O (2009) Dynamic Monte Carlo simulations of binary self-diffusion in ZSM-5. *Microporous Mesoporous Mater* 125:149–159
193. Tuma C, Sauer J (2006) Treating dispersion effects in extended systems by hybrid MP2:DFT calculations – protonation of isobutene in zeolite ferrierite. *Phys Chem Chem Phys* 8: 3955–3965
194. Myers AL, Prausnitz JM (1965) Thermodynamics of mixed gas adsorption. *AIChE J* 11: 121–127
195. Huang P, Carter EA (2008) Advances in correlated electronic structure methods for solids, surfaces, and nanostructures. *Annu Rev Phys Chem* 59:261–290
196. E W, Vanden-Eijnden E (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem* 61:391–420

Real-World Predictions from Ab Initio Molecular Dynamics Simulations

Barbara Kirchner, Philipp J. di Dio, and Jürg Hutter

Abstract In this review we present the techniques of ab initio molecular dynamics simulation improved to its current stage where the analysis of existing processes and the prediction of further chemical features and real-world processes are feasible. For this reason we describe the relevant developments in ab initio molecular dynamics leading to this stage. Among them, parallel implementations, different basis set functions, density functionals, and van der Waals corrections are reported. The chemical features accessible through AIMD are discussed. These are IR, NMR, as well as EXAFS spectra, sampling methods like metadynamics and others, Wannier functions, dipole moments of molecules in condensed phase, and many other properties. Electrochemical reactions investigated by ab initio molecular dynamics methods in solution, on surfaces as well as complex interfaces, are also presented.

Keywords Ab initio molecular dynamics simulations · Always stable predictor-corrector algorithm · Associated liquids · Basis set · Born–Oppenheimer molecular dynamics simulations · Car–Parrinello molecular dynamics simulations · Catalysis · Collective variable · Discrete variable representation · Dispersion · Effective core potential · Enhanced sampling · Fictitious mass · First-principles molecular dynamics simulations · Free energy surface · Hartree–Fock exchange · Ionic liquids · Linear scaling · Metadynamics · Nudged elastic band · Numerically tabulated atom-centered orbitals · Plane waves · Pseudopotential · Rare event · Relativistic electronic structure · Retention potential · Self consistent field · SHAKE algorithm ·

B. Kirchner (✉) and P.J. di Dio
Wilhelm-Ostwald Institute of Physical and Theoretical Chemistry, University of Leipzig, Linnéstr. 2,
04103 Leipzig, Germany
e-mail: bkirchner@uni-leipzig.de

J. Hutter
Physical Chemistry Institute, University of Zürich, Winterthurerstr. 190, CH-8057 Zürich,
Switzerland
e-mail: hutter@pci.uzh.ch

Solvent effect · String method · van der Waals interaction · Wannier orbitals · Water · Wavelets

Contents

1	Introduction	111
2	Ab Initio Molecular Dynamics Simulations in a Nutshell	112
2.1	Molecular Dynamics Simulations: Basics	112
2.2	Obtaining the Forces and Integrating the Equations of Motion	113
2.3	Born–Oppenheimer Molecular Dynamics Simulations	114
2.4	Car–Parrinello Molecular Dynamics Simulations	115
2.5	Generalization of the Car–Parrinello and Born–Oppenheimer Molecular Dynamics Approaches	117
3	Faster, Larger, and More Accurate: Recent Developments	121
3.1	Massively Parallel Implementation	121
3.2	Basis Sets	122
3.3	New Developments in Accuracy	124
3.4	New Integration Schemes	129
3.5	Enhanced Sampling	130
3.6	Properties: IR, NMR and EXAFS	133
4	Applications in Chemical Engineering	136
4.1	Wavefunction Analysis	136
4.2	Properties of the Vapor Phase, Liquids, Mixtures, and Solvent Effects	137
4.3	Chemical Reactions	141
4.4	Electrochemistry	143
5	Summary	147
	References	148

Abbreviations

AIMD	Ab initio molecular dynamics; molecular dynamics with electronic structure calculations on the fly
ASPC	Always stable predictor–corrector algorithm
BOMD	Born–Oppenheimer molecular dynamics; molecular dynamics with electronic structure calculations on the fly, diagonalization in each step
CPMD	Car–Parrinello molecular dynamics; molecular dynamics with electronic structure calculations on the fly, orthogonalization in each step otherwise the coefficients of the wavefunction are propagated like the nuclear positions
CV	Collective variables
DFT	Density functional theory; static quantum chemical method using functionals of the electronic density to account for electron correlation
DVR	Discrete variable representation
ECP	Effective core potential also called pseudopotential
FBR	Finite basis representation
FES	Free energy surface
FFT	Fast Fourier transformation

GGA	Generalized gradient approximation (GGA) functional, a functional that depends on density and its gradient
HF	Hartree–Fock, static quantum chemical method
IR	Infrared red
loc	Local functional depending only on $\mathbf{r}$
MD	Molecular dynamics, simulation method
MEP	Minimum energy path
MFEP	Minimum free energy path
MLWC	Maximally localized Wannier centers
MLWO	Maximally localized Wannier orbitals
MTD	Metadynamics, method to calculate rare events
NAO	Numerically tabulated atom-centered orbitals
NEMD	Non-equilibrium molecular dynamics
nl	Non-local, functional depending not only on $\mathbf{r}$ but also on $\mathbf{r}'$
NPT	NPT ensemble: isothermal-isobaric ensemble; constant particle (N), pressure (P), and temperature (T) simulation
PBC	Periodic boundary conditions
QM/MM	Hybrid quantum-mechanical/molecular-mechanical calculations
RPA	Random phase approximation
SCF	Self consistent field
vdW	van der Waals, dispersion forces; usually not well-described in DFT

1 Introduction

Ab initio molecular dynamics (AIMD) simulations combine classical molecular dynamics simulations with electronic structure calculations on the fly. The theoretical foundations for ab initio molecular dynamics were laid with the work of Ehrenfest [1] and Dirac [2] at the beginning of the twentieth century. Dirac developed the theory of time-dependent self consistent field (SCF) equations for nuclear and electronic motion and Ehrenfest derived mixed classical-quantum mechanical (time-dependent electronic structure) equations [3]. In 1985 it was the seminal article of Roberto Car and Michele Parrinello [4] which initiated the use and further development of ab initio molecular dynamics simulations. The authors intended to derive a new method which is able to “(1) compute ground-state electronic properties of large and/or disordered systems using state-of-the-art electronic structure calculations; (2) perform AIMD simulations where the only assumptions are the validity of classical mechanics to describe ionic motion and the Born–Oppenheimer (BO) approximation to separate nuclear and electronic coordinates” [4]. For this purpose Car and Parrinello made use of the extended Lagrangian technique, previously invented to simulate systems under constant pressure [5, 6]. This ingenious method solved the problem of the expensive self-consistent solution to the electronic structure problem along the molecular dynamics trajectory. By showing a feasible route to extensive ab initio molecular dynamics

simulations of condensed systems, Car and Parrinello kick started and dominated the field. Their method stays at the beginning of all new developments in the field.

Section 2 provides a brief overview of the AIMD methodology mainly in the representation of Car–Parrinello molecular dynamics simulations. Born–Oppenheimer molecular dynamics (BOMD) simulations (time-independent electronic structure) are introduced in a generalized formulation based on the work by Niklasson [7, 8]. This will be followed (Sect. 3) by some recent methodological advancements which allow for computationally more efficient simulations with better statistical sampling and which use more accurate electronic structure methods. After this, some examples from applied chemistry studied from AIMD will be given in Sect. 4.

2 Ab Initio Molecular Dynamics Simulations in a Nutshell

2.1 Molecular Dynamics Simulations: Basics

Molecular dynamics (MD) is an application of classical mechanics using computer simulations. Good introductions can be found in many textbooks, for example the excellent book by Tuckerman [9]. In order to carry out MD, equations describing the motion of molecules are needed. These equations of motion can be derived for example from the classical Lagrangian $\mathcal{L}$, a function of the kinetic (K) and the potential energy (U):

$$\mathcal{L}(\mathbf{R}_I, \dot{\mathbf{R}}_I) = \underbrace{K(\mathbf{p}_I)}_{\text{kin. energy}} - \underbrace{U(\mathbf{R}_I)}_{\text{potential}} = \sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 - U(\mathbf{R}_I) \quad (1)$$

with $\mathbf{R}_I$ and M_I being position and mass of particle I . The momentum $\mathbf{p}_I$ is related to the velocity $\dot{\mathbf{R}} = \mathbf{p}_I/M_I$. The equations of motion are then obtained from the Euler–Lagrange relation:

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} = \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I}. \quad (2)$$

This means that the nuclei (atoms) are treated as classical particles, a well founded and tested approximation [10, 11]. Applying the Euler–Lagrange equation (2) to the Lagrangian $\mathcal{L}$ (1) leads to Newton’s second law:

$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_I. \quad (3)$$

Using Newton's second law (3) the acceleration ($\ddot{\mathbf{R}}_I$) of the particles can be obtained from the forces $\mathbf{F}_I$ acting on the particles.

2.2 *Obtaining the Forces and Integrating the Equations of Motion*

Traditional molecular dynamics simulations use pre-defined analytical potentials. The potentials most commonly work with the pairwise additivity approximation ($U(R_{IJ})$). This means that an analytical expression of the pair potential, a potential between each set of atoms I and J , is parameterized such that good structural and/or thermodynamics quantities can be expected [12]. Alternatively, electronic structure calculations are carried out for a pair of particles as a function of distance and the analytical expression has to be fitted to these energy points on the potential energy surface [12]. It also means that the Born–Oppenheimer approximations have to be valid, i.e., a separation of nuclear and electronic variables is possible and coupling terms (non-diagonal and diagonal) can be neglected [13, 14]. From the analytical potentials the forces are then obtained by taking the derivatives with respect to the positions

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

In order to propagate the atoms, a small discrete time step Δt is introduced and a numerical step-by-step integration of the equations of motion is carried out. Taking the Taylor series expansion in Δt gives

$$\mathbf{R}_I(t + \Delta t) = \mathbf{R}_I(t) + \Delta t \dot{\mathbf{R}}_I(t) + \frac{1}{2} \Delta t^2 \ddot{\mathbf{R}}_I(t) + \frac{1}{6} \Delta t^3 B_I(t) + \dots \quad (5)$$

and

$$\dot{\mathbf{R}}_I(t + \Delta t) = \dot{\mathbf{R}}_I(t) + \Delta t \ddot{\mathbf{R}}_I(t) + \frac{1}{2} \Delta t^2 B_I(t) + \dots \quad (6)$$

The time evolution of the system is followed by applying integration algorithms (the so-called integrator) in an MD computer program. One can obtain these integrators from the Taylor expansion around $t + \Delta t$ and $t - \Delta t$ and by combining the resulting equations. The following form is the velocity Verlet (Strömer–Verlet) integrator:

$$\mathbf{R}_I(t + \Delta t) = \mathbf{R}_I(t) + \dot{\mathbf{R}}_I(t) \Delta t + \frac{\mathbf{F}_I(t)}{2M_I} \Delta t^2, \quad (7)$$

$$\dot{\mathbf{R}}_I(t + \Delta t) = \dot{\mathbf{R}}_I(t) + \frac{F_I(t + \Delta t) + F_I(t)}{2M_I} \Delta t. \quad (8)$$

It is apparent how the new positions $\mathbf{R}_I(t + \Delta t)$ can be calculated from the current positions $\mathbf{R}_I(t)$, velocities $\dot{\mathbf{R}}_I(t)$, and forces $\mathbf{F}_I(t)$. Similarly, the new velocities can be obtained from knowledge of current velocities and forces $\mathbf{F}_I(t)$ as well as from the new forces which are available as soon as the new positions (7) are calculated. An overview over how integration algorithms are derived is provided in [10].

In many molecular systems it is desirable to freeze fast degrees of freedom. This can be necessary in order to allow the integration of the slower motions using larger time steps. Or the freezing of fast degrees of freedom might be necessary if the quantum nature of such degrees of freedom (e.g., bond stretch vibrations including hydrogen atoms) are important. A technique developed [10, 15] to handle properly such constraints to the molecular structure in molecular dynamics simulation is based on undetermined multipliers. The constraint conditions with the undetermined multipliers are added to the Lagrangian of (1). The constraint condition then gives rise to additional (constraint) forces $\mathbf{G}_I$ in the equation of motion

$$M_I \ddot{\mathbf{R}}_I = F_I + G_I. \quad (9)$$

The constraint forces depend linearly on the multipliers which have to be determined in accordance with the numerical integration scheme. This usually leads to nonlinear equations which can in special cases be solved directly. However, the most common algorithm, called SHAKE [15], solves the equations iteratively, until self consistency between input and output multipliers is achieved.

In order to avoid surface effects for condensed phase simulations, periodic boundary conditions are applied. The central computational box is replicated infinitely in all dimensions. A detailed description can be found in the textbooks of Allen and Tildesley [10] as well as of Frenkel and Smit [11].

2.3 Born–Oppenheimer Molecular Dynamics Simulations

Instead of using a pre-parameterized potential, the potential can be calculated on the fly using electronic structure theory within the Born–Oppenheimer approximation. In such calculations the potentials are obtained by solving a time-independent quantum chemical electronic structure problem:

$$U(\mathbf{R}_I) = \min_{\{\Psi\}} E[\{\Psi\}; \mathbf{R}_I]. \quad (10)$$

The Schrödinger equation with an optimized electronic wavefunction Ψ at current nuclear position $\mathbf{R}_I$ is solved in order to provide the forces for calculating the next step.

2.4 Car–Parrinello Molecular Dynamics Simulations

The basic idea of Car–Parrinello molecular dynamics simulations can be expressed by “A two-component quantum-classical problem is mapped onto a two-component purely classical problem with employing the constraints that quantum mechanics has to be fulfilled at all times” [3].

This leads to two separate energy scales at the expense of losing the physical time information of the quantum subsystem dynamics. The corresponding Lagrangian (1) reads

$$\mathcal{L} = \underbrace{\sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \mu \left\langle \dot{\phi}_i | \dot{\phi}_i \right\rangle}_{\text{kin. energy}} - \underbrace{\varepsilon^{el}[\{\phi_i\}; \mathbf{R}_I]}_{\text{potential}}. \quad (11)$$

The second term of (11) describes the fictitious kinetic energy of the electrons. The term contains an arbitrary parameter (fictitious mass parameter) μ with appropriate units of energy times a squared time. The explicit form used in (11) is for orbital based methods.

How this mass parameter has to be chosen is extensively discussed in [3]. A critical point of view about the fictitious mass parameter and about arguments used for the justification of the CPMD approach is given in [16]. The dot in this Lagrangian indicates the time derivative; thus it is apparent that the wavefunction fulfils the same task as the nuclear position variable. The potential is now a functional of the electronic energy E^{el} plus the constraints which are enforced in order to satisfy quantum mechanics, i.e., the orbitals which are altered during time evolution are supposed to stay orthonormal; see second term of (12). The additional constraint is introduced by the standard Lagrange multipliers approach, where the Λ_{ij} are the Lagrange multipliers and δ_{ij} is the Kronecker delta:

$$\mathcal{E}[\{\phi_i\}; \mathbf{R}_I] = E^{el} + \sum_{ij} \Lambda_{ij} \left(\left\langle \phi_i | \phi_j \right\rangle - \delta_{ij} \right). \quad (12)$$

As the electronic energy is a function of the nuclear positions as well as function of the orbitals ϕ_i , its derivative is taken with respect to the nuclear positions but also with respect to the wavefunction. The Euler–Lagrange equations then read

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} = \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I} \quad \text{and} \quad \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \langle \dot{\phi}_i |} = \frac{\partial \mathcal{L}}{\partial \langle \phi_i |}. \quad (13)$$

This leads again to the equations of motion, in this case to the Car–Parrinello equations of motion, given below:

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E^{el}}{\partial \mathbf{R}_I} + \sum_{ij} \Lambda_{ij} \frac{\partial}{\partial \mathbf{R}_I} \langle \phi_i | \phi_j \rangle, \quad (14)$$

$$\mu |\ddot{\phi}_i\rangle = -\frac{\delta E^{el}}{\delta \langle \phi_i |} + \sum_j \Lambda_{ij} |\phi_j\rangle. \quad (15)$$

The forces on the nuclei are given by

$$\mathbf{F}(\mathbf{R}_I) = -\frac{\partial E^{el}}{\partial \mathbf{R}_I} + \sum_{ij} \Lambda_{ij} \frac{\partial}{\partial \mathbf{R}_I} \langle \phi_i | \phi_j \rangle \quad (16)$$

and the forces on the electrons are given by

$$\mathbf{f}_i = \mathbf{f}(\phi_i) = -\frac{\delta E^{el}}{\delta \langle \phi_i |} + \sum_j \Lambda_{ij} |\phi_j\rangle. \quad (17)$$

The integration algorithm for these equations of motion is given by the scheme below. The resemblance to (7) is immediately apparent. Because we follow the line of [3] these equations are valid for the special case of Kohn–Sham density functional theory in the plane wave basis set representation as electronic structure method. This leads to the simplification that the second term of (14) vanishes, because the orthonormality constraint is independent of the nuclear positions. Furthermore, the wavefunction (now Kohn–Sham orbitals in the plane wave representation) are replaced by the expansion coefficients $\mathbf{c}_i$ of the plane waves. The possibility of different integrators for equations of motion was alluded to above; the following scheme for the expansion coefficients is also based on the velocity Verlet algorithm. The constraint is enforced by the Rattle algorithm [10, 11]; therefore, the new “positions” of the coefficients *without applying constraints* $\tilde{\mathbf{c}}_i(t + \Delta t)$ read

$$\tilde{\mathbf{c}}_i(t + \Delta t) = \mathbf{c}_i(t) + \dot{\mathbf{c}}_i(t) \Delta t + \frac{\mathbf{f}_i(t)}{2\mu} \Delta t^2, \quad (18)$$

and then the constraints are corrected which is expressed in the second term of

$$\mathbf{c}_i(t + \Delta t) = \tilde{\mathbf{c}}_i(t + \Delta t) + \sum_j \frac{\Lambda_{ij}^c}{2\mu} \Delta t^2 \mathbf{c}_j(t). \quad (19)$$

The Lagrange multipliers Λ_{ij}^c of the constraints depend now only on the electronic part. For their determination see [3]. Of course the nuclei are also propagated, their positions being obtained according to (7). From these new “positions,” i.e., new nuclear positions and new coefficients, the forces on the nuclei $\mathbf{F}(\mathbf{R}_I)$ and those on the electrons $\mathbf{f}_i$ are obtained. Again, the “velocities” of the coefficients are derived as

$$\dot{\mathbf{c}}'_i(t + \Delta t) = \dot{\tilde{\mathbf{c}}}_i(t + \Delta t) + \frac{\mathbf{f}_i(t + \Delta t)}{2\mu} \Delta t. \quad (20)$$

They are corrected afterwards by determining the constraints

$$\dot{\mathbf{c}}_i(t + \Delta t) = \dot{\mathbf{c}}'_i(t + \Delta t) + \sum_j \frac{\Lambda_{ij}^c}{2\mu} \Delta t \mathbf{c}_j(t + \Delta t). \quad (21)$$

The difference from classical force field based simulations where the forces are calculated from pre-defined pair potentials is that the forces are derived from the global potential energy surface of an electronic structure theory. The vastly higher computational costs of an electronic structure calculation restrict the system size and the length of trajectories accessible by ab initio molecular dynamics simulations. However, it becomes clear that CPMD and AIMD are important steps towards general predictive methods, due to their independence from parameterizations.

2.5 Generalization of the Car–Parrinello and Born–Oppenheimer Molecular Dynamics Approaches

In order to allow for higher order symplectic or geometric integration schemes, Anders Niklasson et al. introduced a Lagrangian generalization of the time-reversible Born–Oppenheimer molecular dynamics simulations [7, 8].

Integrators in molecular dynamics simulations are supposed to be accurate, i.e., they should enforce the exact trajectory being followed as closely as possible. They should provide stability, meaning that the constants of motion, e.g., the total energy in the microcanonical ensemble, are preserved. Nevertheless, the integrators should be efficient, which means that a minimum number of force calculations are needed in order to save computer time. The best numerical methods are based on

symplectic and time-reversible integrators. For integrator algorithms see Sect. 2.2, in which the Strömer–Verlet integrator was introduced. However, in general, long time stability is more important than short time accuracy.

For ab initio molecular dynamics simulations it is important to understand how the errors in the forces affect the long term MD stability of the simulations. While the error in the energy [$\min_{\phi} E[\{\phi\}, \mathbf{R}_I]$, see (10)] is, due to the variational principle, of second order in the error $\delta\phi$ of the wave functions, the error in the force ($dE[\{\phi\}, \mathbf{R}_I]/d\mathbf{R}_I$) is of first order in $\delta\phi$. This suggests that MD stability can only be achieved with numerically highly accurate wavefunctions.

In the following the indices for the nuclei and the electrons will be omitted. Applying the extended Lagrangian method introduced by Niklasson [7, 8] a general expression for the AIMD Lagrangian can be written as

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}M\dot{\mathbf{q}}^2 + \frac{1}{2}\mu\dot{\mathbf{x}}^2 - E(\mathbf{q}, \mathbf{y}) + k\mu G(\|\mathbf{x} - \mathbf{y}\|). \quad (22)$$

$\mathbf{q}$ and $\mathbf{x}$ are now generalized coordinates of the nuclei and electrons, respectively. The vector $\mathbf{y}$ expresses the wave function after complete or partial optimization:

$$\mathbf{y} = F(\mathbf{q}, \mathbf{x}). \quad (23)$$

$G(\|\mathbf{x} - \mathbf{y}\|)$ is a retention potential that ensures that the propagated wavefunction $\mathbf{x}$ stays close to the optimized wavefunction $\mathbf{y}$ and μ is a mass parameter, $k = \omega^2$ is the force constant of the retention potential.

From the generalized Lagrangian follows the equations of motion

$$M\ddot{\mathbf{q}} = -\frac{\partial E}{\partial \mathbf{q}} - \frac{\partial E}{\partial \mathbf{y}} \frac{\partial F}{\partial \mathbf{q}} + k\mu \frac{\partial G}{\partial \mathbf{y}} \frac{\partial F}{\partial \mathbf{q}} \quad (24)$$

and

$$\mu\ddot{\mathbf{x}} = -\frac{\partial E}{\partial \mathbf{y}} \frac{\partial F}{\partial \mathbf{x}} + k\mu \left[\frac{\partial G}{\partial \mathbf{q}} + \frac{\partial G}{\partial \mathbf{y}} \frac{\partial F}{\partial \mathbf{x}} \right]. \quad (25)$$

In this notation the Car–Parrinello molecular dynamics scheme is obtained with the condition

$$\mathbf{y} = \mathbf{x} \quad \Rightarrow \quad G(\|\mathbf{x} - \mathbf{y}\|) = 0. \quad (26)$$

This leads directly to the Car–Parrinello Lagrangian [see (11)]

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}, \mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2}M\dot{\mathbf{q}}^2 + \frac{1}{2}\mu\dot{\mathbf{x}}^2 - E(\mathbf{q}, \mathbf{x}) \quad (27)$$

and, accordingly, the equations of motion; see (14) and (15). The accuracy of CPMD simulations with respect to Born–Oppenheimer surface as well as the efficiency, through the maximal time step δt , is controlled by the fictitious mass μ . The two conditions, accuracy and efficiency, have conflicting requirements on the mass and usually a rather large μ value is selected. As the fictitious mass changes the dynamics of the system, a renormalization of all dynamic quantities, for example the vibrational spectrum, is needed. However, the stability of the CPMD trajectory is excellent, because the calculations of the forces can be carried out easily to high precision. The efficiency of CPMD calculations is strongly system dependent as the maximal time step is dictated by the electronic energy gap.

Born–Oppenheimer molecular dynamics is obtained from the generalized Lagrangian with the conditions

$$\mathbf{y} = \arg \min_{\mathbf{x}} E(\mathbf{q}, \mathbf{x}) \quad \text{and} \quad \mu = 0. \quad (28)$$

The Born–Oppenheimer Lagrangian is thus simply

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} M \dot{\mathbf{q}}^2 - E(\mathbf{q}, \mathbf{y}). \quad (29)$$

Next the respective equations of motion are decoupled:

$$M \ddot{\mathbf{q}} = - \frac{\partial E}{\partial \mathbf{q}} \quad \text{and} \quad \ddot{\mathbf{x}} = -k \frac{\partial G}{\partial \mathbf{x}}.$$

This set of equations suggests a simple scheme where the nuclei are propagated on the Born–Oppenheimer surface according to the forces calculated from the derivatives of the optimized Kohn–Sham energy functional. Furthermore, an independent dynamical system can be applied to propagate an initial guess for the wavefunction. Accuracy and stability of the nuclei propagation are seemingly independent from the details, initial guess, and optimization method of the electronic structure calculation. The variational nature of the energy functional guarantees a quadratic dependence of the error in energy on the residual error in wavefunction. However, the derivative with respect to nuclear coordinates will show a linear dependence on the error in wavefunction. A rather tight convergence of wavefunction will be needed and this will require a rather large number of SCF cycles. It is therefore natural to try to reduce the number of SCF cycles by using a cleverly chosen initial wavefunction. A first simple choice would be to use the last optimized wavefunction along the trajectory. However, simulations using this approach have been shown to possess large drifts in total energy, even for tight SCF convergence criteria. This problem has been analyzed by Niklasson and coworkers [7, 8] in a series of papers. They realized that the incomplete SCF convergence leads to errors in the energy and forces of the Born–Oppenheimer system. Because the forces are calculated under the assumption of complete

optimization, i.e., the wavefunction derivatives are neglected, energy and forces are also no longer fully consistent. This would constitute, for energies sufficiently close to the Born–Oppenheimer energy, no real problem if these errors did not depend on the optimization method and initial guess wavefunction. Initial wavefunctions selected in a non-time reversible way, e.g., by taking the last optimized wavefunction, will transfer this property with the force error to the nuclear dynamics. The nuclear dynamics is therefore no longer time reversible, despite a seemingly time reversible integration algorithm, and very poor stability of integration results. To cure this problem a time reversible propagation of the initial wavefunction has to be chosen.

A time reversible BO molecular dynamics scheme based on the propagation of one-particle density matrices has been proposed by Niklasson [8]. The equation of motion for the density matrix is

$$\ddot{P} = \omega^2(D - P). \quad (30)$$

Here, D is the self-consistent, optimized density matrix. Using a time-reversible Verlet scheme we get an explicit integration of the form

$$P(t + \delta t) = 2P(t) - P(t - \delta t) + \delta t^2 \omega^2 [D(t) - P(t)]. \quad (31)$$

With $\kappa = \delta t^2 \omega^2 = 2$ as the original form proposed, using stability analysis the largest possible value of κ can be determined. A larger value of κ is desirable, as for a given time step δt a larger κ corresponds to a stiffer harmonic potential keeping the propagated density matrix P closer to the optimized density matrix D . The largest value of κ that is consistent with stability is the one that guarantees that the distance $D(t) - P(t)$ does not diverge. For the Verlet family of integrators this optimal value is in fact $\kappa = 2$.

The propagation of wavefunctions expanded in atom centered basis functions needs special care. It is best to use an extrapolated contra-covariant density matrix PS as a projector on to the occupied subspace

$$C^p(t + \delta t) = \sum_{m=0}^{K-1} B_m C(t - m\delta t) C^T(t - m\delta t) S(t - m\delta t) C(t - \delta t), \quad (32)$$

where $S(t)$ is the overlap matrix and $C(t)$ are the orbital expansion coefficients at time t . An approximate time reversible predictor–corrector method proposed by Kolafa [17], always stable predictor–corrector (ASPC), originally proposed for classical polarizable force fields, can be used. The extrapolation coefficients for this method are

$$B_k = (-1)^m (m+1) \frac{\binom{2K+2}{K-m}}{\binom{2K}{K}}. \quad (33)$$

The coefficients B_k ensure time reversibility to $O(\delta t^{K+3})$. The corrector associated with this method is

$$C(t) = \omega \arg \min_C E_{\text{KS}}(\mathbf{R}, C) + (1 - \omega)C^p(t), \quad (34)$$

with $\omega = \frac{K+1}{2K+1}$. Variations of the above two methods are possible and have been discussed together with, for example, a noise dissipation algorithm [18–20] for additional stability.

3 Faster, Larger, and More Accurate: Recent Developments

Advanced applications in AIMD require the treatment of systems with hundreds of atoms and an extended sampling of configurations. These simulations are only possible using large computational resources and the most advanced algorithms. Tapping the power of massively parallel computers was therefore an important goal in many AIMD code projects. Further important improvements in AIMD simulations came from better sampling algorithms and the emergence of improved electronic structure methods.

3.1 *Massively Parallel Implementation*

Modern high performance computer architectures are all based on massively parallel assemblies of multi-core CPUs with high speed networks. In order to take advantage of the available computing power it is necessary to revise and adapt algorithms constantly and to update computer codes. The AIMD community using plane wave based codes played a leading role in the efficient usage of parallel computers. These codes are dominated by a small number of computational kernels, e.g., three-dimensional fast Fourier transforms (FFT) and matrix multiplication, which can be mapped efficiently on distributed memory architectures. Early parallelization strategies focused either on a parallelization by bands (orbitals) or on the 3d-FFT [21, 22]. The FFT route proved to be more successful but mixed schemes were also explored in order to extend the range of optimal scaling. Efficient scaling to tens of thousands of processors was achieved with these implementations [23]. The recent trend to multi-core systems led to compute nodes with many compute elements. This caused an adaptation of algorithms using multi-level parallelization [24–26]. Typically a coarse-grain distributed memory level using the MPI library is used for inter-node parallelization and a fine-grain, loop level, parallelization using OpenMP takes advantage of the compute cores within the nodes. Several leading AIMD codes, e.g., CPMD [3, 22], Qbox [23], and Quantum ESPRESSO [26] are following this approach.

Implementations using massively parallel algorithms are no longer restricted to codes using plane wave basis sets. Atomic orbital based [27] or local basis sets based on regular grids [28] were successfully adapted to modern computers. These implementations are able to reduce scaling with system size and are therefore more efficient to simulate larger systems. However, this also leads to more complex algorithms and data structures which in turn make parallelization more difficult. This increased complexity is also responsible for the fact that these codes typically scale less efficiently to large numbers of processors.

Currently, reduced scaling methods using localized basis sets or the plane wave method are both capable of performing AIMD simulations for systems of thousand atoms at a rate of several picoseconds a week. The plane wave codes reach this performance by making use of their excellent scaling while the localized basis set codes profit from algorithmic efficiencies. In the future, plane wave codes will still be of importance for system sizes up to a thousand atoms, but for considerably larger systems the local basis set methods will dominate. For a further selection on reviews about high performance computing on vector systems we refer the interested reader to [29] and references therein.

3.2 *Basis Sets*

Simulations using the plane wave-pseudopotential framework were completely dominating the field of AIMD in the early years [3]. Plane waves are especially suited to be used with the Car–Parrinello method and also have other advantages, e.g., their orthogonality, control by a single parameter, or the absence of basis set superposition errors [30]. However, the original plane wave approach has an intrinsic cubic scaling with system size. For large system sizes, local basis set methods are more advantageous as they can lead to a reduced and ultimately linear scaling.

3.2.1 **Finite Difference and Discrete Variable Representation Methods**

Fattebert and Gygi proposed a real-space finite differences implementation for $O(N)$ density functional theory molecular dynamics simulations [31]. They showed that the discretization error can be reduced systematically by adapting the mesh spacing. Linear scalability was demonstrated with increasing system size using a localized orbital scheme. The authors were able to demonstrate energy-conserving AIMD with plane wave accuracy in $O(N)$ operations. Similar achievements were possible using spline function type basis sets within the CONQUEST code [32]. AIMD simulations at the complete basis set limit were demonstrated successfully by Lee and Tuckerman [33]. The authors used a discrete variable representation (DVR) approach. DVRs are local analytic functions defined at regular grid points. In contrast to finite difference methods, the DVR is a basis set approach, and

the kinetic energy can be evaluated exactly. DVRs and other grid based local basis functions try to combine the advantages of plane waves, e.g., the absence of basis set superposition errors and Pulay forces, with the screening abilities of local basis set methods. Another such approach was implemented in the ONETEP program [28]. Hine and coworkers showed that they are able to perform linear-scaling DFT with tens of thousands of atoms [34]. Their linear scaling approach was based on non-orthogonal generalized Wannier functions. The authors stressed the importance of linear scaling DFT developments for AIMD, as this is the only way to simulate system sizes in DFT that go beyond model systems.

3.2.2 Wavelets

Plane waves are highly suitable for electronic structure calculations of periodic and homogeneous systems, but are much less efficient for localized systems. For inhomogeneous or isolated systems other basis sets might be more appropriate. Still, it is for computational reasons desirable to work with orthogonal basis sets. Genovese et al. proposed Daubechies wavelets for DFT calculations, because they form a systematic orthogonal and smooth basis, localized both in real and Fourier spaces and they therefore meet both the requirements of precision and localization found in many applications [35]. The authors were able to show the principal features of an electronic structure pseudopotential method based on wavelets within the ABINIT code. The strength of the wavelet basis is reflected in the fact that the matrix elements, the kinetic energy, and nonlocal pseudopotential operators can be calculated analytically. The basis shows systematic convergence properties, very good computational performances, and an excellent efficiency for parallel calculations. For example, the gain in CPU time for large systems of several hundred atoms compared to a plane wave program was observed to be proportional to the significant reduction in the number of degrees of freedom [35].

3.2.3 Atomic Orbital Basis Methods

Gaussian type orbitals are the most popular basis in quantum chemistry calculations. For AIMD simulations, however, plane waves were used much more often. This was related to the difficulty in finding efficient algorithms for periodic systems based on atomic orbital basis sets. The Gaussian and plane waves (GPW) method [36] addressed this problem with a dual basis set approach. While the Kohn–Sham orbitals were expanded in Gaussians, plane waves were used for the electronic density. This allowed for a rapid calculation of the Hartree potential, similarly to plane wave codes, using fast Fourier transforms. VandeVondele et al. [27] published a general overview of the GPW method as part of the CP2K program. All computational aspects of the method are discussed. It was shown that the computational cost of calculating the Kohn–Sham matrix scaled linearly with system size with a very small pre-factor. Wavefunction optimization with the

orbital transformation technique resulted in good parallel performance outperforming traditional diagonalization methods. The efficiency of this method enabled the use of large Gaussian basis sets. Energy conserving Born–Oppenheimer dynamics were shown to be possible, and a highly efficient scheme was obtained using an extrapolation of the density matrix [27].

A comparison of plane waves and Gaussian basis sets was carried out in 2007 [30]. This was conducted in the framework of density functional theory for the hydrogen bond description with the water dimer as an example. Molecular dynamics simulations enforcing the self-dissociation reaction of the water dimer to study the influence of the basis set onto the reaction showed strongly varying results of the calculated forces for a chosen cutoff along the reaction coordinates. The basis set superposition errors of the dimer interaction energy was analyzed along the free-energy surface. Based on the analysis along the trajectories a qualitative and quantitative estimate depending on the particular point of the free-energy surface was provided.

In 2008, Artacho et al. presented developments and applicability of the Siesta method for a large variation of systems [37]. Within the Siesta code the plane wave basis for the electron density is combined with numerical atomic orbitals of finite support. In their article, Artacho et al. demonstrate linear scalability of the Siesta program using a system with more than 4,000 atoms [37].

Blum et al. suggested the application of numerically tabulated atom-centered orbitals (NAOs) in AIMD. These basis sets are implemented in the ab initio molecular simulations package FHI-aims. In benchmark calculations the authors showed an $O(N)$ scalability and a good parallelization [38].

3.3 *New Developments in Accuracy*

Real-world predictions do not just rely on a sufficiently large system size and fast calculations. Accuracy of the calculations plays an equally important role in applications. As most of the AIMD codes are carried out within the framework of density functional theory, the errors connected with this electronic structure method have to be reduced. For instance, frequencies calculated by Gaigeot et al. using the BLYP functional had to be down shifted by up to 100 cm^{-1} compared to frequencies calculated with hybrid functionals (e.g., B3LYP) or with wavefunction based ab initio calculations [39]. Gaigeot et al. stated that 5–10% underestimation of frequencies is typical for the BLYP functional. The amplitudes of methyl groups δ (C–H) bands were underestimated in their calculations which they attributed to C–H–water interactions being more sensitive to dispersion than to electrostatics forces. Therefore they estimated that this deficiency could be related to the lack of a proper dispersion term in DFT calculations. In 2008, Cohen et al. discussed the deficiencies of DFT in a short communication [40]. Approximations to the unknown exchange–correlation functional lead to major failures in DFT, e.g., underestimation of chemical reaction barriers and band gaps, errors in dissociation

energies of molecular ions as well as charge transfer excitation energies. Binding energies and the response to electric fields are often overestimated. Many of these errors are due to the self-interaction error [40]. This problem is partly cured by the inclusion of Hartree–Fock exchange in hybrid functionals. Another apparent deficiency is due to static correlation which is difficult to describe within a single determinant Kohn–Sham approach [40]. Finally, it is not possible to capture an intrinsically non-local interaction like dispersion using only local correlation functionals.

In order to improve on accuracy, new functionals had to be developed that are, on one hand, more exact but that are, on the other hand, still feasible for large scale periodic computations. Especially challenging was the inclusion of exact-exchange needed for hybrid functionals while more pragmatic solutions to the dispersion problem were found.

3.3.1 New Functionals

In 2003, Heyd et al. proposed to use a screened Coulomb operator in the Hartree–Fock exchange part of hybrid functionals [41]. The Coulomb operator was split into short-range (SR) and long-range (LR) components,

$$\frac{1}{r} = \underbrace{\frac{\text{erfc}(\omega r)}{r}}_{\text{SR}} + \underbrace{\frac{\text{erf}(\omega r)}{r}}_{\text{LR}}, \quad (35)$$

with $\text{erfc}(\omega r) = 1 - \text{erf}(\omega r)$ and ω being an adjustable parameter. In the next step a new hybrid functional was proposed which performs the exact exchange mixing only for short-range interactions in both HF and DFT. For the density functional part the authors developed a special screened operator version of the PBE functional. The performance of this functional was very promising, i.e., for molecular systems it yielded an accuracy comparable to the best established hybrid methods, such as B3LYP and PBE0. For periodic boundary condition-calculations a performance similar to established functionals were found while, at the same time, significant reductions in the computational costs were observed [41].

A variety of databases for testing and designing new density functionals together with a suite of new functionals were developed by Zhao and Truhlar in 2008 [42]. The new density functionals were termed M06-class (and, earlier, M05-class) functionals, for which the authors enforced some fundamental exact constraints such as the uniform-electron-gas limit and the absence of self-interaction energy. The M06-suite functionals depended on spin densities, on spin density gradients, on spin kinetic energy densities, and, for nonlocal (hybrid) functionals, on Hartree–Fock exchange. This led to the design of four new functionals:

1. The hybrid meta functional M06 was found to perform well for transition metals, main group thermochemistry, medium-range correlation energy, and barrier heights.
2. Another hybrid meta functional, M06-2X, failed for transition metals but showed excellent performance for main group chemistry. It predicted accurate valence and Rydberg electronic excitation energies, and it was an excellent functional for aromatic–aromatic stacking interactions.
3. M06-L was not as accurate as M06 for barrier heights but was the most accurate functional for transition metals and was the only local functional (no Hartree–Fock exchange) with better average performance than B3LYP.
4. M06-HF showed good performance for valence, Rydberg, and charge transfer excited states with minimal sacrifice of ground-state accuracy.

In 2008, Spencer and Alavi [43] introduced a scheme for calculating the exact exchange energy in periodic solids within a Kohn–Sham or Hartree–Fock approach in such a way that the need to treat the integrable singularities via an auxiliary function was removed. In the exchange integrals, the authors used a modified Coulomb potential, which tended to the exact potential as the number of k points increased. It also showed no singularities and was very simple to implement. A comparison of this approach to the auxiliary function scheme for diamond, graphite, and two allotropes of silicon carbide showed a rapid convergence with the number of wave vectors [43].

Guidon et al. [44] used the truncated Coulomb operator, as presented in [43], for a robust and accurate scheme for large condensed phase systems. This scheme was based on previous work [45], where the authors suggested a linear scaling implementation of an exact exchange functional using periodic boundary conditions. Atom-centered Gaussian basis functions were employed, allowing for efficient screening with the density matrix. The new scheme showed increased stability if applied with large and flexible basis sets.

In 2010, Paier et al. showed the applicability of the random phase approximation (RPA) to ground-state correlation [46]. RPA correlation has the advantage that it is compatible with exact Hartree–Fock-type exchange and that it describes van der Waals interactions exceptionally well. The inclusion of second-order screened exchange rectifies problems with one-electron self-interaction errors, which lead to disturbingly large correlation energies in stretched bond situations. A large number of molecular benchmark results obtained using full-range as well as long-range corrected hybrids incorporating second-order screened exchange correlation were presented [46].

3.3.2 Correcting for van der Waals Interaction

To treat the dispersion problem of density functional theory (i.e., the failure to describe van der Waals interaction accurately) is now an active field of research. It is especially important for AIMD to treat these kinds of interactions more

accurately, because in large and condensed phase systems the chemistry often relies solely on the intermolecular forces of such type.

Several articles on corrections of van der Waals interactions applying the seamless approach in density functional theory have appeared in the literature [47–50]. In an early article of a series, Dion et al. developed and applied a van der Waals density functional in order to treat situations for which nonlocal, long-ranged interactions, such as van der Waals (vdW) forces, were influential. The authors suggested the following form:

$$E_c^{\text{nl}} = \frac{1}{2} \int d^3r d^3r' n(\mathbf{r}) \phi(\mathbf{r}, \mathbf{r}') n(\mathbf{r}') \quad (36)$$

for the nonlocal correlation energy part in which $\phi(\mathbf{r}, \mathbf{r}')$ is some given, general function depending on $\mathbf{r} - \mathbf{r}'$ and the densities n in the vicinity of $\mathbf{r}$ and $\mathbf{r}'$. This truly non-linear functional has been applied successfully to many systems and recently a second generation further improved version has been proposed [51].

Röthlisberger and coworkers proposed to add an effective atom centered nonlocal term to the exchange-correlation potential in order to cure the lack of London dispersion forces in standard density functional theory [52, 53]. In particular, the authors constructed an effective potential consisting of optimized nonlocal terms dependent on higher angular momentum for all atoms in the system. They modeled van der Waals forces by an atom-electron interaction, mediated by appropriate nonlocal effective core potential (ECP) projectors, which were obtained from an optimization scheme. Röthlisberger and coworkers stated that this scheme has some advantages over empirical pair potential corrections: “... First, the improved electronic properties (dipole moment, quadrupole moment, and polarizability) indicate that, due to the non-locality of the ECP projectors, the valence wavefunctions reproduce more of the characteristics of dispersion interactions than a simple additive atom-atom based correction. Second, properly calibrated and transferable atomic dispersion calibrated ECPs no longer need any artificial a priori assignment of interacting groups or atoms” [52].

Among the approaches discussed here, the most simple and straightforward and thus the most practical approach is that followed by Stefan Grimme [54, 55]. Grimme defined the dispersion corrected total energy $E_{\text{MF-D}}$ as

$$E_{\text{MF-D}} = E_{\text{MF}} | E_{\text{disp}}, \quad (37)$$

where E_{MF} is the Hartree–Fock or DFT mean-field energy and E_{disp} is an empirical dispersion correction expressed as

$$E_{\text{disp}} = -s_6 \sum_{i=1}^{N_{\text{at}}-1} \sum_{j=i+1}^{N_{\text{at}}} \frac{C_6^{ij}}{R_{ij}^6} f_{\text{dmp}} R(R_{ij}). \quad (38)$$

With N_{at} being the number of atoms in the system, C_6^{ij} denotes the dispersion coefficient for an atom pair ij , s_6 is a global scaling factor that only depends on the DFT functional used, and R_{ij} is an interatomic distance. To avoid near-singularities the damping function f_{dmp} was added:

$$f_{\text{dmp}}(R) = \frac{1}{1 + e^{-\alpha(R/R_0 - 1)}}. \quad (39)$$

Here R_0 is the sum of van der Waals radii and α is a scaling factor [54]. The C_6 coefficients were partly taken from the literature [54], but also newly averaged over possible hybridization states of the individual atoms. Mixing rules of the following kind:

$$C_6^{ij} = \frac{2 \cdot C_6^i \cdot C_6^j}{C_6^i + C_6^j} \quad (40)$$

were applied. This approach was termed DFT-D2 by Grimme [56].

As the fragment densities of hydrogen bonded systems significantly overlap, these kinds of interactions are well described by standard DFT. However, if errors of 10–30% need to be corrected, Grimme recommended his dispersion correction scheme. As a consequence, Grimme mentioned the fact that steep damping functions need to be applied in order to retain the original DFT description as closely as possible in hydrogen bonded systems [54]. An improvement of the original approach followed in 2006 [55], where Grimme stated that the following shortcomings were addressed:

1. Consistent atomic parameters (C_6 coefficients) were only available for elements H, C-Ne, but studies of supramolecular structures or problems in material science require parameters for elements from the whole periodic table.
2. Test calculations for molecules with third-row elements showed systematic errors.
3. Adding the dispersion energy to the KS-DFT energy led to inconsistencies for “normal” thermochemistry, e.g., atomization energies: the dispersion correction is zero for the free atom and always nonzero (and large) for the molecule.

In order to account for these problems Grimme reduced the scaling factor from 1.22 to 1.10, which improved computed intermolecular distances for systems with heavier atoms [55]. Smaller values of α from the damping functions were chosen which provided larger corrections at intermediate distances and at negligible dispersion energies for typical covalent bonding situations. Furthermore, Grimme applied a new combination rule:

$$C_6^{ij} = \sqrt{C_6^i \cdot C_6^j} \quad (41)$$

for mixed atom situations, which yielded much better results but required a new fitting of s_6 . Altogether, Grimme obtained much improved results and the main reason for the higher accuracy was that the short range part of the density functional was adjusted to the presence of the long-range correction and double-counting effects could be avoided. In 2010, Grimme and coworkers [56] suggested a consistent and accurate ab initio parameterization of the DFT-D approach for the 94 elements H-Pu termed DFT-D3. The main ingredients of the new approach were atom-pairwise specific dispersion coefficients and cutoff radii which were both calculated from first principles. Three-body terms were also considered [56].

3.3.3 Explicit Relativistic Description

In 2009, the scalar-relativistic Douglas–Kroll–Hess method was combined with Born–Oppenheimer molecular dynamics simulations [57]. Using relativistic densities in a nonrelativistic gradient routine was found to be a valid approximation of relativistic gradients. An excellent agreement between optimized structures and geometries obtained from numerical gradients was observed with an error smaller than 0.02 pm. Hydrogen halide dimers (HX)₂ with X=F, Cl, Br, and I served as small test systems for AIMD simulations [57]. Relativistic effects were observed. In particular, the amplitude of motion was larger, the frequency of motion was smaller, and the distances were larger in the relativistic picture. Several localization schemes were evaluated for different interatomic and intermolecular distances. The errors of these localization schemes were small for geometries which were similar to the equilibrium structure. They became larger for smaller distances, introducing a slight bias towards closed packed configurations.

3.4 New Integration Schemes

Using a time reversible BO scheme as discussed in Sect. 2.5 allows for efficient and stable dynamics. However, a sufficient accuracy of the nuclear forces is still needed. In systems with slow convergence this will require many self-consistency cycles and therefore it will slow down the simulation. For such systems, an algorithm that would allow for less stringent convergence criteria on the forces could bring considerable gains in efficiency.

Based on their previous work on molecular dynamics using noisy forces, Kühne et al. [58] demonstrated a novel AIMD method that combines Langevin dynamics for the nuclei with time reversible dynamics for the electronic degrees of freedom and incomplete SCF convergence. The method is based on the observations [59] that the error of the nuclear forces has a Gaussian distribution, that the autocorrelation function of the force errors decays rapidly, and that the force errors show a random distribution with respect to the velocities. From these observations it was

deduced that the canonical distribution can be sampled using a Langevin-type equation:

$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_{\text{inc}} - \gamma \dot{\mathbf{R}}_I + [\Xi_I + \Xi_I^{\text{inc}}], \quad (42)$$

where γ is a Langevin friction coefficient, Ξ_I a random noise term, and $\mathbf{F}_{\text{BO}} = \mathbf{F}_{\text{inc}} + \Xi_I^{\text{inc}}$. The force from the incomplete SCF optimization is denoted here by $\mathbf{F}_{\text{inc}}$ and the corresponding fully converged force by $\mathbf{F}_{\text{BO}}$. Sampling of the Boltzmann distribution requires that the fluctuation dissipation theorem is obeyed:

$$\langle \Xi_I(0) \Xi_I(t) \rangle = 6\gamma M_I k_B T \delta(t). \quad (43)$$

In applications, the friction term is split into two contributions $\gamma = \gamma_D + \gamma_L$, where γ_L is taken to be an arbitrary constant and γ_D is determined by requiring that the correct temperature $T = \frac{1}{3k_B} \langle M_I \dot{\mathbf{R}}_I^2 \rangle$ is generated.

If the parameter γ_D can be kept small, the Langevin method not only generates the correct canonical distribution but also dynamical properties are accurately described [60]. In order to achieve this goal, the error of the forces can be consistently reduced by noting that the last self-consistent cycle can be interpreted as a Harris functional. The missing force $\mathbf{F}_{\text{BO}} - \mathbf{F}_{\text{inc}}$ can then be approximated to a high degree by

$$- \int d\mathbf{r} \left\{ \left[\left(\frac{\partial V_{\text{XC}}[\rho_{\text{in}}]}{\partial \rho_{\text{in}}} \right) \Delta \rho + V_{\text{H}}[\Delta \rho] \right] \nabla_I \rho_{\text{in}} \right\}, \quad (44)$$

where $\Delta \rho$ is the difference between input ρ_{in} and output density, and V_{XC} and V_{H} are the exchange-correlation and Hartree potential, respectively.

Using the Langevin method, Kühne et al. were able to accelerated their benchmark calculations on liquid SiO_2 by one to two orders of magnitude. In 2009, Kühne et al. simulated liquid water with the new method [60]. Oxygen–oxygen radial distribution functions agreed well with other approaches. Because of the acceleration they were able to estimate for the first time reliably the diffusion coefficient and shear viscosity of liquid water.

3.5 Enhanced Sampling

3.5.1 Elastic Band and String Methods

Of special importance to AIMD are the enhanced sampling methods termed elastic band and string methods. Michaelides and coworkers explained in their excellent free-energy method assessment article that special considerations have to be taken when rare event simulations are done within AIMD. As the iterative SCF procedure has an impact on the calculation of the forces in contrast to the exact forces in empirical simulations, the convergence tolerance of the SCF procedure has to be chosen carefully, especially for reactive situations [61].

A very excellent introduction to string methods is found in [62] in which the authors introduced a combined approach of the string method with a sampling technique to determine minimum free energy paths (MFEP). The string method was originally designed to determine reaction pathways without making any a priori assumption about the mechanism [63]. With the string method, an evolution equation for a parameterized curve is constructed, coined a string, which is such that any initial guess for the string converges to a solution as time evolves. The solution has the condition that the force must be everywhere tangent to the minimum energy path (MEP). Maragliano et al. stated [62] that the string method is very robust and efficient at determining minimum MFEP in a given landscape, requiring as input only the calculation of the mean force. The cost of the string method calculations performed locally scaled linearly with the number of points along the discretized string. However, it was independent of the dimension which was in marked contrast with free energy mapping techniques, for which the cost increases exponentially with the number of reaction coordinates used to describe the reaction. In order to eliminate potential difficulties for the string method when being applied in the original state space of the system, the authors removed fast degrees of freedom. They found that working in free energy space was easier as well, because the free energy landscape is in general much smoother than the original potential energy landscape of the system [62]. The applicability of the new technique was demonstrated with the example of the dipeptide alanine isomerization transition. Most importantly, the example showed that the transition mechanism can be described by using the four dihedral angles; however, it was not captured using only two of them [62].

Michaelides and coworkers explained the nudged elastic band (NEB) method [64] as taking a set of system images between the initial and the final states and optimizing them simultaneously in a subspace perpendicular to the imaginary line connecting the images [61]. Additional forces keep the images evenly distributed along the pathway between the initial and final states. According to Michaelides the location of the transition state with the NEB method can be achieved using the climbing image procedure, where the highest energy image is moved with the aid of the modified real force. The modified real force is obtained by flipping the component of the force parallel to some direction [61]. In their study the authors evaluated the performance of a number of methods for locating transition states with DFT. They also introduced three new algorithms. A major conclusion from their work was that the NEB method is relatively fast, especially when just a single (climbing) image is used [61].

3.5.2 Metadynamics

Metadynamics is a method based on MD simulations that allows thorough sampling of a predefined multidimensional configurational space and provides, at the same time, the direct reconstruction of the explored free energy surface (FES) [65–71].

The metadynamics method was introduced in 2002 by Laio and Parrinello as an elegant extension of adaptive bias potential methods [65]. The authors used a coarse-grained non-Markovian dynamics in the space defined by a few collective coordinates s_i . With the aid of a history-dependent potential term the minima of the FES were filled in time, allowing the efficient exploration and accurate determination of the FES as a function of the collective coordinates. Laio and Parrinello demonstrated the applicability of this approach in the case of the dissociation of a sodium chloride molecule in water and in the study of the conformational changes of a dialanine in solution [65].

With the metadynamics approach, the free energy surface $\mathcal{F}(\mathbf{s})$ of a limited set of collective variables s_i can be explored [68]. This is done by introducing an extended Lagrangian with fictitious particles s_α for each CV:

$$\mathcal{L} = \mathcal{L}_{\text{CPMD/BOMD}} + \sum_{\alpha} \frac{1}{2} \mu_{\alpha} \dot{s}_{\alpha}^2 - \sum_{\alpha} \frac{1}{2} k_{\alpha} (S_{\alpha}(\mathbf{r}) - s_{\alpha})^2 - V(t, \mathbf{s}). \quad (45)$$

While the first term on the right-hand side of (45) is the usual AIMD Lagrangian, the second term is the total kinetic energy of the fictitious particles. For large enough masses μ_{α} they are adiabatically separated from the ionic and electronic degrees of freedom. Each fictitious particle s_{α} is connected to its actual collective variable $S(\mathbf{r})$ by a harmonic potential [68]. The history-dependent biasing potential [last term in (45)] is introduced in order to enhance sampling. This biasing potential $V(t, \mathbf{s})$ constitutes a sum of repulsive Gaussian-shaped potential hills:

$$V(t, \mathbf{s}) = \sum_{t_i < t} H \exp \left[-\frac{(\mathbf{s} - \mathbf{s}^i)^2}{2(\Delta W^{\perp})^2} \right] \exp \left[\frac{((\mathbf{s}^{i+1} - \mathbf{s}^i)(\mathbf{s} - \mathbf{s}))^2}{2(\Delta W_i^{\parallel})^4} \right] \quad (46)$$

with $\mathbf{s}^i = \{s_{\alpha}(t_i)\}$ and H the height. $\Delta W_i^{\parallel} = \mathbf{s}^{i+1} - \mathbf{s}^i$ gives the width along the direction of motion and $\Delta W^{\perp}$ the size in the orthogonal direction. In the limit of a long simulation time, the following equation holds [66]:

$$\lim_{t \rightarrow \infty} V(t, \mathbf{s}) = -F(\mathbf{s}). \quad (47)$$

Laio and Parrinello stated [65] that constructing dynamics on an FES that depends on a few collective coordinates allowed one to simplify the complexity of the problem, which depends exponentially on the number of degrees of freedom. The FES will be smoother than the underlying PES and topologically simpler, with a greatly reduced number of local minima [65]. A history-dependent bias potential as defined in (46), but applied in a regular MD simulation without applying the collective variable space, is efficient in finding escapes from the local minima but will not provide quantitative information about the FES [65].

3.6 Properties: IR, NMR and EXAFS

3.6.1 IR Spectra

Due to the costs associated with the electronic structure calculations, AIMD simulations always suffer from short simulation times; see also previous Sect. 3. In 2005, Ifimie and Tuckerman devised a method that allows well-converged results for IR spectra from small AIMD systems and short trajectories [72]. The frequency-(ν)-dependent Beer–Lambert absorptivity coefficient $\alpha(\nu)$ is given as

$$\alpha(\nu) = \frac{\pi\nu[1 - \exp(-\beta\hbar 2\pi\nu)]}{3\hbar V c n(\nu)\epsilon_0} \int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \times \langle \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(\tau) \rangle_{\text{qm}} d\tau \quad (48)$$

with β being $1/(k_b T)$, V the sample's volume, c the speed of light, $n(\nu)$ the index of refraction, ϵ_0 the vacuum permittivity, and $\hat{\mathbf{M}}$ the quantum mechanical total dipole moment operator. Ifimie and Tuckerman applied the harmonic approximation expressed in

$$\int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \times \langle \hat{\mathbf{M}}(0)\hat{\mathbf{M}}(\tau) \rangle_{\text{qm}} d\tau = \frac{\beta\hbar 2\pi\nu}{1 - \exp(-\beta\hbar 2\pi\nu)} \int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \times \langle \mathbf{M}(0)\mathbf{M}(\tau) \rangle_{\text{cl}} d\tau. \quad (49)$$

In the last line “cl” denotes a classical ensemble average, i.e., phase space integration. Next, the authors suggested the application of integration by parts:

$$\frac{2\pi\nu}{c} \epsilon'' = \alpha(\nu)n(\nu) = \frac{1}{6cV\epsilon_0 k_B T} \int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \langle \dot{\mathbf{M}}(0)\dot{\mathbf{M}}(\tau) \rangle_{\text{cl}} d\tau. \quad (50)$$

Applying this expression with the four-term Blackman windowed Fourier transform approach led to sufficient accuracy based on a relatively short trajectory (10 ps), i.e., the authors found excellent agreement between the experimentally obtained spectra for liquid water and ice. Using this approach and decomposing the total dipole moment

$$\mathbf{M} = \sum_A \mu_A \quad (51)$$

the contribution of a molecule A was calculated via cross-correlation:

$$\epsilon'' = \frac{1}{12\pi V\epsilon_0 k_B T} \int_{-\infty}^{\infty} \exp(2\pi i\nu\tau) \langle \dot{\mathbf{M}}(0)\dot{\mu}_A(\tau) \rangle_{\text{cl}} d\tau. \quad (52)$$

Iftimie and Tuckerman demonstrated that the absolute spectrum of an excess proton in water can be accurately obtained by subtracting the spectrum of bulk water from that of an aqueous solution of HCl.

Similarly, Gaigeot and coworkers analyzed the IR spectrum of *N*-methylacetamide (NMA) in gas phase and aqueous solution [39]. Two approaches were tested. The first is based on (49) with individual dipole moments of molecules and they applied the derivative similar to (50). However, the derivative of dipole moment $\mathbf{j}$ is obtained via the following expression:

$$j^\beta(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial q_i^\alpha}(t) \frac{dq_i^\alpha}{dt}(t) = \sum_{i,\alpha} \frac{\partial M^\beta}{\partial q_i^\alpha}(t) v_i^\alpha(t) \quad (53)$$

with q_i being the position of atom i and $\partial M^\beta / \partial q_i^\alpha(\alpha, \beta = x, y, z)$ representing the components of the atomic polar tensor of atom i . Gaigeot and coworkers found that, despite the very short time span of 1 ps, the IR spectrum from the current–current autocorrelation function gave most of the important features of the absorption. That is, all amide bands were present. In contrast, the most intense amide I–amide II band was not correctly reproduced from the same 1 ps time interval when the dipole–dipole correlation function was used. The authors explain that “This improved convergence is most likely an effect of the favorable statistics of velocities. Atomic velocities, in contrast to dipoles, are isotropic and fluctuate very quickly during the dynamics. Therefore, calculations of infrared spectra through current–current correlation functions can be done on shorter timescales of dynamics. This might be particularly important in the case of strong coupling between almost degenerate modes, such as for example the δ (O–H) bending mode of water and the amide I and amide II bands of NMA which both occupy the same $\approx 1,600 \text{ cm}^{-1}$ frequency band.” [39].

Furthermore, it was pointed out by Gaigeot and coworkers that thermalization of all degrees of freedom might be difficult to achieve and can therefore induce errors in calculated infrared intensities. To compare the calculated infrared spectra to experiments (gas and liquid phase), use of two different scaling factors that adjusted the position of the calculated amide III band were made, 1.064 and 1.117 for the gas phase and the solution, respectively [39]. Gaigeot and coworkers state that there is no reason why the scaling factor of gas phase and solution should be the same. The scaling factor depends on frequency and thus might change in a condensed phase environment. Another difference in solution could be an enhanced inertia (giving rise to frequency red-shifts) due to the fictitious electron mass used in the Car–Parrinello molecular dynamics scheme. As also shown by Iftimie and Tuckerman, the fictitious electron mass can contribute to the underestimation of the frequencies, up to $40\text{--}50 \text{ cm}^{-1}$ [72]. If the well-known frequency red-shifts due to the use of the BLYP functional are kept in mind, this leads to an increased underestimation of the frequency positions [39].

3.6.2 NMR and ESR/EPR

The calculation of NMR parameter has been studied extensively; see [3, 73] for general overviews. In 2001, Sebastiani and Parrinello implemented the NMR chemical shift calculation in the plane wave AIMD code CPMD [74]. From this implementation it was possible to treat extended systems within periodic boundary conditions, i.e., the method was applicable to crystalline and amorphous insulators as well as to liquids. The problem of the position operator was solved by the use of maximally localized Wannier functions. Several benchmark calculations showed good agreement with experimental values.

A linear scaling, tested with up to 3,000 basis functions, was implemented in Q-Chem by Ochsenfeld et al. in 2004 [75]. The calculations were dependent on a Hartree–Fock formalism and test calculations with more than 1,000 atoms made.

In 2009, the calculation of the NMR chemical shifts and EPR g tensors was extended to the Gaussian and plane wave code CP2k [76]. Weber et al. performed several test calculations with good agreement with experimental results. Additionally, the NMR shifts in isolated as well as hydrated adenine were calculated.

3.6.3 EXAFS

Near-edge X-ray absorption spectra calculations at the DFT level were also carried out in the framework of AIMD [77–81]. Several test calculations have been carried out: water and CO with different basis sets and core-hole potentials, the C, O, and N K-edges in $(\text{CH}_3)_2\text{CO}$, CH_3COH , and $\text{C}_5\text{H}_5\text{N}$, as well as water and CH_3OH dimers for the sensitivity to weak intermolecular interactions. For the basis set dependence the 6-31G^{**}, 6-311G^{**}, 6-311++G(2d,2p), 6-311++G(3fd,2dp), Iglö-III, Roos-ADZ-ANO, Roos-ATZ-ANO, aug-cc-pVDZ, aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z basis sets were compared, and it was observed that the EXAFS spectra significantly varied with the basis set in number of signals, signal position, as well as signal shape. Even with the largest basis set the experimental O K-edge in water was only marginally described by the BLYP exchange functional. The same was found for CO. For the dependence on the core-hole potential, a comparison for H_2 and CO molecules with the aug-cc-pV5Z basis set and the BLYP functional were made. Using full core-hole potentials, the entire spectrum was shifted by several eV to higher energies and, similar to the basis set choice, the choice of the functional largely influenced the spectrum. Despite these deficiencies, EXAFS calculations of $(\text{CH}_3)_2\text{CO}$, CH_3COH , and $\text{C}_5\text{H}_5\text{N}$ showed a resemblance between theoretical and experimental spectra for the different atoms, and therefore an alignment depending on these calculations was possible [80]. Weaker interactions were investigated at water–water and methanol–methanol dimers. In both calculations the weak hydrogen bonds significantly changed the spectra for the acceptor and the donor in accordance with chemical intuition and experiment, allowing for an assignment of the experimental results to different coordinations

and clusters. In the computed EXAFS spectrum a systematic error with respect to the experimental spectrum was obtained. In a subsequent study from 2008 the different dependencies of the calculated EXAFS spectra were studied for liquid water and hexagonal ice within the supercell approach [81]. Several configurations of AIMD simulations were produced and asymmetrically coordinated water molecules were observed. For example, such water molecules with only one hydrogen bond showed well defined spectral lines which significantly differed from the ice signals.

For a review of calculations of the X-ray adsorption spectra (XAS) which especially focused on the transition potential approach and its application to water, see the recent work of Leetmaa et al. [82].

4 Applications in Chemical Engineering

In this section we discuss several studies in which AIMD was applied to special chemical problems, reactions, and industrial processes.

4.1 Wavefunction Analysis

Many schemes were adapted to analyze the wavefunction (electronic structure) in AIMD simulations. The most important ones are the Wannier analysis based on maximally localized Wannier functions (MLWF) [83], the electron localization function (ELF)[84], the Fukui function [85], and the nucleus-independent chemical shift maps [74].

The usefulness of Wannier functions was demonstrated by Silvestrelli et al. [86] in a study of amorphous silicon. The authors were able to describe the bond structure using the Wannier functions. The investigation of defect configurations was possible with a novel degree of accuracy through the Wannier approach [86]. Another application of Wannier functions was published by Fitzhenry et al. investigating silicon–carbon alloys [87]. In this study the bond structure was resolved by the application of Wannier functions and Fitzhenry et al. were able to identify, classify, and quantify the types of bonding present in the alloy. They were able to observe three-center bonding and a temperature dependent flipping of bonds during the simulation [87]. In 2005, Bühl et al. investigated the ionic liquid (see also Sect. 4.2.2) 1,3-dimethylimidazolium chloride ([Mmim][Cl]) at 438 K using CPMD [88]. Population analyses showed noticeable charge transfer from anions to cations and Wannier functions demonstrated this specifically for the CH ... Cl hydrogen bonds. Another important tool of the Wannier analysis is the derivation of local dipole moments. The applications of dipole moment calculations is discussed in Sect. 4.2.1.

The electron localization function (ELF) was applied to investigate a system of 30 AlCl_3 molecules with one $[\text{Emim}][\text{Cl}]$ ion pair [89]. It was found that, due to the decrease in electron deficiencies, large anionic clusters formed.

Aromaticity and homoaromaticity of a parent barbaralane and a tetraphosphabarbaralane of C_{2v} -symmetry were visualized by means of three-dimensional nucleus-independent chemical shift maps [90]. In combination with CPMD simulations the fluxional character of tetraphosphabarbaralane was revealed and it was shown that the ionic motion at room temperature leaves the aromaticity in this case unchanged [90].

4.2 *Properties of the Vapor Phase, Liquids, Mixtures, and Solvent Effects*

AIMD is well suited for describing several properties of the vapor phase, liquids, mixtures, and solvent effects. Solvent effects are especially very well described by AIMD if the molecules actively solvate the solutes, because the electronic structure is explicitly described by AIMD and changes according to the solvent-solute interaction will be well captured.

4.2.1 From Gas Phase to Liquid Phase

Differences between gas phase molecules and molecules in condensed phases have been summarized previously [91]. Chemical reactivity can be highly influenced by the chemical environment and, therefore, chemical reactivity of an isolated molecule in vacuum is not always a good model for a molecule surrounded by other “active” or solvent molecules. A first step to study solvent effects is to consider the dipole moment of molecules in gas phase as well as in condensed phase.

The dipole moment of liquid *water* was investigated by several authors [92–94]. Silvestrelli and Parrinello calculated dipole moments of a single water molecule (1.87 D), a dimer (2.1 D), a trimer (2.4 D), as well as liquid water (2.95 D) [92]. In a subsequent study with refined methods they obtained a dipole moment of 3.0 D for liquid water from AIMD simulations [93]. In 2004, Kuo and Mundy reported a study of the aqueous liquid–vapor interface where water was simulated in such a fashion that in one simulations box the water molecules moved freely from the dense bulk phase into the low density vapor phase, i.e., the number of molecules surrounding a water molecule changed smoothly [94]. In this study, Kuo and Mundy found a molecular dipole moment at the vapor/liquid interphase of approximately 2.4 D which changed smoothly to a value of 3.0 D in the bulk phase.

Together with other water properties, the temperature change of the water dipole moment was investigated by McGrath et al. in 2006 [95]. The authors observed

a dipole moment of gas phase water of 1.8 D at 323 K and 2.1 D at 523 K, while in the liquid phase the dipole moments changed to 3.0 D at 323 K and 2.5 D at 523 K. This demonstrates not only the dependence on the chemical environment but also on the temperature.

Besides water, *methanol* was investigated with respect to its changing dipole moment [96]. Handgraaf et al. found – despite little alterations in the Wannier center positions – a dipole moment increase of methanol from 1.73 D in the gas phase for a single molecule to 2.54 D in the liquid phase.

N-Methylacetamide was investigated by Whitfield et al. in 2006 [97]. For the gas phase molecules a dipole moment of 3.74 D was found and in the liquid phase the dipole moments had a value of approximately 6 D. AIMD simulations also show for this liquid a broad distribution of molecular dipole moments. The average AIMD value is considerably higher than the dipole moment of 4 D that is used in classical force field simulations of this liquid.

In associating liquids the molecular dipole moments increase by 40–60% compared to the isolated molecule. These solvents will therefore strongly affect the chemical reactivity of solute molecules. Classical force field simulations neglecting polarization will not be able to capture these changes.

4.2.2 Liquids: Water, Ionic Liquids, and Others

Water serves as an ideal test system for different calculations, because a wide range of experimental as well as theoretical data are available [98–107].

One of the first water AIMD study was undertaken by Laasonen et al. in 1993 [98]. The authors applied a gradient corrected exchange functional in order to capture accurately the hydrogen bonding in the liquid. The simulation results were in good agreement with available experimental data.

Three gradient-corrected density functionals – B, BP, and BLYP – in liquid water simulations were tested by Sprik et al. in 1996 [100]. The authors observed from the structural and dynamical properties that hydrogen bonding was too weak with the Becke (B) functional, while hydrogen bonding was too strong if the BP functional was applied. The BLYP functional provided the best agreement with experiment.

Another functional assessment was carried out by VandeVondele et al. in 2005 [104]. The influence of the temperature was investigated within the different functionals (BLYP, PBE, TPSS, OLYP, HCTH120, and HCTH407). The BLYP, PBE, and TPSS functionals gave similar results, while OLYP, HCTH120, and HCTH407 showed a more diffusive dynamics and a lower structuring of the liquid. The BLYP and PBE functionals were again compared in a study by Schmidt et al. in 2009 [106].

Ionic liquids are liquids at or near room temperature which are composed entirely of ions [108]. Their special properties enable a wide range of application and many theoretical [109, 110] as well as experimental [108, 111, 112] investigations

have appeared in the last few years in the literature. AIMD simulations were performed as well to investigate their extraordinary behavior [88, 110, 113–118].

In 2005, AIMD simulations on dimethylimidazolium chloride [Mmim][Cl] carried out by Del Pópolo et al. showed significant differences compared to both the classical simulations and the neutron diffraction results [113]. In particular, and unlike the gas-phase ion pair, chloride ions tended to be located near a ring C–H proton in a position suggesting hydrogen bonding. It should be noted, that these results might be due to the choice of the applied functional. No dispersion correction was used and later it was shown that GGA functionals favor the hydrogen bonded states [119, 120].

Bhargava and Balasubramanian found that the apparently good agreement between the pair correlation functions from classical MD and AIMD conceal subtle but crucial, differences [114]. The radial pair distribution functions between the most acidic proton of the [Mmim] cation and the chloride anion were extremely different regarding location and width of the peaks. Furthermore, differences between AIMD and MD in the spatial distribution of chloride ions around the cation were found. The data were explained in terms of the formation of a hydrogen bond between the most acidic hydrogen of the imidazolium ring and the chloride ion. Size effects were excluded by simulations of 32 ion pairs with traditional MD simulations. The cation–anion hydrogen bond present in the melt was observed as a red-shift in the C–H stretching frequency.

The structural and dynamical hydrogen bonding in the IL [Emim][SCN] was investigated by Thar et al. in 2009 [117]. The geometric picture indicated a superior role for the most acidic hydrogen bond as compared to the two other hydrogen atoms at the rear. Despite the structural picture, the hydrogen bond dynamics at the most acidic hydrogen atom was found to decay faster than the corresponding dynamics at the other ring hydrogen atoms. Neglecting the directionality in the hydrogen bond analysis provided dynamics which reflected the geometrical analysis. Two movements were identified. First, a fast (<0.3 ps) hopping of the anion above and below the imidazolium ring and, second, a translational motion of the anion away from the cation in-plane of the imidazolium ring (5–10 ps).

The first AIMD simulation of an IL applying dispersion-corrected functionals [52] was carried out on the protic ionic liquid monomethylammonium nitrate [118]. On average, 1.8 of 3 possible hydrogen bond contacts formed, leaving one free acceptor and donor site similar to water. Furthermore, like water, monomethylammonium nitrate exhibited a fast fluctuating hydrogen bond network. However, the hydrogen bond network of monomethylammonium nitrate and water also showed some important structural differences. While the hydrogen bonds in water arrange in parallel fashion, the hydrogen bonds of monomethylammonium nitrate prefer angles of 0° , 90° , and 180° . The ion dynamics of monomethylammonium nitrate was described by a model of ions rattling in long living cages [118].

Other liquids, like liquid ammonia NH_3 [121], formamide HCONH_2 [122], and liquid hydrogen fluoride HF [123], as well as more exotic liquids, like liquid

deuterium D₂ [124, 125], melted carbon, graphite or diamond C [126–130], melted aluminum chloride AlCl₃ [131], and liquid phosphorus P [132], were examined.

Liquid metals and alloys were investigated as well from AIMD because of their importance in physics, chemistry, industry, electronics, and engineering. These studies contain Li [133], Na [134], Al-Si alloy [135], Si [136, 137], K-Pb [138], Fe-Si [139], Ni [140], Cu [141], GaAs [142], Ge [143–145], As₂Se₃ [146], Se [147], Zr [148], CdTe [142], CsPb [149], and Hg [150]. For a review see the article by Kresse [151].

4.2.3 Properties of Mixtures and Solvent Effects

In the following section examples of solvated molecules, ions, and ionic liquids in water as well as in methanol are given. Both the effects of the solvent on the solute as well as the opposite effects of the dissolved species on the solvent were considered. In many studies these effects are reflected in altered dipole moments. For an overview over the effects of salts on dipole moments see [91].

Solutes in water are of interest, because many chemical reactions are carried out in water and this liquid influences the solutes and chemical reactivity greatly. Hydrogen chloride HCl was investigated by Laasonen and Klein [152]. Several AIMD simulations were performed with additional water molecules. Starting from an HCl molecule in water, dissociation appeared forming H₃O⁺ and Cl[−] ions. Two different configurations for the proton were reported: an H₃O⁺ ion and an H₅O₃⁺ ion. An excellent overview of the solvated proton [153] and the hydroxyl ion in water was published by Marx [154]. The solvated electron was also investigated extensively from AIMD [155–158].

In a study from 2004, dimethylsulfoxide (DMSO) in water and its changing dipole moment was investigated [159]. From the CPMD simulations the dipole moment increase of DMSO from 3.97 D (isolated molecule [160]) to 7.39 D liquid was observed. The temperature dependence of the dipole moment for isolated DMSO was negligible; it increased from the geometry optimized value of 3.97 D to 4.08 K at 319 K.

The solvent effects of uracil in water [161], ethene and ethanol in water [162], as well as hydrogen in water [163] were discussed in detail elsewhere [91].

Solvent effects were found as well in the IR study of *N*-methylacetamide (NMA) (*cis* and *trans*) in aqueous solution carried out by Gaigeot and coworkers in 2005 [39]. From geometry optimization of *N*-methylacetamide microsolvated with a few water molecules, formation of bridges between the carbonyl functional group and the amide group would be expected. However, no such arrangement was found in the aqueous solution. A very noticeable effect of the solvent on the vibrational density of states was that the amide I stretching motion exhibited a red-shift (87 cm^{−1} for *trans*-NMA and 46–72 cm^{−1} for *cis*-NMA), whereas the amide II was blue-shifted (−8–25 cm^{−1} and −3–38 cm^{−1}, respectively). In general, solvent–solvent hydrogen bonds were observed to be energetically more favorable than solute–solvent hydrogen bonds [39]. Whereas in the gas phase the neutral form

was the most stable, in aqueous solution the zwitterionic state was more stable as has been observed for alanine [164].

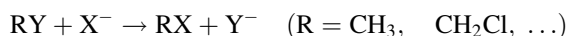
Salts, ions, and ionic liquids in water are widely studied in AIMD. Several anions [165–172], cations [153, 165, 173–182], and ion pairs [164, 183, 184], as well as ionic liquids ion pairs [185] in water were studied using AIMD. In all cases structural as well as dynamical properties of the ion's hydration shell were examined. In some cases the influence of the solvated ions on the water molecules were studied within the Wannier approach. In general, little effect of the halogen ions on the dipole moments of the water molecules in the first hydration shell was observed, while further water molecules remain unaffected. In contrast to this, it was observed that cations increase the dipole moments of the first hydration shell water by approximately 0.2–0.5 D. The second hydration shell and the bulk phase water molecules were mostly unaffected with regard to the dipole moment by the cations as well [91].

4.3 Chemical Reactions

4.3.1 Metal-Free Organic Reactions

In the following, AIMD studies of the S_N2 reaction shall be briefly reviewed. Other metal-free organic reactions like the Diels–Alder and the Wittig reaction have been discussed elsewhere [91].

Between 1999 and 2004, several S_N2 reactions of the type



were investigated from AIMD simulations in vacuum as well as in solution [186–192]. Raugei et al. found that the dipole moment changed drastically along the applied reaction coordinate in a gas phase S_N2 reaction investigation [186]. In a subsequent study, Raugei et al. added one and two water molecules to the reactants, and they observed important hydrogen bonds between the substrate as well as the ion X^- and the water molecules [187].

The complete substitution reaction in water was calculated by Pagliai et al. in 2003 [189]. The authors investigated hydrogen bonds from Wannier functions and the electron localization function (ELF) during the reaction. They found the charge at the transition states to be delocalized and, as a result, weakened and shorter lived hydrogens bonds. Similar results were obtained in other investigations [188, 190].

In 2003 and 2004, Ammal et al. [191] and Yang et al. [192] showed how temperature and dynamical effects can alter the chemical reactions even more than classical concepts in organic chemistry predict.

4.3.2 Metal-Organic Reactions and Catalysis

In 2004 and 2005 the photochemical activation of dinitrogen with transition metal model complexes of the Sellmann type nitrogenase was studied using CPMD [193, 194]. A dinuclear complex – designed to emulate the open-side FeMoco model – was simulated. Several side reactions were observed which have to be suppressed in order to arrive at the reduced species [194]. Chelate effects and their partial dissociation as well as low temperatures led to successful events. An optimized design of the complexes to inhibit side reactions was suggested [194].

In a subsequent study from 2009, the last step in the dinitrogen reduction, i.e., the ammonia-dinitrogen exchange of the Schrock's molybdenum catalyst, was examined [195]. For this purpose the complete Schrock catalyst without any simplifications was simulated with the CPMD approach. Several exchange mechanisms were observed. All constituted the addition-elimination type via a single stable six-coordinate intermediate. No dissociation-addition mechanism occurred in accordance with experiments. Furthermore, a possible detection of the intermediate by a significantly different $\text{N}\equiv\text{N}$ IR mode in the intermediate in comparison to other $\text{N}\equiv\text{N}$ stretching modes in similar complexes was suggested [195].

In 2007, Urakawa et al. investigated the rational design of ruthenium CO_2 hydrogenation catalysts from AIMD simulations [196]. The authors established the concerted CO_2 insertion by a mechanism that involves the rotation of the formate group. Several interesting intermediates were observed along the reactive trajectories. One example was the complex with molecular H_2 coordinated to $[\text{Ru}(\eta^2\text{-H}_2)]$. The most relevant structures were discussed in detail and their relative stability was calculated in terms of the interatomic interactions as well as the associated electronic charge distribution. The free-energy profiles reconstructed by MTD were consistent with experimental results and provided a more precise interpretation of the observed behavior. Urakawa et al. concluded that the reaction proceeds more easily by the *trans*-isomer route and H_2 insertion into the formate complex, which is the rate-determining step of the reaction. On the basis of the disclosed reaction pathways, a procedure that predicts the activity of catalysts with different ligands was proposed.

Another catalytic reaction was studied in 2007, namely the C–C and C–H reductive eliminations at Pt(IV) complexes [197]. The octahedral Pt(IV) complexes of the formula $\text{L}_2\text{Pt}(\text{CH}_3)_3\text{X}$ (with $\text{X}=\text{H}$ or CH_3) contained as L_2 diphosphine model ligands of dppe and dppbz. These two different chelating diphosphine ligands are dppe (bis(diphenylphosphino)-ethane, $\text{PPh}_2(\text{CH}_2)_2\text{PPh}_2$) and dppbz (*o*-bis(diphenylphosphino)benzene, *o*- $\text{PPh}_2(\text{C}_6\text{H}_4)\text{PPh}_2$), of which the latter is a less fluxional ligand compared to dppe because of its benzene backbone. Due to the difference in rate constants for the C–H (no influence) and the C–C (large influence) reductive elimination it was assumed earlier that a dissociative mechanism takes place for C–C the reductive elimination and a direct mechanism for the C–H reductive elimination. This so-called Crumpton–Bregel and Goldberg rule

was thoroughly investigated from MTD. Free energy activations were calculated for the C–H and C–C reductive elimination but also for the dissociation of one arm of the diphosphine ligand. Thereby, Michel et al. estimated the free energy cost thus including entropy effects and the Pt–P distance of the transition state structure. The authors deduced that, from a mechanistic point of view, the C–C reductive elimination occurs through a two-step dissociative pathway with barriers of around 19 and 16 kcal mol^{−1} if the less rigid ligand dppe is used. From kinetic simulations it was shown that this combination of values provides results comparable to a first-order kinetics with a barrier of around 40 kcal mol^{−1}. If the more rigid ligand, dppbz, was treated, the increase of the dissociation cost prevented the system from being reactive. For C–H reductive elimination, two mechanisms were found, the direct one previously postulated and a new one – the concerted mechanism discovered from MTD. In the concomitant mechanism the platinum-phosphorus bond formation occurred simultaneously with the C–H bond formation. Depending on the cost of the phosphine dissociation, the direct or the concomitant mechanism was observed. Thus, the strong influence of the phosphine ligand's basicity as well as the influence of its intrinsic rigidity was detected. A subsequent study was undertaken in 2008 [198].

4.4 *Electrochemistry*

This section of complex electrochemical reactions in solution and on electrodes is divided into three parts regarding the following questions. First, how does the solvent interact with the unbiased and biased metal surface? Second, how does the oxidation/reduction of a single electrochemical active species work in pure solvents? And finally, how does a complex electrochemical reaction proceed in solution and on metal surfaces? Therefore, metal–liquid interfaces are discussed at the beginning, followed by half cell reactions in solvents, and finally complex redox reactions in metal–liquid interfaces are reviewed.

4.4.1 **Metal–Liquid Interfaces**

In 2001, Izvekov and coworkers investigated the Cu(110)–water [199] and the Ag (111)–water [200] interface from AIMD simulations. In both simulations an absorption of water on the surface and a bilayer structure of water was found, in which water was tightly bound to the metal surface in the first shell. Exploration of the interface's electronic structure showed strong coupling of the water molecules and the metal. However, the metal surface remained almost undisturbed in the presence of water, both geometrically as well as electronically.

In 2007 and 2008, Sugino et al. [201] and Otani et al. [202] investigated biased platinum–water interfaces. Sugino et al. found that an orientation of the water molecules emerged due to the negative bias potential of the water–Pt(111) interface

and, furthermore, that the water molecules screened the electric field due to these reorientations (almost completely in the first shell) [201]. Similar results were obtained from the simulations by Otani et al. in which an O-down configuration (oxygen is attached to the Pt surface) was found in the neutral interface, while at the negative biased interface mostly H-down configurations (hydrogen is directed to the Pt surface) occurred.

4.4.2 Redox-Reactions in Solution

Since 2004, several redox and half cell reactions in solution have been studied from AIMD simulations; see Table 1 [203–206].

Please note that this list is far from being complete. It is impossible to discuss all studies in detail but one special case bears going into detail, i.e., reaction (n) of Table 1 will be briefly reviewed along with the main facts of the other studies.

In all studies AIMD simulations of the ions were carried out in solution (aqueous or organic) and the Marcus theory was applied to calculate the electrochemical potential. All electrochemical potentials were in good to very good (error ≤ 0.2 V) agreement with experimental data. The reaction (n) from Table 1 is the redox reaction of two rubredoxin molecules: *Clostridium pasteurianum* rubredoxin CpRd and *Pyrococcus furiosus* rubredoxin PfRd [221]. Sulpizi et al. used X-ray structures for their study in 2007 [221]. Classical molecular dynamics simulations were carried out, and at every 100 ps a configuration was selected in order to perform an electronic structure calculation with the CP2k program.

From these calculations, under application of the Marcus theory, which leads to the formula

Table 1 Selected redox reaction investigated in solution since 2004, where bpy is 2,2'-bipyridine, TH thianthrene, TTF tetrathiafulvalene, Q 1,4-benzoquinone, CpRd *Clostridium pasteurianum* rubredoxin, and PfRd *Pyrococcus furiosus* rubredoxin. For reviews see [203–206]

(a)	$\text{Mn}^{2+} \rightarrow \text{Mn}^{3+} + e^-$	[207]
(b)	$\text{MnO}_4^{2-} \rightarrow \text{MnO}_4^- + e^-$	[208]
(c)	$\text{Cu}^+ \rightarrow \text{Cu}^{2+} + e^-$	[209–211]
(d)	$\text{Ru}^{2+} \rightarrow \text{Ru}^{3+} + e^-$	[212–215]
(e)	$\text{Ru}(\text{bpy})_3^{2+} \rightarrow \text{Ru}(\text{bpy})_3^{3+} + e^-$	[214]
(f)	$[\text{RuCl}_6]^{4-} \rightarrow [\text{RuCl}_6]^{3-} + e^-$	[216]
(g)	$[\text{Ru}(\text{CN})_6]^{4-} \rightarrow [\text{Ru}(\text{CN})_6]^{3-} + e^-$	[216]
(h)	$\text{RuO}_4^{2-} \rightarrow \text{RuO}_4^- + e^-$	[208]
(i)	$\text{RuO}_4^- + \text{H}_2\text{O} + e^- \rightarrow [\text{RuO}_3(\text{OH})_2]^{2-}$	[217]
(j)	$\text{Ag}^+ \rightarrow \text{Ag}^{2+} + e^-$	[209, 210, 218]
(k)	$\text{TH}^{*+} + \text{TTF} \rightarrow \text{TH} + \text{TTF}^{*+}$	[219]
(l)	$\text{TH}^{2+} + \text{TTF}^{*+} \rightarrow \text{TH}^{*+} + \text{TTF}^{2+}$	[219]
(m)	$\text{Q}^- \rightarrow \text{Q} + e^-$	[220]
(n)	$\text{CpRd}^- + \text{PfRd} \rightarrow \text{CpRd} + \text{PfRd}^-$	[221]

$$\Delta A_X = \frac{1}{2} (\langle E_X \rangle_{\text{red}} + \langle E_X \rangle_{\text{ox}}) \quad (X = \text{CpRd, PfRd}), \quad (54)$$

Sulpizi et al. gained a redox potential difference

$$\Delta \Delta A = \Delta A_{\text{CpRd}} - \Delta A_{\text{PfRd}} \quad (55)$$

of -40 mV. The experimental value is -60 mV. The electrochemical properties of the other reactions listed in Table 1 were obtained in a similar fashion. For the smaller systems with only one cation or anion in water a full AIMD treatment is feasible.

4.4.3 Complex Electrochemical Interfaces and Electrochemical Reactions on Surfaces

In this part, complex electrochemical interfaces and electrochemical reactions on surfaces with various molecules in solvents will be discussed. Examples are the oxidation and evolution of *hydrogen* on different transition metal surfaces, the reduction of *oxygen* on several surfaces as well as *carbon monoxide* reactions, and a *complex photoactive reaction* in a solar cell.

Hydrogen under electrochemical conditions was investigated very recently [222, 223]. Santana et al. investigated the electro-oxidation of molecular hydrogen at the Pt(110)–water interface [222]. The Tafel–Volmer mechanism with a homolytic H–H bond cleavage followed by the formation of adsorbed terminal hydrogen atoms and further oxidation of the H atoms was observed by the authors. Furthermore, Santana et al. found the potential dependent activation energies for this process to be in accordance with experimental results.

Skúlason et al. investigated the hydrogen oxidation as well as evolution reaction on a Pt(111) surface under electrochemical conditions [223]. Three steps were examined, the Tafel, Heyrovsky, and Volmer steps. Skúlason et al. found that the rate determining steps on Pt(111) surface consisted of the Tafel–Volmer cascade for the oxidation and the Volmer–Tafel cascade for the evolution. Additionally, the H adsorption energy and energy barriers for the Tafel reaction were calculated for many metals¹ with different faces and steps. Their results suggested that the binding free energy of hydrogen is the most important parameter for describing oxidation and evolution activity of an electrode.

Oxygen and its electroreduction on a Pt(111) surface was studied under electrochemical conditions by Wang and Balbuena in 2004 [224]. They observed a stepwise adsorption of two oxygen atoms with a very low energy barrier (0.08 eV) and no clear barrier for the decomposition was found. Addition of H_3O^+ from the surrounding

¹Au, Ag, Cu, Pt, Ni, Ir, Rh, Co, Ru, Re, W, Mo, and Nb.

water led to a rapid formation of a proton transfer intermediate $\text{H}^+-\text{O}_2 \cdots \text{Pt}(111)$ followed by an electron transfer to $\text{H}-\text{O}-\text{O}-\text{Pt}(111)$. Wang and Balbuena found that the formation of $\text{H}-\text{O}-\text{O}-\text{Pt}(111)$ has a much higher activation barrier (0.4 eV) than its dissociation (0.1 eV) and that, therefore, the rate determining step for the first electron transfer reaction is the electroreduction of O_2 .

In 2008, the oxygen reduction on a $\text{ZrO}_2(111)$ surface was calculated by Okamoto [225]. During the reactions a spontaneous bond cleavage in HOOH suppressed termination of the reduction reaction at the $2e^-$ step. These simulations showed that at least reduction to HO on the surface should be possible and further reactions could only be hindered by OH poisoning the surface.

In 2009, Hirunsit and Balbuena published AIMD simulations of a $\text{Pt}(111)$ - and a Pt-Co -alloy-water interface and oxygen [226]. Different oxygen coverages were investigated as well as surface reconstruction effects due to different coverages of the adsorbed oxygen. Additionally, an electric field (-0.51 to $+0.51$ V/Å) was applied on the surface but no spontaneous water dissociation or oxygen reduction was observed. Only the reorientation of the water molecules from O-down to H-down orientations was observed, as previously found [201, 202] and is already discussed in Sect. 4.4.1.

Gas phase partial and complete reduction of oxygen by different hydrogen covered transition metal² (111) surfaces with static but periodic calculations were examined by Ford et al. in 2010 [227].

Carbon monoxide was investigated on a Pt surface as well as on a Pt-Ru -alloy surface with water by Santana and Ishikawa in 2010 [228]. The simulations revealed new interpretations for the adsorbed CO and water interactions, as well as rationalized observed quantitative relationship between IR intensities and Pt and Pt-Ru -alloy due to water molecules firmly hydrogen bonded to bridging CO molecules. Furthermore, the authors found the linear dependency of the O-H stretching mode with the potential and the CO coverage.

The *photoactive part of dye sensitized solar cells* consists of a wide band gap semiconductor covered by a monolayer of sensitizing dye [229]. The semiconductor can be directly supported by a transparent electrode on one side, while the dye is connected to the back electrode via a liquid electrolyte or a solid hole conducting material. The initial step of the photovoltaic process is a light induced electron injection from the dye into the semiconductor material. This process yields an oxidized dye and an energetic electron. Rapid regeneration (reduction) of the dye by the electrolyte prevents back transfer of the electron or degradation of the photo-oxidized dye. Meanwhile, the energetic electron diffuses away from the dye, passing through the electrode and an external load, finally reaching the counter electrode where it regenerates the electrolyte. From AIMD simulations Schiffmann et al. identified a highly efficient mechanism for the regeneration of the *cis*-bis(isothiocyanato)bis(2,2'-bipyridyl-4,4'-dicarboxylato)-ruthenium(II) sensitizing dye

²Rh, Ir, Ni, Pd, Pt, Cu, Ag, and Au.

(N3) by I^- in acetonitrile. A barrier-free complex formation of the oxidized dye with both I^- and I_2^- , and facile dissociation of I_2^- and I_3^- from the reduced dye, were determined to be key steps in this process. The authors also carried out in situ vibrational spectroscopy and could thus confirm the reversible binding of I_2 to the thiocyanate group. Furthermore, Schiffmann et al. were able to simulate the electrolyte near the interface and found that acetonitrile is able to cover the (101) surface of anatase with a passivating layer that inhibits direct contact of the redox mediator with the oxide [229, 230]. It was also observed that the solvent structure specifically enhances the concentration of I^- at a distance which further favors rapid dye regeneration.

5 Summary

This review serves as an overview of modern aspects concerning methodology as well as applications of ab initio molecular dynamics simulations.

First, a general introduction into the ab initio molecular dynamics simulations technique of the Car–Parrinello type was given. The derivation of forces and equations of motion were explained. In the last part of this introductory section, generalizations according to Niklasson were detailed.

Next, difficulties encountered in ab initio molecular dynamics simulations were discussed. Topics covered were massive parallelization to address computer time problems, basis set considerations, density functionals and van der Waals interactions, relativistic corrections, and new integration schemes. Several simulation techniques used to gain chemical insight were summarized. Enhanced sampling methods, metadynamics and other methods to explore free energy surfaces, reaction pathways and transition states were covered. Simulation of spectra (IR, NMR, EXAFS) from ab initio molecular dynamics simulations was the subject of the remaining paragraphs.

The last section was devoted to a range of real-world applications treated with ab initio molecular dynamics simulations. Results of gas to liquid phase transition simulations, structural and dynamical properties of liquids such as common solvents as well as the emerging neoteric media of ionic liquids were presented. After a short discussion of chemical reactions concerning homogeneous catalysis, we presented an overview of electrochemical reactions and related processes.

We hope this choice of topics showed that, despite some difficulties, ab initio molecular dynamics simulation is nowadays capable of analyzing and predicting real-world processes, especially those which are poorly accessible through experiments or other theoretical techniques.

Acknowledgment This work was supported by the DFG, in particular by the projects KI-768/4-2, KI-768/5-1, KI-768/5-2, KI-768/5-3, KI-768/6-1 and KI-768/7-1.

References

1. Ehrenfest P (1927) *Z Phys A* 45:455–457
2. Dirac PAM (1930) *Proc Camb Phil Soc* 26:376–385
3. Marx D, Hutter J (2009) *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, Cambridge
4. Car R, Parrinello M (1985) *Phys Rev Lett* 55:2471–2474
5. Andersen HC (1980) *J Chem Phys* 72:2384–2393
6. Parrinello M, Rahman A (1980) *Phys Rev Lett* 45:1196–1199
7. Niklasson AMN, Tymczak CJ, Challacombe M (2006) *Phys Rev Lett* 97:123001
8. Niklasson AMN (2008) *Phys Rev Lett* 100:123004
9. Tuckerman ME (2010) *Statistical mechanics: theory and molecular simulations*. Oxford University Press, Oxford
10. Allen MP, Tildesley DJ (1987) *Computer simulation of liquids*. Oxford University Press, Oxford
11. Frenkel D, Smit B (2002) *Understanding molecular simulation – from algorithms to applications*, 2nd edn. Academic, San Diego
12. Huber H, Dyson AJ, Kirchner B (1999) *Chem Soc Rev* 28:121–133
13. Szabo A, Ostlund NS (1996) *Modern quantum chemistry: introduction to advanced electronic structure theory*. Dover Publications, Inc., Mineola, New York
14. McQuarrie DA, Simon JD (1997) *Physical chemistry: a molecular approach*. University Science Books, Sausalito, California
15. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) *J Comput Phys* 23:327–341
16. Tangney P (2006) *J Chem Phys* 124:044111
17. Kolafa J (2004) *J Comput Chem* 25:335–342
18. Niklasson AMN, Steneteg P, Odell A, Bock N, Challacombe M, Tymczak CJ, Holmstroem E, Zheng G, Weber V (2009) *J Chem Phys* 130:214109
19. Odell A, Delin A, Johansson B, Bock N, Challacombe M, Niklasson AMN (2009) *J Chem Phys* 131:244106
20. Steneteg P, Abrikosov IA, Weber V, Niklasson AMN (2010) *Phys Rev B* 82:075110
21. Brommer KD, Larson BE, Needels M, Joannopoulos JD (1993) *Comput Phys* 7:350–362
22. Marx D, Hutter J (2009) *Ab initio molecular dynamics: basic theory and advanced methods*. In: *Modern methods and algorithms of quantum chemistry*. p 301
23. Gygi F (2006) *J Phys Conf Ser* 46:268–277
24. Hutter J, Curioni A (2005) *Parallel Comput* 31:1–17
25. Hutter J, Curioni A (2005) *Chem Phys Chem* 6:1788–1793
26. Giannozzo P, Cavazzoni C (2009) *Il Nuovo Cimento* 32:49–52
27. VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J (2005) *Comput Phys Commun* 167:103–128
28. Haynes PD, Skylaris C-K, Mostofi AA, Payne MC (2006) *Phys Status Solidi B* 243:2489–2499
29. Resch M, Bönisch T, Tiyyagura S, Furui T, Seo Y, Bez W (eds) (2007) *High performance computing on vector systems 2006*. Springer, Berlin
30. Thar J, Reckien W, Kirchner B (2007) *Top Curr Chem* 268:133–171
31. Fattebert J-L, Gygi F (2006) *Phys Rev B* 73:115124
32. Bowler DR, Choudhury R, Gillan MJ, Miyazaki T (2006) *Phys Status Solidi B* 243:989–1000
33. Lee H-S, Tuckerman ME (2006) *J Phys Chem A* 110:5549–5560
34. Hine NDM, Haynes PD, Mostofi AA, Skylaris C-K, Payne M (2009) *Comput Phys Commun* 180:1041–1053
35. Genovese L et al (2008) *J Chem Phys* 129:014109
36. Lippert G, Hutter J, Parrinello M (1997) *Mol Phys* 92:477–488
37. Artacho E et al (2008) *J Phys Condens Matter* 20:064208

38. Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M (2009) *Comput Phys Commun* 180:2175–2196
39. Gaigeot M-P, Vuilleumier R, Sprik M, Borgis D (2005) *J Chem Theory Comput* 1:772–789
40. Cohen AJ, Mori-Sánchez P, Yang W (2008) *Science* 321:792–794
41. Heyd J, Scuseria GE, Ernzerhof M (2003) *J Chem Phys* 118:8207–8215
42. Zhao Y, Truhlar DG (2008) *Acc Chem Res* 41:157–167
43. Spencer J, Alavi A (2008) *Phys Rev B* 77:193110
44. Guidon M, Hutter J, VandeVondele J (2009) *J Chem Theory Comput* 5:3010–3021
45. Guidon M, Schiffmann F, Hutter J, VandeVondele J (2008) *J Chem Phys* 128:214104
46. Paier J, Janesko BG, Henderson TM, Scuseria GE, Grüneis A, Kresse G (2010) *J Chem Phys* 132:094103
47. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) *Phys Rev Lett* 92:246401
48. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2005) *Phys Rev Lett* 95:109902(E)
49. Langreth DC, Dion M, Rydberg H, Schröder E, Hyldgaard P, Lundqvist BI (2005) *Int J Quant Chem* 101:599–610
50. Langreth DC et al (2009) *J Phys Condens Matter* 21:084203
51. Lee K, Murray ED, Kong L, Lundqvist BI, Langreth DC (2010) *Phys Rev B* 82:081101
52. von Lilienfeld OA, Tavernelli I, Röthlisberger U, Sebastiani D (2004) *Phys Rev Lett* 93:153004
53. Tapavicza E, Lin I-C, von Lilienfeld OA, Tavernelli I, Coutinho-Neto M, Röthlisberger U (2007) *J Chem Theory Comput* 3:1673–1679
54. Grimme S (2004) *J Comput Chem* 25:1463–1473
55. Grimme S (2006) *J Comput Chem* 27:1787–1799
56. Grimme S, Antony J, Ehrlich S, Krieg H (2010) *J Chem Phys* 132:154104
57. Thar J, Kirchner B (2009) *J Chem Phys* 130:124103
58. Kühne TD, Krack M, Mohamed FR, Parrinello M (2007) *Phys Rev Lett* 98:066401
59. Dai J, Yuan J (2009) *Europhys Lett* 88:20001
60. Kühne TD, Krack M, Parrinello M (2009) *J Chem Theory Comput* 5:235–241
61. Klimeš J, Bowler DR, Michaelides A (2010) *J Phys Condens Matter* 22:074203
62. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) *J Chem Phys* 125:024106
63. Ren W, Vanden-Eijnden E (2002) *Phys Rev B* 66:52301
64. Mills G, Jónsson H, Schenter GK (1995) *Surf Sci* 324:305–337
65. Laio A, Parrinello M (2002) *Proc Natl Acad Sci USA* 99:12562–12566
66. Iannuzzi M, Laio A, Parrinello M (2003) *Phys Rev Lett* 90:238302
67. Martonňák R, Laio A, Bernasconi M, Ceriani C, Raiteri P, Zipoli F, Parrinello M (2005) *Z Kristallogr* 220:489–498
68. Ensing B, Laio A, Parrinello M, Klein MI (2005) *J Phys Chem B* 109:6676–6687
69. Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M (2005) *J Phys Chem B* 109:6714–6721
70. Ensing B, Vivo MD, Liu Z, Moore P, Klein ML (2006) *Acc Chem Res* 39:73–81
71. Cucinotta CS, Ruini A, Catellani A, Stirling A (2006) *Chem Phys Chem* 7:1229–1234
72. Ifimie R, Tuckerman ME (2005) *J Chem Phys* 122:214508
73. Kaupp M, Bühl M, Malkin VG (eds) (2004) *Calculation of NMR and EPR parameters – theory and application*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim
74. Sebastiani D, Parrinello M (2001) *J Phys Chem A* 105:1951–1958
75. Ochsenfeld C, Kussmann J, Koziol F (2004) *Angew Chem Int Ed* 43:4485–4489
76. Weber V, Iannuzzi M, Giani S, Hutter J, Declerck R, Waroquier M (2009) *J Chem Phys* 131:014106
77. Cavalleri M, Odelius M, Nilsson A, Pettersson LGM (2004) *J Chem Phys* 121:10065–10075
78. Cavalleri M, Odelius M, Nordlund D, Nilsson A, Pettersson LGM (2005) *Phys Chem Chem Phys* 7:2854–2858

79. Odelius M, Cavalleri M, Nilsson A, Pettersson LGM (2006) *Phys Rev B* 73:024205
80. Iannuzzi M, Hutter J (2007) *Phys Chem Chem Phys* 9:1599–1610
81. Iannuzzi M (2008) *J Chem Phys* 128:204506
82. Leetmaa M, Ljungberg M, Lyubartsev A, Nilsson A, Pettersson L (2010) *J Electron Spectrosc Rel Phenom* 177:135–157
83. Wannier GH (1937) *Phys Rev* 52:191–197
84. Savin A, Nesper R, Wengert S, Fässler TF (1997) *Angew Chem Int Ed* 36:1809–1832
85. Fukui K (1981) *Acc Chem Res* 14:363–368
86. Silvestrelli PL, Marzari N, Vanderbilt D, Parrinello M (1998) *Solid State Commun* 107:7–11
87. Fitzhenry P, Bilek MMM, Mariks NA, Cooper NC, McKenzie DR (2003) *J Phys Condens Matter* 15:165–173
88. Bühl M, Chaumont A, Schurhammer R, Wipff G (2005) *J Phys Chem B* 109:18591–18599
89. Kirchner B, Seitsonen AP (2007) *Inorg Chem* 46:2751–2754
90. Kirchner B, Sebastiani D (2004) *J Phys Chem A* 108:11728–11732
91. di Dio PJ, Kirchner B (2011) Stereo-electronic effects from ab initio molecular dynamics simulations. *Top Curr Chem*. Springer (accepted)
92. Silvestrelli PL, Parrinello M (1999) *Phys Rev Lett* 82:3308–3311
93. Silvestrelli PL, Parrinello M (1999) *J Chem Phys* 111:3572–3580
94. Kuo I-FW, Mundy CJ (2004) *Science* 303:658–660
95. McGrath MJ, Siepmann JI, Kuo I-FW, Mundy CJ, VandeVondele J, Hutter J, Mohamed F, Krack M (2006) *J Phys Chem A* 110:640–646
96. Handgraaf J-W, van Erp TS, Meijer EJ (2003) *Chem Phys Lett* 367:617–624
97. Whitfield TW, Crain J, Martyna GJ (2006) *J Chem Phys* 124:094503
98. Laasonen K, Sprik M, Parrinello M, Car R (1993) *J Chem Phys* 99:9080–9089
99. Fois ES, Sprik M, Parrinello M (1994) *Chem Phys Lett* 223:411–415
100. Sprik M, Hutter J, Parrinello M (1996) *J Chem Phys* 105:1142–1152
101. Silvestrelli PL, Bernasconi M, Parrinello M (1997) *Chem Phys Lett* 277:478–482
102. Asthagiri D, Pratt LR, Kress JD (2003) *Phys Rev E* 68:041505
103. Sharma M, Wu Y, Car R (2003) *Int J Quant Chem* 95:821–829
104. VandeVondele J, Mohamed F, Krack M, Hutter J, Sprik M, Parrinello M (2005) *J Chem Phys* 122:014515
105. Lee H-S, Tuckerman ME (2006) *J Chem Phys* 125:154507
106. Schmidt J, VandeVondele J, Kuo I-FW, Sebastiani D, Siepmann JI, Hutter J, Mundy CJ (2009) *J Phys Chem B* 113:11959–11964
107. Lin I-C, Seitsonen AP, Coutinho-Neto MD, Tavernelli I, Röhrlisberger U (2009) *J Phys Chem B* 113:1127–1131
108. Rogers RD, Seddon KR (2003) *Science* 302:792–793
109. Kirchner B (2009) In: Kirchner B (ed) *Ionic liquids*. *Top Curr Chem*, vol 290. Springer, Berlin, pp xi–xiii
110. Kirchner B (2009) In: Kirchner B (ed) *Ionic liquids*. *Top Curr Chem*, vol 290. Springer, Berlin, pp 213–262
111. Wasserscheid P, Welton T (eds) (2003) *Ionic liquids in synthesis*. Wiley-VCH, Weinheim
112. Rogers RD, Seddon KR (eds) (2005) *Ionic liquids III A: fundamentals, progress, challenges, and opportunities*. ACS Symposium Series, vol 901. American Chemical Society, Washington, DC
113. Del Pópolo MG, Lynden-Bell RM, Kohanoff J (2005) *J Phys Chem B* 109:5895–5902
114. Bhargava B, Balasubramanian S (2006) *Chem Phys Lett* 417:486–491
115. Bhargava BL, Balasubramanian S (2007) *J Phys Chem B* 111:4477–4487
116. Bhargava BL, Balasubramanian S (2008) *J Phys Chem B* 112:7566–7573
117. Thar J, Brehm M, Seitsonen AP, Kirchner B (2009) *J Phys Chem B* 113:15129–15132
118. Zahn S, Thar J, Kirchner B (2010) *J Chem Phys* 132:124506
119. Zahn S, Kirchner B (2008) *J Phys Chem A* 112:8430–8435
120. Izgorodina EI, Bernard UL, MacFarlane DR (2009) *J Phys Chem A* 113:7064–7072

121. Diraison M, Martyna GJ, Tuckerman ME (1999) *J Chem Phys* 111:1096–1103
122. Tsuchida E (2004) *J Chem Phys* 121:4740–4746
123. R  thlisberger U, Parrinello M (1997) *J Chem Phys* 106:4658–4694
124. Galli G, Hood RQ, Hazi AU, Gygi F (2000) *Phys Rev B* 61:909–912
125. Bonev SA, Militzer B, Galli G (2004) *Phys Rev B* 69:014101
126. Galli G, Martin RM, Car R, Parrinello M (1990) *Phys Rev B* 42:7470–7482
127. Marks NA, McKenzie DR, Pailthorpe BA, Bernasconi M, Parrinello M (1996) *Phys Rev B* 54:9703–9714
128. Silvestrelli PL, Parrinello M (1998) *J Appl Phys* 83:2478–2485
129. McCulloch DG, McKenzie DR, Goringe CM (2000) *Phys Rev B* 61:2349–2355
130. Wang X, Scandolo S, Car R (2005) *Phys Rev Lett* 95:185701
131. Kirchner B, Seitsonen AP, Hutter J (2006) *J Phys Chem B* 110:11475–11480
132. Senda Y, Shimojo F, Hoshino K (2002) *J Phys Condens Matter* 14:3715–3723
133. Anta JA, Madden PA (1999) *J Phys Condens Matter* 11:6099–6111
134. Foley M, Smargiassi E, Madden PA (1994) *J Phys Condens Matter* 6:5231–5241
135. Ji M, Gong XG (2004) *J Phys Condens Matter* 16:2507–2514
136.   tich I, Car R, Parrinello M (1991) *Phys Rev B* 44:4262–4274
137. Sugino O, Car R (1995) *Phys Rev Lett* 74:1823–1826
138. Senda Y, Shimojo F, Hoshino K (1999) *J Phys Condens Matter* 11:5387–5398
139. Gu T, Qin J, Xu C, Bian X (2004) *Phys Rev B* 70:144204
140. Jakse N, Pasturel A (2004) *J Chem Phys* 120:6124–6127
141. Pasquarello A, Laasonen K, Car R, Lee C, Vanderbilt D (1992) *Phys Rev Lett* 69:1982–1985
142. Godlevsky VV, Derby JJ, Chelikowsky JR (1998) *Phys Rev Lett* 81:4959–4962
143. Kresse G, Hafner J (1994) *Phys Rev B* 49:14251–14269
144. Kulkarni RV, Aulbur WG, Stroud D (1997) *Phys Rev B* 55:6896–6903
145. Chai J-D, Stroud D, Hafner J, Kresse G (2003) *Phys Rev B* 67:104205
146. Shimojo F, Munejiri S, Hoshino K, Zempo Y (2000) *J Phys Condens Matter* 12:6161–6172
147. Shimojo F, Hoshino K, Watabe M, Zempo Y (1998) *J Phys Condens Matter* 10:1199–1210
148. Jakse N, Pasturel A (2003) *Phys Rev Lett* 91:195501
149. de Wijs GA, Pastore G, Selloni A, van der Lugt W (1994) *Europhys Lett* 27:667–672
150. Kresse G, Hafner J (1997) *Phys Rev B* 55:7539–7548
151. Kresse G (1995) *J Non-Cryst Solids* 193:222–229
152. Laasonen K, Klein ML (1994) *J Am Chem Soc* 116:11620–11621
153. Kirchner B (2007) *Chem Phys Chem* 8:41–43
154. Marx D (2006) *Chem Phys Chem* 7:1848–1870
155. Marsalek O, Uhlig F, Frigato T, Schmidt B, Jungwirth P (2010) *Phys Rev Lett* 105:043002
156. Marsalek O, Uhlig F, Jungwirth P (2010) *J Phys Chem C* 114:20489–20495
157. Marsalek O, Frigato T, VandeVondele J, Bradforth SE, Schmidt B, Schtte C, Jungwirth P (2010) *J Phys Chem B* 114:915–920
158. Frigato T, VandeVondele J, Schmidt B, Schtte C, Jungwirth P (2008) *J Phys Chem A* 112:6125–6133
159. Kirchner B, Hutter J (2004) *J Chem Phys* 121:5133–5142
160. Kirchner B, Reiher M (2002) *J Am Chem Soc* 124:6206–6215
161. Gaigeot M-P, Sprik M (2003) *J Phys Chem B* 107:10344–10358
162. van Erp TS, Meijer EJ (2003) *J Chem Phys* 118:8831–8840
163. Kirchner B, Hutter J, Kuo I-FW, Mundy CJ (2004) *Int J Mod Phys B* 18:1951–1962
164. Thar J, Zahn S, Kirchner B (2008) *J Phys Chem B* 112:1456–1464
165. Tuckerman M, Laasonen K, Sprik M, Parrinello M (1995) *J Chem Phys* 103:150–161
166. Yarne DA, Tuckerman ME, Klein ML (2000) *Chem Phys* 258:163–169
167. Kuo I-F, Tobias DJ (2001) *J Phys Chem B* 105:5827–5832
168. Tobias DJ, Jungwirth P, Parrinello M (2001) *J Chem Phys* 114:7036–7044
169. Raugai S, Klein ML (2002) *J Chem Phys* 116:196–202
170. Leung K, Rempe SB (2004) *J Am Chem Soc* 126:344–351

171. Heuft JM, Meijer EJ (2005) *J Chem Phys* 122:094501
172. Heuft JM, Meijer EJ (2005) *J Chem Phys* 123:094506
173. Marx D, Sprik M, Parrinello M (1997) *Chem Phys Lett* 273:360–366
174. Ramaniah LM, Bernasconi M, Parrinello M (1999) *J Chem Phys* 111:1587–1591
175. Lyubartsev AP, Laasonen K, Laaksonen A (2001) *J Chem Phys* 114:3120–3126
176. Lightstone FC, Schwegler E, Hood RQ, Gygi F, Galli G (2001) *Chem Phys Lett* 343:549–555
177. Bakó I, Hutter J, Pálincás G (2002) *J Chem Phys* 117:9838–9843
178. Lightstone FC, Schwegler E, Allesch M, Gygi F, Galli G (2005) *Chem Phys Chem* 6:1745–1749
179. Ikeda T, Hirata M, Kimura T (2005) *J Chem Phys* 122:024510
180. Amira S, Spångberg D, Hermansson K (2006) *J Chem Phys* 124:104501
181. Sa R, Zhu W, Shen J, Gong Z, Cheng J, Chen K, Jiang H (2006) *J Phys Chem B* 110:5094–5098
182. Moret M-E, Tavernelli I, Chergui M, Rothlisberger U (2010) *Chem Eur J* 16:5889–5894
183. Todorova T, Hünenberger PH, Hutter J (2008) *J Chem Theory Comput* 4:779–789
184. Mallik BS, Siepmann JI (2010) *J Phys Chem B* 114:12577–12584
185. Spickermann C, Thar J, Lehmann SBC, Zahn S, Hunger J, Buchner R, Hunt PA, Welton T, Kirchner B (2008) *J Chem Phys* 129:104505
186. Raugei S, Cardini G, Schettino V (1999) *J Chem Phys* 111:10887–10894
187. Raugei S, Cardini G, Schettino V (2001) *J Chem Phys* 114:4089–4098
188. Pagliai M, Raugei S, Cardini G, Schettino V (2001) *Phys Chem Chem Phys* 3:4870–4873
189. Pagliai M, Raugei S, Cardini G, Schettino V (2003) *J Mol Struct (Theochem)* 630:141–149
190. Mugnai M, Cardini G, Schettino V (2003) *J Chem Phys* 118:2767–2774
191. Ammal SC, Yamataka H, Aida M, Dupuis M (2003) *Science* 299:1555–1557
192. Yang S-Y, Fleurat-Lessard P, Hristov I, Ziegler T (2004) *J Phys Chem A* 108:9461–9468
193. Reiher M, Kirchner B, Hutter J, Sellmann D, Hess BA (2004) *Chem Eur J* 10:4443–4453
194. Kirchner B, Reiher M, Hille A, Hutter J, Hess BA (2005) *Chem Eur J* 11:574–583
195. Schenk S, Kirchner B, Reiher M (2009) *Chem Eur J* 15:5073–5082
196. Urakawa A, Iannuzzi M, Hutter J, Baiker A (2007) *Chem Eur J* 13:6828–6840
197. Michel C, Laio A, Mohammad F, Krack M, Parrinello M, Milet A (2007) *Organometallics* 26:1241–1249
198. Michel C, Milet A (2008) *J Mol Struct (Theochem)* 852:54–61
199. Izvekov S, Mazzolo A, VanOpdorp K, Voth GA (2001) *J Chem Phys* 114:3248–3257
200. Izvekov S, Voth GA (2001) *J Chem Phys* 115:7196–7206
201. Sugino O, Hamada I, Otani M, Morikawa Y, Ikeshoji T, Okamoto Y (2007) *Surf Sci* 601:5237–5240
202. Otani M, Hamada I, Sugino O, Morikawa Y, Okamoto Y, Ikeshoji T (2008) *Phys Chem Chem Phys* 10:3609–3612
203. Blumberger J, Tateyama Y, Sprik M (2005) *Comput Phys Commun* 169:256–261
204. VandeVondele J, Ayala R, Sulpizi M, Sprik M (2007) *J Electroanal Chem* 607:113–120
205. VandeVondele J, Sulpizi M, Sprik M (2007) *Chimia* 61:155–158
206. Lynden-Bell R (2007) *Electrochem Commun* 9:1857–1861
207. Moens J, Seidel R, Geerlings P, Faubel M, Winter B, Blumberger J (2010) *J Phys Chem B* 114:9173–9182
208. Tateyama Y, Blumberger J, Sprik M, Tavernelli I (2005) *J Chem Phys* 122:234505
209. Blumberger J, Bernasconi L, Tavernelli I, Vuilleumier R, Sprik M (2004) *J Am Chem Soc* 126:3928–3938
210. Blumberger J, Sprik M (2004) *J Phys Chem B* 108:6529–6535
211. Blumberger J (2008) *J Am Chem Soc* 130:16065–16068
212. Blumberger J, Sprik M (2005) *J Phys Chem B* 109:6793–6804
213. Blumberger J, Sprik M (2006) *Theor Chem Acc* 115:113–126
214. Seidel R, Faubel M, Winter B, Blumberger J (2009) *J Am Chem Soc* 131:16127–16137

215. Oberhofer H, Blumberger J (2009) *J Chem Phys* 131:064101
216. Ayala R, Sprik M (2006) *J Chem Theory Comput* 2:1403–1415
217. Tateyama Y, Blumberger J, Ohno T, Sprik M (2007) *J Chem Phys* 126:204506
218. Blumberger J, Tavernelli I, Klein ML, Sprik M (2006) *J Chem Phys* 124:064507
219. VandeVondele J, Lynden-Bell R, Meijer EJ, Sprik M (2006) *J Phys Chem B* 110:3614–3623
220. Cheng J, Sulpizi M, Sprik M (2009) *J Chem Phys* 131:154504
221. Sulpizi M, Raugé S, VandeVondele J, Carloni P, Sprik M (2007) *J Phys Chem B* 111:3969–3976
222. Santana JA, Mateo JJ, Ishikawa Y (2010) *J Phys Chem C* 114:4995–5002
223. Skúlason E, Tripkovic V, Björketun ME, Gudmundsdóttir S, Karlberg G, Rossmeisl J, Bligaard T, Jónsson H, Nørskov JK (2010) *J Phys Chem C* 114:18182–18197
224. Wang Y, Balbuena PB (2004) *J Phys Chem B* 108:4376–4384
225. Okamoto Y (2008) *Appl Surf Sci* 255:3434–3441
226. Hirunsit P, Balbuena PB (2009) *Surf Sci* 603:3239–3248
227. Ford DC, Nilekar AU, Xu Y, Mavrikakis M (2010) *Surf Sci* 604:1565–1575
228. Santana JA, Ishikawa Y (2010) *Electrochim Acta* 56:945–952
229. Schiffmann F, VandeVondele J, Hutter J, Urakawa A, Wirz R, Baiker A (2010) *Proc Natl Acad Sci USA* 107:4830–4833
230. Schiffmann F, Hutter J, VandeVondele J (2008) *J Phys Condens Matter* 20:064206

Nanoscale Wetting Under Electric Field from Molecular Simulations

Christopher D. Daub, Dusan Bratko, and Alenka Luzar

Abstract Applying an electric field is a well-established experimental method to tune surface wettability. As accessible experimental length scales become shorter, the modification of interfacial properties of water using electric field must come to grips with novel effects existing at the nanoscale. We survey recent progress in understanding these effects on water interfacial tension and on water-mediated interactions using molecular simulations. We highlight the key role of external conditions in determining the system's response to applied electric field. We further discuss the role of appropriate boundary conditions in modeling polar fluids subject to collective polarization. The work reviewed here broadens the basic understanding of applied and internal field effects that can operate in condensed phase systems, from modulating local hydrophilicity/hydrophobicity of engineered and biological surfaces, to surface manipulation in nanofluidic devices.

Keywords Confined water · Interfacial hydrogen bonds · Nanofluidics · Nanoparticle ordering · Nanoscale electrowetting

Contents

1	Introduction	156
2	Continuum Thermodynamics	157
2.1	Confinement Effects on Liquids	157
2.2	Combined Effects of Confinement and Electric Field: Electrocapillarity	159

C.D. Daub, D. Bratko and A. Luzar (✉)

Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284-2006, USA
e-mail: dbratko@vcu.edu; aluzar@vcu.edu

3	Molecular Thermodynamics	161
3.1	Water in Hydrophobic Confinement and Applied Field	161
3.2	Resilience of the Hydrogen Bond Network in Polarized Water	164
4	New Effects at the Nanoscale	166
4.1	Effects of Field Direction and Polarity on the Wetting Properties	167
4.2	Wetting Free Energy	168
4.3	Water-Mediated Ordering of a Nanomaterial	170
5	Conclusions	173
	References	174

1 Introduction

Modulation of solid/liquid interfacial tension by the applied electric field is currently enjoying explosive growth in a wide range of applications: from electrospray ionization and ink-jet printing to electrical control of optical devices [1, 2]. Understanding the influence of an applied electric field on interfacial properties of water is of great interest to workers in the field of microfluidics [3, 4], in particular the electrowetting on dielectric (EWOD) [5]. There is also great interest from the biology perspective since strong fields E (water dipole energy $E \cdot d$ comparable to thermal energy $k_B T$) arise in ion channels of cell membranes [6–10], in membrane electroporation [11, 12], and at the active site on an enzyme [13]. Recent experiments [14–17] investigated the effect of electric field on contact angle, which also potentially impacts the stability of liquid–liquid interfaces [18] and may be pertinent to carbon nanotube sieves of O (1 nm) thickness [19]. There are excellent review articles on electrowetting from macroscopic perspective that the reader is referred to [16, 20–23].

The advent of micro- and nanoporous materials sparked renewed interest in wetting techniques including electrowetting in nanomaterials whose high surface-to-volume ratio makes these media especially difficult to permeate with water. Rapid developments in nanofluidics warrant a transition from continuum to molecular level descriptions [24]. Computer simulation offers unique possibilities for investigating molecular-level phenomena difficult to probe experimentally [25]. In this review chapter we focus on nanoscale effects that can currently be probed best via molecular simulations. These tools give us the predictive power to discover novel effects operating at short length scales.

The chapter is organized as follows. We start with macroscopic thermodynamic predictions and discuss the phase behavior of confined liquids in general in the absence of applied electric field. The primary focus is on capillary evaporation, a phenomenon that can be reversed in the presence of the electric field. The reader is directed to extensive excellent reviews [26] of capillary condensation. Next we focus on the combined effect of confinement and electric field on liquids structure and thermodynamics, water in particular, its stability against evaporation, and resilience of the hydrogen bond network in polarized water. We devote increased attention to issues of external conditions, as they determine how the system responds to applied electric field. We concentrate on systems maintaining

equilibrium with external bath at ambient temperature and pressure. Next we review novel nanoscale effects in electrowetting where polarity and field direction come into play. We briefly describe the method for calculating wetting free energies (and associated contact angles) in a field-exposed nanopore from pressure tensor calculations. We describe a novel mechanism for nanoparticle alignment, based on the notable dependence of surface free energy on the angle between applied field and nanoparticle surface. We end with conclusions and future possibilities.

2 Continuum Thermodynamics

2.1 Confinement Effects on Liquids

Spatial confinement can have significant effects on the phase behavior of a confined fluid compared to its bulk counterpart. Classical examples are the shift of the bulk gas–liquid condensation in adsorbing confinement towards lower pressures (capillary condensation), or shift towards higher pressures [27] in a lyophobic confinement (capillary evaporation) [28] and the corresponding shift of the vapor–liquid critical point [29–31]. In a simulation study of strongly coupled dipolar fluids (spherical particles with permanent point dipoles), Klapp and Schoen showed that the presence of confining walls can promote long-range parallel ordering of the dipoles [32, 33].

Confining a liquid between weakly attractive lyophobic surfaces (characterized by contact angles above 90°) at a sufficiently small separation will lead to spontaneous evaporation. This thermodynamic process is controlled by competition between bulk energetics (that favors the liquid phase) and surface energetics (that favors the vapor phase). The liquid-to-vapor transition occurs when the grand potential of the confined liquid and confined vapor are comparable [28, 34, 35]:

$$\Omega_l \sim -PV + 2A_w\gamma_{wl} = \Omega_v \sim -P_vV + 2A_w\gamma_{wv} + A\gamma \quad (1)$$

where $V = A_w D$ is the volume of the confined region, $A_w \propto L^2$ the wetted area of the wall, and $A \propto LD$ the area of the liquid/vapor interface. The A term is relevant because of finite lateral size. For an incompressible fluid the difference in the bulk pressure, P , and the pressure of the coexisting vapor, P_v , can also be replaced by $\rho\Delta\mu$ [36, 37], where ρ is the number density of the liquid and $\Delta\mu$ is the difference in the chemical potential of bulk liquid from the value at liquid–gas coexistence.

From the above equality, the general expression for the critical threshold distance for spontaneous expulsion of a liquid confined between surfaces of the *finite* lateral size, L , follows [28]:

$$D_c = 2\Delta\gamma / (P - P_v + b\gamma/L), \quad (2)$$

where

$$\Delta\gamma = \gamma_{wl} - \gamma_{vv} = -\gamma\cos\theta_c \quad (3)$$

is the Young equation [38] relating the difference in wall/vapor and wall/liquid surface tension to the surface tension of the free liquid–vapor interface, γ , and the contact angle θ_c . b is a geometry dependent constant of the order of unity. Equation (2) shows that for (laterally) small confinements, the critical separation approaches the confinement lateral size, $D_c = O(L)$. Berne and coworkers have used an analogous equation (2) to study capillary-evaporation induced collapse of ellipsoidal hydrophobic particles [34, 39]. For macroscopic surfaces $L \rightarrow \infty$, and $P \gg P_v$, (2) reduces to the well-known Kelvin equation [40],

$$D_c \simeq 2\Delta\gamma/P. \quad (4)$$

The thermodynamic effects of finite size and the kinetic barriers, ΔG^* , for the formation of vapor phase have been fully developed [41–46]. Macroscopic thermodynamics predicts that when we have two non-polar surfaces immersed in a liquid and bring them closer together, at a critical distance, D_c , liquid will be replaced by vapor (2). Due to a considerable free energy barrier for confinement-induced evaporation, however, the liquid phase is often metastable below D_c [36, 41, 43, 45, 47]. Coarse-grained simulations confirmed [45] macroscopic scaling predictions [48, 49]:

$$\Delta G^* \propto D^2, \quad \Delta G^* \propto 1/\cos\theta_c. \quad (5)$$

Combining the known result for the magnitude of the activation barrier and the evaporation rate of molecular water in a specified molecular confinement [41] with the above scaling results enables predictions of kinetic viability of expulsion of water over a range of length scales and between arbitrary physically and chemically modified hydrophobic surfaces with contact angles above 90° . In many practical situations the activation barrier for evaporation can lead to a strong metastability of confined liquid phase. For extended, strongly hydrophobic nanopores ($\theta_c \sim 135^\circ$), widths above ~ 1.6 nm proved sufficient to suppress capillary evaporation kinetically over practical simulation times [36, 41].

Extrapolation from intermediate to lower contact angles (closer to 90°) enables estimating ΔG^* at conditions where critical cavity sizes and barrier magnitudes make simulations impractical even for coarse-grained models [45]. This regime is important because low contact angles slightly above 90° are characteristic of many so-called hydrophobic surfaces observed in nature. If the relation $\Delta G^* \sim 1/\cos\theta_c$ is approximately valid, reducing the contact angle from, say 110 – 100° nearly doubles the barrier to evaporation. In view of barrier values for $\theta_c = 109^\circ$ (Fig. 6 in [45]), it is clear that in most naturally occurring systems we cannot expect to observe spontaneous evaporation except from a small molecular-sized confinement.

These predictions [45] have been confirmed in studying the capillary evaporation events within hydrophobic pockets of melittin dimers [50].

When the liquid-to-vapor transition is suppressed kinetically, the loosening of water structure characteristic of hydrophobic interfaces (so-called soft interfaces [51, 52]) can still be inferred from the rise in compressibility, κ , within the solvation layer. The increase in surface compressibility of water has been quantified from density fluctuations [36] and direct density dependence on the pressure [43]. Compressibility next to hydrophilic surfaces, on the other hand, remains virtually indistinguishable from that of bulk water [43, 53]. Local compressibility has also been shown to offer a viable measure of hydrophobicity at a molecular level [52, 54]. The issue will be addressed in the following sections as we describe the changes in surface compressibility revealed in a simulated electrowetting experiment.

2.2 Combined Effects of Confinement and Electric Field: Electrocapillarity

If an electric field is applied across the planar confinement, additional contributions reducing the wetting surface free energy (3) arise. Electrocapillarity can reverse the sign of $\cos\theta_c$, leading to *electrowetting* of a lyophobic surface. Ignoring any field dependence of liquid/vapor (γ) and solid/liquid (γ_{sl}) surface tensions, within continuum approximations, the macroscopic relation due to Lippmann [55] describes electrocapillarity by

$$\cos\theta_c(V) = \frac{\gamma_{sl} - \gamma_{sv}}{\gamma} - \frac{W_{el}(V)}{\gamma} = \cos\theta_o - \frac{W_{el}(V)}{\gamma}. \quad (6)$$

Here $W_{el}(V)$ is the change in electrostatic energy per unit area, associated with surface spreading of the liquid wetting both walls, V is the voltage across the interface, and θ_o is the contact angle in the absence of electric field. Precise form of W_{el} depends on system geometry and material properties but is generally presumed to be proportional to the areal electric capacitance of the interface, c , and the potential drop across the interface squared, $W_{el} \sim -cV^2/2$ [16, 20]. Inserting Lippmann's effective surface tension into Young's equation (3), we obtain the electrowetting Young-Lippmann equation:

$$\cos\theta_c = \cos\theta_o + cV^2/2\gamma \quad (7)$$

which describes the low voltage behavior of the contact angle [20, 56].

In macroscopic experiments, electrocapillarity effects usually originate in a thin surface layer where electric field is strong [20, 38]. In microscopic pores, on the other hand, the field permeates throughout the whole aqueous slab. Notwithstanding its

weak ionization, pure water in a nanopore can be considered to behave as a dielectric of relative permittivity ε because the double layer screening length associated with water ions exceeds the nanopore width by orders of magnitude. The field E_o can stem from charges on the plates of a capacitor [15] or can be attributed to charges in the nearby environment. Water can enter the confinement from external field-free reservoir at ambient temperature and pressure, hence the state of confined water is fully described by specifying the values of chemical potential μ , volume $A_w D$ ($A_w \gg D^2$ is the plate area and D inter-plate separation), temperature T , and field strength E .

For a uniform field, the continuum level approximation for the difference between electrical energies of water-filled (l) and empty (e) pores, treated as capacitors with areal capacitances $c_e = \varepsilon_o/D$ and $c_l = \varepsilon_r \varepsilon_o/D$, gives [7]

$$W_{el} = W_l - W_e \cong \frac{D}{2} (\varepsilon_r \varepsilon_o E^2 - \varepsilon_o E_o^2) \approx -\frac{\varepsilon_o D}{2} E_o^2. \quad (8)$$

In (8), $E_o = V_o/D$ is the applied electric field across the slit before reduction due to water polarization ($E_o \rightarrow E = O(E_o/\varepsilon_r)$). Equations (6) and (7) suggest an expression for apparent contact angle $\cos\theta_c \sim \cos\theta_c^o + \varepsilon_o D E_o^2/4\gamma$. Here, W_{el} is associated with volume rather than with the surface layer alone. Equation (6) also presumes that bare surface tensions, γ_{ab} ($a, b = s, l, v$) remain unaffected by the field. While the latter is usually true for the solid–vapor term, γ_{sv} , the alignment of water molecules in the field can modify molecular interactions at the surface and hence further affect γ_{sl} and γ_{lv} , an effect confirmed by simulations (see Sect. 4.2).

Favorable interaction of a polar liquid with electric field also results in increased liquid density ρ . To the first order, electrostriction is given by

$$d \ln \rho = \frac{\kappa \rho}{8\pi} \frac{\partial \varepsilon}{\partial \rho} d(E^2), \quad (9)$$

where E is local field and κ is isothermal compressibility [57, 58]. In common with predictions for other geometries [16], in weak fields (7) and (8) imply that the change in contact angle and relative increase in local density vary in proportion to field squared. Equations (6) and (7), combined with an appropriate estimate of areal capacitance of the surface, provide the basis for techniques to tune hydrophobic/hydrophilic surfaces electrically [20, 59, 60] while (8) offers a qualitative explanation of the role of ions in gating of biological channels [6, 7, 61].

As already discussed, in narrow hydrophobic confinements, unfavorable surface energetics can trigger capillary evaporation if $D < D_c$ [(2) and (4)] [36, 41, 44, 46, 48, 49], provided the kinetic barriers are surmountable [45]. Dzubiella and Hansen discussed a generalization of Kelvin equation that incorporates the effect of electric field [62]. The energy density change due to the field augments the PV term, which tends to suppress liquid expulsion, thus reducing the threshold separation D_c . For an extended planar confinement subject to a weak electric field E_o , the generalized relation takes the form [63]

$$D < -2\gamma_{lv}\cos\theta_0/\left(P + \frac{\varepsilon_0 E_0^2}{4\gamma_{lv}}\right). \quad (10)$$

The extended expression enables studies of phase instability in dispersions of charged particles like ionic colloids. In the context of this review, it helps to explain electrostriction or electrowetting in a confinement maintaining equilibrium with a field-free aqueous bath. Brunet et al. discussed the use of electric field to tune mixing/demixing equilibria in a multicomponent system [64, 65].

3 Molecular Thermodynamics

3.1 Water in Hydrophobic Confinement and Applied Field

Stabilization of liquid phase under an applied field conforms to experimental observations of electrostriction as the field attracts more water into the exposed region. The behavior is captured in several simulation studies, both in the bulk [66, 67] and confined [7, 53, 61, 66, 68] regimes. On the other hand, some studies reported field induced depletion or evaporation in bulk [69] and confined [57] water; hence it is of interest to discuss the underlying differences between these works.

Some of these differences can be attributed to different thermodynamic conditions (Fig. 1). Studying confined water in a field-exposed confinement, open to water flow from the external, field-free bath, is an entirely different situation from the case of an isobaric system, closed to water exchange. In the former scenario, water in confinement responds to applied field by increasing the density

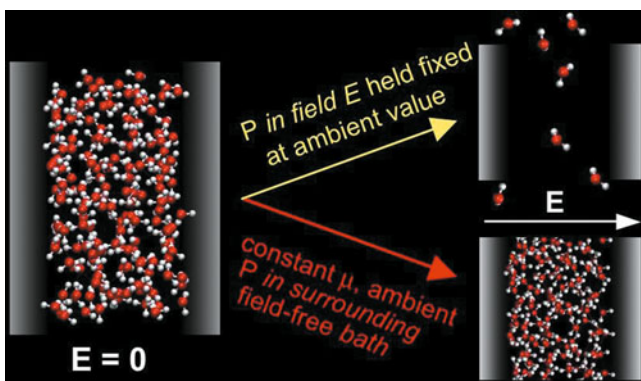


Fig. 1 Imposed external conditions lead to different responses of confined water to electric field. *Left*: field free water in a hydrophobic confinement. *Right*: field-exposed systems. *Top*: isobaric, mass conserving system of variable lateral dimensions. *Bottom*: isochoric confinement subject to applied field, and open to exchange of water with field-free aqueous bath

to equalize its chemical potential with that in the external bath. In the closed isobaric system, the amount of water is fixed and volume adjusts to maintain pressure constant despite the application of the field.

The response of the volume V in a mass-conserving isobaric (N, P, T) system to the application of field $E_o = |E_o|$ can be described by

$$\begin{aligned} \frac{\partial \langle V \rangle}{\partial E_o} &= \frac{\partial}{\partial E_o} \left(\Delta^{-1} \sum_{V_j} \sum_{\text{states } i} V_j e^{-\frac{U_i}{k_B T} - \frac{pV_j}{k_B T}} \right) \\ &= -(k_B T)^{-1} \left(\langle V \frac{\partial U}{\partial E_o} \rangle - \langle V \rangle \langle \frac{\partial U}{\partial E_o} \rangle \right) \end{aligned} \quad (11)$$

$$\text{where } \Delta = \sum_{V_j} \sum_i e^{-\frac{U_i}{k_B T} - \frac{pV_j}{k_B T}}.$$

Each of states i corresponds to a distinct configuration $[\mathbf{r}_N, \boldsymbol{\Omega}_N]$ consisting of positions $\mathbf{r}$ and orientations $\boldsymbol{\Omega}$ of all N particles. The angle brackets denote the ensemble average. The slope $\partial U / \partial E_o$ measures the ease with which the molecules align with the field, and can increase with fluid dilution. All N molecules are exposed to the field and any structural rearrangement takes place only to find the best compromise between molecular alignment with the field and orientation-dependent interactions among molecules. The density of a mass-conserving isobaric polar fluid can therefore decrease under applied field E_o . The prediction agrees with the (N, P, T) simulation in [69] but is not transferable to other external conditions.

Electrowetting experiments typically involve transfer of water from a field-free region or region with weak field to a region under strong field to maximize field/dipole interaction [20]. Pressure is therefore neither fixed nor uniform, but the chemical potential of an equilibrated system is uniform. These thermodynamic conditions are best described by grand canonical (μ, V, T) statistics with fixed volume (V), temperature, and chemical potential μ . The field dependence of the mean number of molecules, $\langle N \rangle$, in the field-exposed region is given by

$$\begin{aligned} \frac{\partial \langle N \rangle}{\partial E_o} &= \frac{\partial}{\partial E_o} \left[\Xi^{-1} \sum_N \sum_{\text{states } i} N e^{-\frac{U_i}{k_B T} - \frac{\mu N}{k_B T}} \right] \\ &= -\frac{1}{k_B T} \left(\langle N \frac{\partial U}{\partial E_o} \rangle - \langle N \rangle \langle \frac{\partial U}{\partial E_o} \rangle \right) \end{aligned} \quad (12)$$

$$\text{with } \Xi = \sum_N \sum_i e^{-\frac{U_i}{k_B T} - \frac{\mu N}{k_B T}}.$$

Electric field E_o affects the energy U through orientation-dependent interaction with molecular dipoles d , $U(N) = -E_o \sum_{i=1}^N d_i \cos \theta_i = -NE_o d \cos \theta$:

$$\begin{aligned} \partial N / \partial E_o &= \frac{d}{k_B T} \left(\left\langle N^2 \frac{\partial (E_o \overline{\cos \theta_{N,i}})}{\partial E_o} \right\rangle - \left\langle N \right\rangle \left\langle N \frac{\partial (E_o \overline{\cos \theta_{N,i}})}{\partial E_o} \right\rangle \right) \\ &= \frac{d}{k_B T} \left(\left\langle N^2 \overline{\cos \theta_{N,i}} \right\rangle - \left\langle N \right\rangle \left\langle N \overline{\cos \theta_{N,i}} \right\rangle \right), \end{aligned} \quad (13)$$

where $d=|d|$. $\overline{\cos \theta_{N,i}}$ measures the average dipole alignment, $dE_o/|d||E_o|$, of N molecules in the system in configuration (N,i) , because $\cos \theta_{N,i} \geq 0$ for all representative configurations, and the *product* $(N \overline{\cos \theta_{N,i}})$ generally increases with N . The density of a dipolar liquid in an open system will therefore rise with increasing field strength E_o as predicted by continuum analyses [57, 70, 71] and seen in electrostriction experiments.

Simulation studies for bulk and confined systems at different external conditions have been compared over a range of applied electric fields. Open (μ, V, T) systems invariably show density increase under applied field, both in bulk and confined phases, with or without Ewald periodic conditions. Bigger effects are restricted to hydrophobic confinements, as water fills the initially depleted interfacial layers. On the other hand, dilution and eventual evaporation are observed in mass conserving, isobaric systems when constancy of pressure is enforced in the presence of the field. Representative results are collected in Fig. 2.

3.1.1 Electric Fields in Modeled Systems

A few comments pertaining to simulation studies of field-exposed aqueous systems discussed in this review are in order to aid in evaluating the results, presented in Fig. 2 and in subsequent sections.

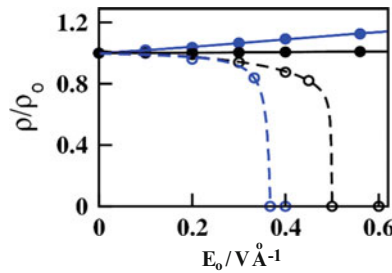


Fig. 2 Simulated water density dependence on the strength of (unscreened) applied field $0 \leq E_o \leq 0.6 \text{ V } \text{\AA}^{-1}$. The actual field, E , is lowered by orientational polarization of water, and spans the range $0 \leq E \leq 0.025 \text{ V } \text{\AA}^{-1}$. In confinement, dielectric screening renders the field both weaker and nonuniform. *Blue*: 1.64 nm wide hydrophobic confinement (wall contact angle 135°), *black*: bulk water. *Solid lines and circles* describe open systems (varying N) under the field, in equilibrium with field-free water bath. *Dashed lines and open symbols*: closed (mass conserving) isobaric systems. To keep pressure constant, these systems expand under the field. Note that different conditions (μ, V, T) (*solid symbols*) vs (N, P, T) (*open symbols*) correspond to *contrasting physical situations*, and *not* the same situation described using different ensembles

First, the reported field strengths warrant explanation. In the absence of tin-foil boundary conditions [72, 73], the actual strength of electric field spanning the aqueous phase, $\mathbf{E}(r)$, differs significantly from the unperturbed “applied” field $\mathbf{E}_o$ due to the opposing effect of water polarization. The actual, dielectrically screened field $\mathbf{E}(r)$ is generally nonuniform and is of the order ϵ_r^{-1} weaker than the unscreened field $\mathbf{E}_o$. In the bulk phase, modeled by Ewald summation with vacuum boundary conditions [72], the exact relation (for the absolute values of the field) is $E = 3E_o/(\epsilon_r + 2)$, where ϵ_r is the relative permittivity [74]. While constant-voltage simulation techniques have been presented [75, 76], in most cases, E_o represents input information for a simulation experiment and the screened, position dependent, field $\mathbf{E}(r)$ can be determined during the simulation. In the latter case, a viable estimate of the average field $\bar{E}$ across the system can be obtained from the observed polarization of water, measured in terms of the cosine of the alignment angle, $\langle \cos \theta_{N,i} \rangle$ [77–79]. When compared with a laboratory measurement, a simulation with fixed applied field E_o is akin to an experiment on a system between electrified surfaces (or capacitor plates) with fixed charge densities. The common fixed-voltage experiment, on the other hand, corresponds to preselecting the average field across the system. The implications of the two different constraints have been discussed by Jia and Hentschke [80]. In capturing electrostatic screening in a globally polarized system, simulations employing Ewald periodic conditions are generally superior to cutoff-based techniques. While the use of a distance cutoff on the intermolecular interactions mostly gives a satisfactory qualitative description of the system’s responses to the field, dielectric screening is typically underestimated, and the average field $\bar{E}$, and apparent voltage V across the system, $V \sim D\bar{E}$ are overestimated in this approach [68]. For clarity, both the exact value of the unscreened input field E_o and the approximate average of the actual field will be listed in most cases we discuss below.

The second comment concerns the usage of rigid water models such as the SPC/E model [81], which by design cannot undergo ionization or react chemically. We note that actual fields considered here are much too weak to polarize significantly, let alone decompose water when any flow of electric current is prevented by proper insulation. Top end insulators like polymer and silica films with dielectric strength of up to $5 \times 10^8 \text{ V m}^{-1}$ can provide more than adequate insulation.

3.2 Resilience of the Hydrogen Bond Network in Polarized Water

For an open system, described by (13), any field-induced density depletion, $(\partial N / \partial E_o)_{\mu VT} < 0$, could only be expected in case of dramatic rise in orientational polarizability of the molecules upon dilution. A mean-field analysis [69, 82] of a water-mimicking Ising model in electric field explored the assumption that molecular dipole alignment perturbs hydrogen bonding. Over an interval of intermediate field strengths, the model-system featured a density drop akin to the

reported evaporation from an open, field-exposed confinement [57] in equilibrium with unperturbed bath. Repeated observations of electrostriction (density increase under field) in other simulation studies of open systems [6, 7, 61, 63, 66–68] question the hypothesis [69] of strongly negative correlation between attractive water–water interactions (dominated by hydrogen bonding), and water’s ability to align with the applied field.

The issue was addressed in Monte Carlo and Molecular Dynamics simulations directly monitoring water structure and the extent of hydrogen bonding as the molecules became increasingly aligned by the applied field [68]. Below, we report results for fields E_0 up to $0.8 \text{ V } \text{\AA}^{-1}$, corresponding to actual fields of up to $0.03 \text{ V } \text{\AA}^{-1}$, thus spanning the range of fields detectable near charged electrodes, ion channels, ionic biomolecules or assemblies [6, 79, 83–93]. Within the above range of fields, no significant field-induced changes in water hydrogen bond populations, in atom–atom distribution functions or in water’s tetrahedral coordination were observed [68]. The resilience of hydrogen-bond strength and their population were confirmed by:

1. Direct population statistics: the number of bonds does not diminish upon polarization by the field.
2. The average energy of the bonds remained essentially equal to that in field-free water.
3. The preserved *free-energy* advantage of forming bonds was evidenced by unperturbed equilibrium between the populations of formed and ruptured bonds. Any weakening in the *free-energy* of the bonds in the field-oriented water would be reflected in at least some increase in the fraction of ruptured bonds, which did *not* happen.

Simulation results demonstrate that high alignment of aqueous dipoles can be achieved without serious penalties in the number and free-energies of hydrogen bonds or distortion of tetrahedral coordination. An onset of the transition from tetrahedral toward tightly-packed, highly coordinated “electrofrozen” structure can be observed at extreme fields of about $2 \text{ V } \text{\AA}^{-1}$ and higher [79, 94, 95]. Remarkable persistence of the hydrogen-bond network under aligning electric field equal or below $1 \text{ V } \text{\AA}^{-1}$ is evidenced by several studies [79, 88, 96–98] showing no evidence that field alignment would be conducive to unraveling of the hydrogen bonded network. For uniform phases, recent Gibbs ensemble simulations by Sieppman and coworkers [67] explicitly confirm that orientational polarizability of water molecules in liquid phase exceeds that in vapor.

A common Molecular Dynamics approach in studies of confined systems at fixed chemical potential employs an isobaric, bulk-like bath surrounding the confinement. The conditions inside a field-exposed confinement (e.g., inside a capacitor) should therefore be close to (μ, V, T) ones of GCMC cases [66, 68], suggesting similar electrostriction behavior. Occurrence of a field-induced expulsion of water has, however, been reported in a Molecular Dynamics study of this kind [57]. To secure proper barostat performance under strongly nonuniform and anisotropic pressure fields, and to account for the discontinuity between field/no

field regimes (impulse on crossing molecules) in (N, P, T) Molecular Dynamics are still open technical challenges. Avoiding explicit boundaries between field and no-field regions, Grand Canonical [61, 66, 68] and Gibbs ensemble [67] Monte Carlo approaches represent natural choices for molecular studies of electrowetting under nonuniform applied field. Local Molecular Field (LMF) theory [99] is poised to offer a promising alternative to molecular simulations in electrically nonuniform systems.

4 New Effects at the Nanoscale

Traditionally, electrowetting has been considered a relatively non-specific phenomenon defined in terms of a material's macroscopic properties. The effect of applied field on $\cos\theta_c$ has generally been found to be proportional to voltage squared, independent of field polarity and direction, as in (7). Nanosized aqueous confinements we review next, however, behave very differently from macroscopic systems. Due to molecular anisotropy of water, the electric field effect on the surface tension at aqueous interfaces *depends on the angle* between the field and the surface. The field-induced alignment competes with orientational preferences of interfacial water molecules relative to a wall to maximize hydrogen bonding [100–103]. The two trends are reconciled when the field is parallel to the interface. Perpendicular field, on the other hand, results in asymmetric wettability of opposing confinement surfaces (Fig. 3). These nanoscale effects, in contrast to conventional macroscopic experiments, derive primarily from the properties of the first solvation layer, and are negligible in systems of macroscopic dimensions. Analogous direction dependence is observed in sessile nanodroplets' response to electric field [104]. Preferred alignment of confinement surfaces with the field suggests a novel mechanism whereby the applied electric field can orient nanoparticles even in the absence of charges or dielectric contrast.

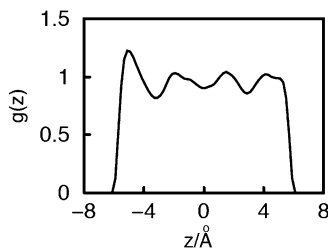


Fig. 3 Simulated density profile of water in an extended hydrophobic (hydrocarbon-like) nanopore subject to normal electric field of strength $0.2 \text{ V } \text{\AA}^{-1}$ (average actual field $E \sim 0.005 \text{ V } \text{\AA}^{-1}$) in equilibrium with a field-free aqueous reservoir. Model: SPC/E water. Method: GCMC with slab-corrected Ewald sums for laterally periodic boundary conditions

Examples discussed below illustrate field-direction effects in selected scenarios. Avoiding possible ambiguities related to the angle between the surface and the applied field in the presence of surface roughness, these examples focus on smooth interfaces described by two forms of substrate/water interaction. The integrated, laterally invariant Lennard-Jones (9–3) potential [102] was used in nanopores as it allows lateral scaling, thus facilitating surface free energy calculations [53, 66]. The Grand Canonical Transition Matrix Monte Carlo approach developed by Errington and coworkers enables extensions to nonplanar surfaces [105, 106]. Interfacial behavior at smooth surfaces was consistent with that observed on molecular substrates mimicking graphitic surfaces in sessile drop [104], thin films [107], and suspended nanoparticle calculations [108].

4.1 Effects of Field Direction and Polarity on the Wetting Properties

Molecular simulations of water equilibrium between nanopores under electric field and unperturbed bulk phase demonstrated notable differences between electrowetting at the nanoscale and in macroscopic systems [53, 66, 68]. Due to the coupling between surface/water and field/water forces, the water/wall surface tension depends on the alignment between the surface and the field. This, in turn, reflects in angle-dependent uptake of water in the pore. The transition from hydrophobic to hydrophilic behavior of paraffin-like nanopores upon imposition of electric field is enhanced when the field is parallel with confinement walls. In a narrow, 1.64 nm wide planar pore (a width just above kinetic threshold to capillary evaporation from a hydrocarbon-like pore [41]), simulated by GCMC, electrostriction (solid blue curve in Fig. 2), is about twice the strength in parallel vs normal fields. In a normal field, the coupling of field-induced alignment and orientational preferences of interfacial water molecules *relative to a wall* renders solvation layers at opposing walls completely different. As shown in Fig. 3, the hydrophobic wall under incoming field (pointing into aqueous phase) features a pronounced density peak of water/wall distribution function, $g(z)$, in the first solvation layer, reminiscent of hydrophilic hydration. The opposite wall, on the other hand, shows essentially no peak, in analogy with hydrophobic solvation in the absence of the field [66, 68]. Polarity dependence of water density profiles have also been reported between oppositely charged colloids [63] and between parallel graphene sheets under strong electric field [57]. When the field is applied along the walls, prominent hydration peaks form at both walls [66]. The sensitivity to field direction and polarity is explained by competition between spontaneous water molecule-surface orientations and molecular dipole alignment with the field. As do all polar fluids, aqueous dipoles prefer polarization parallel to walls. This general tendency acts in concert with orientational ordering imposed by hydrogen bonding. Under the *combined* effects, the optimal orientation of interfacial water molecules corresponds to dipoles almost aligned with the wall, but pointing slightly into the

liquid phase. Electrowetting is therefore most effective when the field polarizes water along the wall, but is considerably weaker when the field tends to turn water dipoles *toward* the wall. Consistent with the above picture, hydrogen bond populations, monitored as a function of field angle relative to the walls, are enhanced in parallel fields and depleted especially at interfaces with field pointing toward the wall [66]. Analogous preference for the interface/field alignment has been observed and discussed in recent simulation studies of nanodroplet elongation [109] and aqueous film evaporation [110] in the field. The surprisingly strong effect of field direction and polarity on surface wetting is a signature of the nanoscale regime where surface molecules represent a statistically significant constituency.

4.2 Wetting Free Energy

Field-enhanced wettability can be quantified in terms of wetting surface free energy, $\sigma(E)$, here defined as the sum $\Delta\gamma + W_{\text{el}}(E) = -\gamma \cos\theta_c(E)$ [(3) and (6)]. For smooth surfaces, σ inside an open nanopore of fixed width D has been shown [66] to relate to the lateral component of the pressure tensor, $P_{\parallel}$:

$$\sigma = \frac{\partial\Omega}{\partial A} = -\frac{P_{\parallel}D}{2}. \quad (14)$$

Here, Ω is the grand potential of the wetted part of the confinement atop the wetted area A with volume AD . A recent study reported systematic GCMC calculations of $\sigma(E)$ in hydrocarbon-like nanopores [66]. To estimate contact angles under the field also required calculations of surface tension (γ_{lv}) as a function of the field strength. Calculations for a free-standing aqueous slab were performed using the conventional relation $\sigma(E) = (P_{\perp} - P_{\parallel})D/2$. A novel finite-difference technique for the calculation of pressure tensor components determined energy differences ΔU_{α} , associated with uniform scaling of molecular coordinates α ($\alpha = z$ or x, y) and volume change ΔV_{α} [66]. $\Delta U_{\alpha} = U_{\alpha}(V + \Delta V_{\alpha}/2) - U_{\alpha}(V - \Delta V_{\alpha}/2)$ comprised changes in intermolecular and water-wall interactions. As described in Supporting Information to [66], pressure tensor components were obtained from the relation

$$P_{\alpha\alpha} = \rho kT + \lim_{\Delta V_{\alpha} \rightarrow 0} \frac{kT \ln \langle \exp(-\frac{\Delta U_{\alpha}}{kT}) \rangle}{\Delta V_{\alpha}} = \rho kT - \lim_{\Delta V_{\alpha} \rightarrow 0} \langle \frac{\Delta U_{\alpha}}{\Delta V_{\alpha}} \rangle. \quad (15)$$

Related finite-difference techniques have been studied in a number of contexts involving fluids with hard-core [111] and soft potentials [112–116]. The central finite-difference approximation, analyzed systematically in [115], was implemented [66] through scaling by a factor of $f = 1 \pm \varepsilon$ with $\varepsilon = 10^{-5}$ in forward and backward directions. Within the range $10^{-6} \leq \varepsilon \leq 10^{-4}$, no significant dependence on ε has been detected and exponential and linearized forms of (15) produced identical

results. The calculated normal component of the pressure tensor, $P_{zz} \equiv P_{\perp}$, matched the wall pressure calculated directly from wall/water forces as described elsewhere [36]. Consistency of the pressure tensor approach was also verified by direct surface free energy calculations using thermodynamic integration [117]. An interesting variant of thermodynamic integration with phantom wall approach was introduced by Muller-Plathe and coworkers [118]. Pressure, compressibility, and wetting free energies followed the trend described in discussing electrostriction and water structure in the presence of the field. As water is driven into the pore, the pressure rises approximately in proportion to field squared; however, the rise is steeper in a parallel field [66]. The change is accompanied with decreasing interfacial compressibility, approaching that of bulk water in the strong field limit (Fig. 4). Compressibility next to hydrophilic surfaces, on the other hand, proves essentially insensitive to applied field. The nonlinear dependence of compressibility on surface contact angle, highly sensitive at high θ_c and almost insensitive at $\theta_c < 90^\circ$, is illustrated in Fig. 5 [119]. The observed asymmetric dependence conforms to

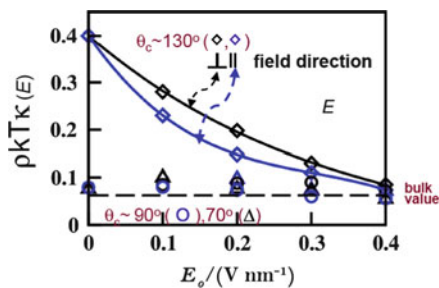


Fig. 4 Reduced compressibility, $\rho k_B T \kappa$, in a planar nanopore of width $D = 1.64$ nm as a function of applied (nonscreened) electric field E_o from Grand Canonical Monte Carlo simulations in SPC/E water for three wall/water contact angles $\theta_c = 135^\circ$ (diamonds), 93° (circles), or 69° (triangles). Actual field E ranging from 0 to $0.0095 \text{ V \AA}^{-1}$. Reduction in compressibility inside a hydrophobic pore, $\theta_c = 135^\circ$, is indicative of electrostriction in the field. Electrostriction is stronger in parallel than normal field

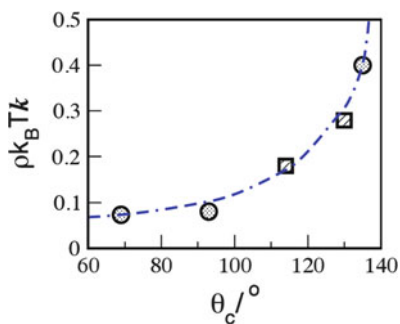


Fig. 5 Reduced compressibility, $\rho k_B T \kappa$, of water in a planar nanopore of width $D = 1.64$ nm as a function of surface contact angle, θ_c in SPC/E water [53, 66, 119]. Contact angle of chemically homogeneous pore walls was varied through surface chemistry (circles at $\theta_c = 135^\circ$, 93° , or 69°), or by applying electric field across a hydrophobic pore (squares at 114° and 129°)

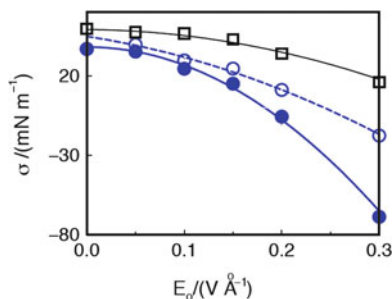


Fig. 6 Wetting free energy $\sigma(E_0)$ in 1.64 nm wide hydrophobic pore as a function of applied field E_0 (actual normal field E up to $0.008 \text{ V } \text{\AA}^{-1}$). Empty circles: perpendicular field, solid symbols: parallel field, and squares: water/air surface tension of a free-standing aqueous slab of identical thickness. Field direction was not significant in the latter case

reported sensitivity to minority polar groups on hydrophobic surface and marginal influence of hydrophobic groups in a polar context [52, 117, 120, 121].

Calculated wetting free energies, shown in Fig. 6, reveal the prominent effect of field direction. In a 1.64 nm wide pore and strong field $E_0 \sim 0.3 \text{ V } \text{\AA}^{-1}$, the wetting surface free energy in parallel field is nearly 50 mN m^{-1} lower than the average over both walls in a normal field. Again, the fields E_0 in Fig. 6 correspond to external field before reduction due to dielectric screening, which renders the field across the pore nonuniform and over an order of magnitude weaker ($\bar{E}$ up to $0.008 \text{ V } \text{\AA}^{-1}$). Based on the structures of hydration layers (Fig. 3), the solvation of the confinement wall with outgoing normal field is similar to that observed at both walls in the parallel field. The wall with incoming normal field, on the other hand, remains only weakly affected by the field. In sufficiently strong normal field, this asymmetry renders one wall strongly hydrophilic, the other hydrophobic. This situation, known as a Janus interface, shows very interesting behavior experimentally [122]. Janus interface can be produced, for example, by applying voltage of $\sim 0.1 \text{ V}$ ($E_0 \sim 0.2 \text{ V } \text{\AA}^{-1}$, average field $E \sim 0.005 \text{ V } \text{\AA}^{-1}$) across a 2 nm wide confinement, without modifying the surfaces themselves.

4.3 Water-Mediated Ordering of a Nanomaterial

A crucial step in the manufacture of many complex materials is the orientation of the constituents in a solvent such that they can be deposited on a substrate with a desired orderly structure. Several methods have been considered for an efficient solute orientation in a solvent [123, 124], but they all rely on the presence of a nanoparticle permanent dipole or considerable dielectric contrast between the particle and the medium [125, 126]. A newly proposed method of orienting nanoparticles [53] exploits the coupling between the field-alignment of polar solvent molecules and anisotropic solvent–solute interactions due to solvent

molecules' electrostatic and geometrical asymmetry. This coupling can lead to directional solvent–solute forces [66] and torques, which favor certain solute orientations, irrespective of the electric nature of the solute itself. A major contribution to the torque comes from the preference of water to maximize its hydrogen bonding [66, 104], augmenting the general tendency of dipolar fluids to spontaneously polarize parallel to the interface [32, 33, 127, 128]. According to macroscopic electrostatics, electric field will generally align anisotropic particles with dielectric constant *different* from that of the medium. The free energy F of an object of volume V_2 and permittivity ϵ_2 in the medium ϵ_1 , and associated aligning torque $\tau(\phi)$, is [125]

$$F = \frac{1}{2} \int_{V_2} (\epsilon_2 - \epsilon_1) \vec{E}_1 \cdot \vec{E}_2(\vec{r}) d^3\vec{r}, |\tau| \sim \left| \frac{\partial F}{\partial \phi} \right| \quad (16)$$

where E_1 is the field in the absence of the particle, E_2 the perturbed field, and ϕ the angle of rotation. At the molecular and nanometer scale, however, significant additional orientational forces operate *without* requiring the nanoparticle to have a dipole moment, or strongly contrasting permittivity. For example, this effect could orient a solute of permittivity close to that of the solvent in a large local electric field next to a DNA polyion (actual field $E = O(10^{-2})\text{V } \text{\AA}^{-1}$, $E_o = O(1)\text{V } \text{\AA}^{-1}$) [83].

4.3.1 Nanopore Geometry

Because of the coupling between orientational forces on surface water molecules and their aligning with the field, the wetting free energy of nanopore walls depends on the angle ϕ between the walls and the direction of the field. In a planar pore, the wetting free energy, $\sigma(E, \phi)$, is lowest in the parallel orientation, gradually increasing as the angle ϕ approaches 90° . Figure 7 illustrates this dependence for planar nanopores of two widths, 1.64 and 2.7 nm. The applied field across the pore before accounting for dielectric screening equaled $0.2 \text{ V } \text{\AA}^{-1}$, with the average actual field

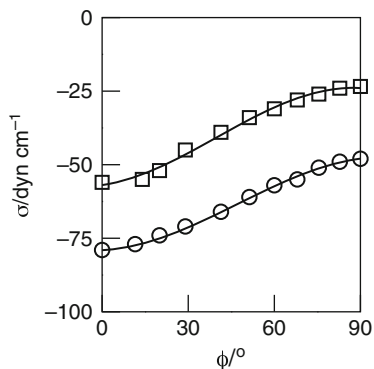


Fig. 7 Wetting free energies inside planar confinements of widths 1.64 (*upper*) and 2.7 nm (*lower curve*) as functions of the angle between the field and the walls. Field strength: $E_o = 0.2 \text{ V } \text{\AA}^{-1}$, average actual field $E \sim 0.005 \text{ V } \text{\AA}^{-1}$ (at $\phi = 90^\circ$), wall contact angle 93°

of $O(10^{-2}) \text{ V } \text{\AA}^{-1}$. The comparison between wetting free energies at different pore widths shows that the effect of the field grows with the thickness of the polarized water slab; however, the free energy change with the angle ϕ is essentially *independent* of the pore width. Given the smaller width amounts to about four layers of water molecules, it is clear that the angle dependence is dominated by pure surface-layer effects. The variation of the free energy with the angle produces an aligning torque of (absolute) magnitude

$$\tau(E, \phi) = -(\partial F / \partial \phi) \sim -A_{\text{wl}} \partial \sigma(E, \phi) / \partial \phi \quad (17)$$

where F is the free energy and A_{wl} is the area of solid wall/liquid interface. The torque on nanopore walls arises solely due to anisotropic water/wall interactions and independently of any direct interaction between the wall material and the field. Importantly, because a major part of angular forces on water molecules reflects orientational preferences of hydrogen-bonding, a nonzero torque can exist even when there is no dielectric contrast between the pore material and water.

4.3.2 Dispersed Nanoparticles

The orienting effect, discussed above, is present in other geometries including nonspherical nanoparticles in a dispersion. As orientational forces between the solvent molecules and the particle surfaces couple with those imposed by the applied field, they augment the classical effect (16), and enhance the trend to align the nanoparticle surface with the field. This expectation, based on surface thermodynamics calculations for aqueous confinements [53], is confirmed in molecular dynamics simulations of freely rotating nano-platelets suspended in water under external field [108]. Orientational forces are found to *exceed* continuum theory predictions [129] by a factor close to two [108], providing a direct measure of the molecular mechanism neglected in macroscopic theories. Enhanced torques can considerably facilitate the use of electric field in tuning suspension structure and thus, for a supersaturated regime, also the structure of any emerging crystalline phase. For materials science as well as for the design of electro-mechanical sensors, it is essential to estimate also the *dynamics* of nanoparticle orientation. Reorientation time of a 2–3 nm wide graphene-like platelet under the (actual) field of $\sim 0.03 \text{ V } \text{\AA}^{-1}$ is of $O(10^2)$ ps. A very interesting result is an approximate balancing between increased hydrodynamic friction and the electric torque upon particle size scaling. The field-induced reorientation dynamics therefore depends only weakly on the particle size and remains fast $O(10^2\text{--}10^3)$ ps even for comparatively big $O(10)$ nm particles; these results can be extrapolated to even bigger sizes not accessible by molecular dynamics simulations with explicit solvent. Apart from the torque enhancement due to hydration-shell molecules, the observed dynamic behavior conforms well to predictions [129] from classical hydrodynamics.

5 Conclusions

We have reviewed computational studies of neat water under combined effect of confinement and electric field. Molecular simulations in these systems were able to demonstrate remarkable differences between field-enhanced wetting *at the nanoscale* and in macroscopic systems. In particular, they highlighted the coupling between interfacial hydrogen bonds [100–103] and molecular alignment in the electric field. This coupling introduces a dependence of wetting on field direction and polarity in contrast to the conventional picture in macroscopic systems; system size plays a crucial role. The observed anisotropy in field-induced wetting is a new nanoscale phenomenon that has so far been elusive as, in the majority of current experimental setups, surface molecules represent a very low fraction of the total number of molecules affected by the field [130]. It may find applications, for example in the design of electrowetting techniques in fabrication and property tuning nanomaterials. Likewise, these effects may play a role in function of membrane proteins that are voltage sensitive, like pumps, transporters, and channels (Roux B, private communication, [131]), including artificial ones [132, 133].

Another novel mechanism that originates from molecular anisotropy of polar solvent molecules such as water reveals strong field-induced orientational forces acting on apolar surface through water mediation, which operate regardless of the presence or absence of solute/solvent permittivity difference. The findings have applications in nanomaterials engineering, where direct interactions between dipolar nanoparticles and applied electric field have been used to control and explain nanoassembly processes [134]. The new mechanism [53, 108] can be used in a similar way, regardless of the electrostatic nature of a nanoparticle. The water-mediated torques can act in concert with direct electrostatic interactions, and can be of similar magnitude when the particles are of nanosize. The response to the applied field takes place at an attractively short time scale [108]. Therefore, the mechanism can be considered in the development of chemical and biosensors.

The examples presented highlight the importance and predictive power of molecular modeling techniques in providing new insights into microscopic scale phenomena not fully accessible in experiment. The novelty of the results should have a broad impact in very active research fields of nano- and bioengineering, physics at the cell scale [135], etc. The future outlook calls for natural extensions of the work reviewed here: exciting new physics can emerge upon inclusion of aqueous salt solutions [136, 137] and evaluating the dynamic response of nanoconfined water to field change [138], a critical dynamic property for electro-switchable nanofluidic or optical devices.

Acknowledgements We thank Kevin Leung for his contribution to some of the work reviewed here. We gratefully acknowledge the financial support from the National Science Foundation (CHE-0718724) and the U.S. Department of Energy (DE-SC-0004406).

References

1. Berge B, Peseux J (2000) Variable focal lens controlled by an external voltage: an application of electrowetting. *Eur Phys J E* 3(2):159–163
2. Hendriks BHW, Kuiper S, Van As MAJ, Renders CA, Tukker TW (2005) Electrowetting-based variable-focus lens for miniature systems. *Opt Rev* 12(3):255–259
3. Stone HA, Stroock AD, Ajdari A (2004) Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annu Rev Fluid Mech* 36:381–411
4. Squires TM, Quake SR (2005) Microfluidics: fluid physics at the nanoliter scale. *Rev Mod Phys* 77(3):977–1026
5. Cho SK, Moon H (2008) Electrowetting on dielectric (EWOD): new tool for bio/micro fluids handling. *Biochip J* 2(2):79–96
6. Dzubiella J, Allen RJ, Hansen JP (2004) Electric field-controlled water permeation coupled to ion transport through a nanopore. *J Chem Phys* 120(11):5001–5004
7. Dzubiella J, Hansen JP (2005) Electric-field-controlled water and ion permeation of a hydrophobic nanopore. *J Chem Phys* 122(23):234706
8. Rasaiah JC, Garde S, Hummer G (2008) Water in nonpolar confinement: from nanotubes to proteins and beyond. *Annu Rev Phys Chem* 59:713–740
9. Li JY, Gong XJ, Lu HJ, Li D, Fang HP, Zhou RH (2007) Electrostatic gating of a nanometer water channel. *Proc Natl Acad Sci* 104(10):3687–3692
10. Bostick D, Berkowitz ML (2003) The implementation of slab geometry for membrane-channel molecular dynamics simulations. *Biophys J* 85(1):97–107
11. Tieleman DP, Leontiadou H, Mark AE, Marrink SJ (2003) Simulation of pore formation in lipid bilayers by mechanical stress and electric fields. *J Am Chem Soc* 125(21):6382–6383
12. Tarek M (2005) Membrane electroporation: a molecular dynamics simulation. *Biophys J* 88(6):4045–4053
13. Suydam IT, Snow CD, Pande VS, Boxer SG (2006) Electric fields at the active site of an enzyme: direct comparison of experiment with theory. *Science* 313(5784):200–204
14. Bateni A, Susnar SS, Amirfazli A, Neumann AW (2004) Development of a new methodology to study drop shape and surface tension in electric fields. *Langmuir* 20(18):7589–7597
15. Bateni A, Laughton S, Tavana H, Susnar SS, Amirfazli A, Neumann AW (2005) Effect of electric fields on contact angle and surface tension of drops. *J Colloid Interface Sci* 283(1):215–222
16. Shapiro B, Moon H, Garrell RL, Kim CJ (2003) Equilibrium behavior of sessile drops under surface tension, applied external fields, and material variations. *J Appl Phys* 93(9):5794–5811
17. Quinn A, Sedev R, Ralston J (2003) Influence of the electrical double layer in electrowetting. *J Phys Chem B* 107(5):1163–1169
18. Daikhin LI, Kornyshev AA, Urbakh M (1999) The effect of electric field on capillary waves at the interface of two immiscible electrolytes. *Chem Phys Lett* 309(3–4):137–142
19. Holt JK, Park HG, Wang YM, Stadermann M, Artyukhin AB, Grigoropoulos CP, Noy A, Bakajin O (2006) Fast mass transport through sub-2-nanometer carbon nanotubes. *Science* 312(5776):1034–1037
20. Mugele F, Baret JC (2005) Electrowetting: from basics to applications. *J Phys Condens Matter* 17(28):R705–R774
21. Shamai R, Andelman D, Berge B, Hayes R (2008) Water, electricity, and between ... on electrowetting and its applications. *Soft Matter* 4(1):38–45
22. Mugele F (2009) Fundamental challenges in electrowetting: from equilibrium shapes to contact angle saturation and drop dynamics. *Soft Matter* 5(18):3377–3384
23. Mugele F, Duits M, van den Ende D (2009) Electrowetting: a versatile tool for drop manipulation, generation, and characterization. *Adv Colloid Interface Sci* 161(1–2):115–123
24. Bocquet L, Charlaix E (2010) Nanofluidics, from bulk to interfaces. *Chem Soc Rev* 39(3):1073–1095

25. Guan L, Qi GC, Liu S, Zhang H, Zhang Z, Yang YL, Wang C (2009) Nanoscale electrowetting effects studied by atomic force microscopy. *J Phys Chem C* 113(2):661–665
26. Gelb LD, Gubbins KE, Radhakrishnan R, Sliwinski-Bartkowiak M (1999) Phase separation in confined systems. *Rep Prog Phys* 62(12):1573–1659
27. Luzar A, Bratko D, Blum L (1987) Monte-Carlo simulation of hydrophobic interaction. *J Chem Phys* 86(5):2955–2959
28. Lum K, Luzar A (1997) Pathway to surface-induced phase transition of a confined fluid. *Phys Rev E* 56(6):R6283–R6286
29. Nakanishi H, Fisher ME (1983) Critical-point shifts in films. *J Chem Phys* 78(6):3279–3293
30. Evans R (1990) Fluids adsorbed in narrow pores – phase-equilibria and structure. *J Phys Condens Matter* 2(46):8989–9007
31. Schoen M, Diestler DJ (1998) Analytical treatment of a simple fluid adsorbed in a slit-pore. *J Chem Phys* 109(13):5596–5606
32. Klapp SHL, Schoen M (2002) Spontaneous orientational order in confined dipolar fluid films. *J Chem Phys* 117(17):8050–8062
33. Klapp SHL, Schoen M (2004) Ferroelectric states of a dipolar fluid confined to a slit-pore. *J Mol Liq* 109(2):55–61
34. Huang X, Margulis CJ, Berne BJ (2003) Dewetting-induced collapse of hydrophobic particles. *Proc Natl Acad Sci* 100(21):11953–11958
35. Parker JL, Claesson PM, Attard P (1994) Bubbles, cavities, and the long-ranged attraction between hydrophobic surfaces. *J Phys Chem* 98(34):8468–8480
36. Bratko D, Curtis RA, Blanch HW, Prausnitz JM (2001) Interaction between hydrophobic surfaces with metastable intervening liquid. *J Chem Phys* 115(8):3873–3877
37. MacDowell LG (2003) Formal study of nucleation as described by fluctuation theory. *J Chem Phys* 119(1):453–463
38. Israelachvili JN (2011) *Intermolecular and surface forces*, 3rd edn. Academic, London
39. Berne BJ, Weeks JD, Zhou RH (2009) Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem* 60:85–103
40. Evans R, Parry AO (1990) Liquids at interfaces – what can a theorist contribute. *J Phys Condens Matter* 2:SA15–SA32
41. Leung K, Luzar A, Bratko D (2003) Dynamics of capillary drying in water. *Phys Rev Lett* 90:065502
42. Koishi T, Yasuoka K, Ebisuzaki T, Yoo S, Zeng XC (2005) Large-scale molecular-dynamics simulation of nanoscale hydrophobic interaction and nanobubble formation. *J Chem Phys* 123(20):204707
43. Giovambattista N, Rossky PJ, Debenedetti PG (2006) Effect of pressure on the phase behavior and structure of water confined between nanoscale hydrophobic and hydrophilic plates. *Phys Rev E* 73(4):041604
44. Leung K, Luzar A (2000) Dynamics of capillary evaporation. II. Free energy barriers. *J Chem Phys* 113(14):5845–5852
45. Luzar A (2004) Activation barrier scaling for the spontaneous evaporation of confined water. *J Phys Chem B* 108(51):19859–19866
46. Luzar A, Leung K (2000) Dynamics of capillary evaporation. I. Effect of morphology of hydrophobic surfaces. *J Chem Phys* 113(14):5836–5844
47. Lum K, Chandler D (1998) Phase diagram and free energies of vapor films and tubes for a confined fluid. *Int J Thermophys* 19(3):845–855
48. Yushchenko VS, Yaminsky VV, Shchukin ED (1983) Interaction between particles in a nonwetting liquid. *J Colloid Interface Sci* 96(2):307–314
49. Yaminsky VV (2000) Molecular mechanisms of hydrophobic transitions. *J Adhes Sci Technol* 14(2):187–233
50. Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG (2008) Hydrophobicity of protein surfaces: separating geometry from chemistry. *Proc Natl Acad Sci* 105(7):2274–2279

51. Huang DM, Chandler D (2002) The hydrophobic effect and the influence of solute-solvent attractions. *J Phys Chem B* 106(8):2047–2053
52. Acharya H, Ranganathan S, Jamadagni SN, Garde S (2010) Mapping hydrophobicity at the nanoscale: applications to heterogeneous surfaces and proteins. *Faraday Discuss* 146:353
53. Bratko D, Daub CD, Luzar A (2009) Water-mediated ordering of nanoparticles in an electric field. *Faraday Discuss* 141:55–66
54. Sarupria S, Garde S (2009) Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins. *Phys Rev Lett* 103(3):037803
55. Lipmann G (1875) Relations entre les phenomenes electriques et capillaires. *Ann Chim Phys* 5:494
56. Buehrle J, Herminghaus S, Mugele F (2003) Interface profiles near three-phase contact lines in electric fields. *Phys Rev Lett* 91(8):086101
57. Vaitheeswaran S, Yin H, Rasaiah JC (2005) Water between plates in the presence of an electric field in an open system. *J Phys Chem B* 109(14):6629–6635
58. Frank HS (2023) *J Chem Phys* 155:23
59. Mugele F, Klingner A, Buehrle J, Steinhauser D, Herminghaus S (2005) Electrowetting: a convenient way to switchable wettability patterns. *J Phys Condens Matter* 17(9):S559–S576
60. Krupenkin T, Taylor JA, Kolodner P, Hodes M (2005) Electrically tunable superhydrophobic nanostructured surfaces. *Bell Labs Tech J* 10(3):161–170
61. Vaitheeswaran S, Rasaiah JC, Hummer G (2004) Electric field and temperature effects on water in the narrow nonpolar pores of carbon nanotubes. *J Chem Phys* 121(16):7955–7965
62. Dzubiella J, Hansen JP (2003) Reduction of the hydrophobic attraction between charged solutes in water. *J Chem Phys* 119(23):12049–12052
63. Dzubiella J, Hansen JP (2004) Competition of hydrophobic and Coulombic interactions between nanosized solutes. *J Chem Phys* 121(11):5514–5530
64. Brunet C, Malherbe JG, Amokrane S (2009) Controlling the composition of a confined fluid by an electric field. *J Chem Phys* 131(22):221103
65. Brunet C, Malherbe JG, Amokrane S (2010) Binary mixture adsorbed in a slit pore: field-induced population inversion near the bulk instability. *Phys Rev E* 82(2):021564
66. Bratko D, Daub CD, Leung K, Luzar A (2007) Effect of field direction on electrowetting in a nanopore. *J Am Chem Soc* 129(9):2504–2510
67. Maerzke KA, Siepmann JI (2010) Effects of an applied electric field on the vapor-liquid equilibria of water, methanol, and dimethyl ether. *J Phys Chem B* 114(12):4261–4270
68. Bratko D, Daub CD, Luzar A (2008) Field-exposed water in a nanopore: liquid or vapour? *Phys Chem Chem Phys* 10(45):6807–6813
69. England JL, Park S, Pande VS (2008) Theory for an order-driven disruption of the liquid state in water. *J Chem Phys* 128(4):044503
70. Landau LD, Lifshitz EM, Pitaevskii LP (1948) *Electrodynamics of continuum media*. Pergamon, Oxford
71. Høye JS, Stell G (1980) Statistical-mechanics of polar fluids in electric fields. *J Chem Phys* 72:1597
72. Frenkel D, Smit B (2002) *Understanding molecular simulation, from algorithms to applications*. Academic, San Diego
73. Allen MP, Tildesley DJ (1987) *Computer simulation of liquids*. Oxford University Press, New York
74. Frolich H (1990) *Theory of dielectrics*, 2nd edn. Oxford University Press, Oxford
75. Siepmann JI, Sprik M (1995) Influence of surface-topology and electrostatic potential on water electrode systems. *J Chem Phys* 102(1):511–524
76. Reed SK, Lanning OJ, Madden PA (2007) Electrochemical interface between an ionic liquid and a model metallic electrode. *J Chem Phys* 126(8):084704
77. Yeh IC, Berkowitz ML (1999) Ewald summation for systems with slab geometry. *J Chem Phys* 111(7):3155–3162

78. Yeh IC, Hummer G (2004) Diffusion and electrophoretic mobility of single-stranded RNA from molecular dynamics simulations. *Biophys J* 86(2):681–689
79. Sutmann G (1998) Structure formation and dynamics of water in strong external electric fields. *J Electroanal Chem* 450(2):289–302
80. Jia R, Hentschke R (2009) Dipolar particles in an external field: molecular dynamics simulation and mean field theory. *Phys Rev E* 80(5):051502
81. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91(24):6269–6271
82. England JL, Pande VS (2010) Charge, hydrophobicity, and confined water: putting past simulations into a simple theoretical framework. *Biochem Cell Biol* 88(2):359–369
83. Wagner K, Keyes E, Kephart TW, Edwards G (1997) Analytical Debye-Huckel model for electrostatic potentials around dissolved DNA. *Biophys J* 73(1):21–30
84. Polk C (2000) Biological applications of large electric fields: some history and fundamentals. *IEEE Trans Plasma Sci* 28(1):6–14
85. Siu SWI, Bockmann RA (2007) Electric field effects on membranes: Gramicidin A as a test ground. *J Struct Biol* 157(3):545–556
86. Schweighofer KJ, Xia XF, Berkowitz ML (1996) Molecular dynamics study of water next to electrified Ag(111) surfaces. *Langmuir* 12(16):3747–3752
87. Kornyshev AA, Lee DJ, Leikin S, Wynveen A (2007) Structure and interactions of biological helices. *Rev Mod Phys* 79(3):943–996
88. Schweighofer KJ, Benjamin I (1995) Electric-field effects on the structure and dynamics at a liquid/liquid interface. *J Electroanal Chem* 391(1–2):1–10
89. Bratko D, Dolar D (1984) Ellipsoidal model of poly-electrolyte solutions. *J Chem Phys* 80(11):5782–5789
90. Luzar A, Bratko D (1990) Electric double-layer interactions in reverse micellar systems – a Monte-Carlo simulation study. *J Chem Phys* 92(1):642–648
91. Bratko D, Woodward CE, Luzar A (1991) Charge fluctuation in reverse micelles. *J Chem Phys* 95(7):5318–5326
92. Wu JZ, Bratko D, Prausnitz JM (1998) Interaction between like-charged colloidal spheres in electrolyte solutions. *Proc Natl Acad Sci* 95(26):15169–15172
93. Wu JZ, Bratko D, Blanch HW, Prausnitz JM (1999) Monte Carlo simulation for the potential of mean force between ionic colloids in solutions of asymmetric salts. *J Chem Phys* 111(15):7084–7094
94. Svishchev IM, Kusalik PG (1994) Crystallization of liquid water in a molecular-dynamics simulation. *Phys Rev Lett* 73(7):975–978
95. Zangi R, Mark AE (2004) Electrofreezing of confined water. *J Chem Phys* 120(15):7123–7130
96. Suresh SJ, Satish AV, Choudhary A (2006) Influence of electric field on the hydrogen bond network of water. *J Chem Phys* 124(7):074506
97. Suresh SJ (2007) Disruption of hydrogen bond structure of water near charged electrode surfaces. *J Chem Phys* 126(20):204705
98. Kiselev M, Heinzinger K (1996) Molecular dynamics simulation of a chloride ion in water under the influence of an external electric field. *J Chem Phys* 105(2):650–657
99. Rodgers JM, Weeks JD (2008) Local molecular field theory for the treatment of electrostatics. *J Phys Condens Matter* 20:494206
100. Luzar A, Svetina S, Zeks B (1983) The contribution of hydrogen-bonds to the surface-tension of water. *Chem Phys Lett* 96(4):485–490
101. Luzar A, Svetina S, Zeks B (1985) Consideration of the spontaneous polarization of water at the solid liquid interface. *J Chem Phys* 82(11):5146–5154
102. Lee CY, McCammon JA, Rossky PJ (1984) The structure of liquid water at an extended hydrophobic surface. *J Chem Phys* 80(9):4448–4455
103. Du Q, Freysz E, Shen YR (1994) Surface vibrational spectroscopic studies of hydrogen-bonding and hydrophobicity. *Science* 264(5160):826–828

104. Daub CD, Bratko D, Leung K, Luzar A (2007) Electrowetting at the nanoscale. *J Phys Chem C* 111(2):505–509
105. Grzelak EM, Errington JR (2010) Nanoscale limit to the applicability of Wenzel's equation. *Langmuir* 26(16):13297–13304
106. Grzelak EM, Errington JR (2008) Computation of interfacial properties via grand canonical transition matrix Monte Carlo simulation. *J Chem Phys* 128(1):014710
107. Hu GH, Xu AJ, Xu Z, Zhou ZW (2008) Dewetting of nanometer thin films under an electric field. *Phys Fluids* 20(10):102101
108. Daub CD, Bratko D, Ali T, Luzar A (2009) Microscopic dynamics of the orientation of a hydrated nanoparticle in an electric field. *Phys Rev Lett* 103:207801
109. Cramer T, Zerbetto F, Garcia R (2008) Molecular mechanism of water bridge buildup: field-induced formation of nanoscale menisci. *Langmuir* 24(12):6116–6120
110. Cheung DL (2010) Molecular simulation of nanoparticle diffusion at fluid interfaces. *Chem Phys Lett* 495(1–3):55–59
111. Eppenga R, Frenkel D (1984) Monte-Carlo study of the isotropic and nematic phases of infinitely thin hard platelets. *Mol Phys* 52(6):1303–1334
112. Bratko D, Jonsson B, Wennerstrom H (1986) Electrical double-layer interactions with image charges. *Chem Phys Lett* 128(5–6):449–454
113. Harismiadis VI, Vorholz J, Panagiotopoulos AZ (1996) Efficient pressure estimation in molecular simulations without evaluating the virial. *J Chem Phys* 105(18):8469–8470
114. Gloor GJ, Jackson G, Blas FJ, de Miguel E (2005) Test-area simulation method for the direct determination of the interfacial tension of systems with continuous or discontinuous potentials. *J Chem Phys* 123(13):134703
115. de Miguel E, Jackson G (2006) The nature of the calculation of the pressure in molecular simulations of continuous models from volume perturbations. *J Chem Phys* 125(16):164109
116. Brumby PE, Haslam AJ, de Miguel E, Jackson G (2011) Subtleties in the calculation of the pressure and pressure tensor of anisotropic particles from volume-perturbation methods and the apparent asymmetry of the compressive and expansive contributions. *Mol Phys* 109:169–189
117. Wang J, Bratko D, Luzar A (2011) Probing surface tension additivity on chemically heterogeneous surfaces: a molecular approach. *Proc Natl Acad Sci USA* 108(16):6374–6379
118. Leroy F, dos Santos D, Muller-Plathe F (2009) Interfacial excess free energies of solid-liquid interfaces by molecular dynamics simulation and thermodynamic integration. *Macromol Rapid Commun* 30(9–10):864–870
119. Bratko D (2010) General discussion. Fig. 2. *Faraday Discuss* 146:367–393
120. Giovambattista N, Debenedetti PG, Rossky PJ (2009) Enhanced surface hydrophobicity by coupling of surface polarity and topography. *Proc Natl Acad Sci* 106(36):15181–15185
121. Giovambattista N, Debenedetti PG, Rossky PJ (2007) Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity. *J Phys Chem C* 111(3):1323–1332
122. Zhang XY, Zhu YX, Granick S (2002) Hydrophobicity at a Janus interface. *Science* 295(5555):663–666
123. Park J, Lu W (2007) Orientation of core-shell nanoparticles in an electric field. *Appl Phys Lett* 91(5):053113
124. Zhang X, Zhang ZL, Glotzer SC (2007) Simulation study of dipole-induced self-assembly of nanocubes. *J Phys Chem C* 111(11):4132–4137
125. Stratton JA (1941) *Electromagnetic theory*. McGraw-Hill, New York
126. Tsori Y (2009) Colloquium: phase transitions in polymers and liquids in electric fields. *Rev Mod Phys* 81(4):1471–1494
127. Froltsov VA, Klapp SHL (2007) Dielectric response of polar liquids in narrow slit pores. *J Chem Phys* 126(11):114703
128. Gramzow M, Klapp SHL (2007) Capillary condensation and orientational ordering of confined polar fluids. *Phys Rev E* 75(1):011605

129. Lyklema J (1991) Fundamentals in interface and colloid science, vol I: Fundamentals. Academic, London
130. Lu WY, Kim T, Han AJ, Chen X, Qiao Y (2009) Electrowetting effect in a nanoporous silica. *Langmuir* 25(16):9463–9466
131. Anishkin A, Akitake B, Kamaraju K, Chiang CS, Sukharev S (2009) Hydration properties of mechanosensitive channel pores define the energetics of gating. *J Phys Condens Matter* 22(45):4120
132. Garate JA, English NJ, MacElroy JMD (2009) Carbon nanotube assisted water self-diffusion across lipid membranes in the absence and presence of electric fields. *Mol Simul* 35(1–2): 3–12
133. Garate JA, English NJ, MacElroy JMD (2009) Static and alternating electric field and distance-dependent effects on carbon nanotube-assisted water self-diffusion across lipid membranes. *J Chem Phys* 131(11):114508
134. Grzelczak M, Vermant J, Furst EM, Liz-Marzan LM (2010) Directed self-assembly of nanoparticles. *ACS Nano* 4(7):3591–3605
135. Sun S, Wong JTY, Zhang TY (2011) Molecular dynamics simulations of phase transition of lamellar lipid membrane in water under an electric field. *Soft Matter* 7(1):147–152
136. Girault HH (2006) Electrowetting: shake, rattle and roll. *Nat Mater* 5(11):851–852
137. Daub CD, Bratko D, Luzar A. Electric control of wetting by salty nanodrops: molecular dynamics simulations (submitted)
138. von Domaros M, Wang J, Bratko D, Kirchner B, Luzar A. Dynamics at a Janus interface (work in progress)

Molecular Simulations of Retention in Chromatographic Systems: Use of Biased Monte Carlo Techniques to Access Multiple Time and Length Scales

Jake L. Rafferty, J. Ilja Siepmann, and Mark R. Schure

Abstract The use of configurational-bias Monte Carlo simulations in the Gibbs ensemble allows for the sampling of phenomena that occur on vastly different time and length scales. In this review, applications of this simulation approach to probe retention in gas and reversed-phase liquid chromatographic systems are discussed. These simulations provide an unprecedented view of the retention processes at the molecular-level and show excellent agreement with experimental retention data.

Keywords Gas chromatography · Molecular simulation · Monte Carlo · Retention mechanism · Reversed-phase liquid chromatography

Contents

1	Introduction	182
2	Multiscale Monte Carlo Techniques	183
3	Gas Chromatography	185

J.L. Rafferty

Department of Chemistry and Chemical Theory Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455-0431, USA

J.I. Siepmann (✉)

Departments of Chemistry and of Chemical Engineering and Materials Science and Chemical Theory Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455-0431, USA

e-mail: Siepmann@umn.edu

M.R. Schure

Theoretical Separation Science Laboratory, The Dow Chemical Company, 727 Norristown Road, Box 0904, Spring House, PA 19477-0904, USA

e-mail: MSchure@Dow.com

4	Reversed-Phase Liquid Chromatography	187
4.1	Mobile Phase Solvent Effects and Small Molecule Retention Mechanisms	188
4.2	Shape Selectivity in RPLC	196
5	Conclusions	198
	References	199

1 Introduction

A multiscale method refers to a computational technique that spans multiple levels of length (resolution) and/or timescales. Most commonly, the term refers to multiple levels of theory being utilized in an energy or force calculation (e.g., a small part of the system is treated using expensive electronic structure calculations, whereas less expensive molecular mechanics force fields are used for the surrounding) or the presence of multiple timesteps in a molecular dynamics simulation where forces arising from short-range interactions are calculated more frequently than those arising from long-range interactions. The term “multiscale” is also often used to describe a set of separate calculations or simulations where one level provides information used in the next level.

The main goal in a molecular simulation is to sample the important regions of the statistical-mechanical phase space, i.e., those configurations that contribute most to the thermodynamic averages for the system. Sampling inefficiencies occur when these important regions of phase space are separated by free energy barriers, which manifest themselves in long relaxation times. In a Monte Carlo simulation, special (sometimes termed “unphysical” or biased) moves can be deployed that allow one to hop over these kinetic barriers. As multiple different move types can be employed in a Monte Carlo simulation, multiple timescales can also be accessed, and the most efficient computation of equilibrium properties is achieved when the vastly different timescale for different types of motion (e.g., bond stretching vs transfer of a molecule over large distances) are merged into a common computational timescale. The Monte Carlo technique also allows for the use of open ensembles where multiple phases (each handled in a separate simulation box) are thermodynamically connected via special Monte Carlo moves but do not share a direct interface. These phases may also represent different length scales, e.g., a phase confined in a nanopore and a bulk phase.

In this review we discuss the application of multiscale Monte Carlo simulations to explore various chromatographic systems. Chromatography is a collective term for a set of techniques used for the separation of chemical mixtures. The common theme in these techniques is that the mixture to be separated is dissolved in a mobile phase which is passed through a stationary phase usually placed in an elongated column. The separation of the mixture is based upon the differential partitioning of the components in the mixture between the two phases. When a molecule has a higher affinity for the stationary phase, it is retained in this phase longer and can be separated from the less retained components that spend a relatively longer time in the mobile phase. In gas chromatography (GC), the mobile phase is an inert gas and

the stationary phase is either a solid or a liquid film adhering to the inside of a capillary tube. In reversed-phase liquid chromatography (RPLC) the mobile phase is an aqueous/organic solvent and the stationary phase is most often composed of alkyl chains chemically bonded to the surfaces of highly porous silica particles.

2 Multiscale Monte Carlo Techniques

To enable the simulation of chromatographic systems, a combination of two Monte Carlo techniques that allows for the sampling of processes across different length and timescales is used. These are the Gibbs ensemble Monte Carlo (GEMC) and the configurational-bias Monte Carlo (CBMC) methods. The Gibbs ensemble [1, 2] involves the concurrent simulation of multiple phases that are thermodynamically coupled through special Monte Carlo moves, but do not require consideration of explicit interfaces between these phases. One of the special GEMC moves allows for direct particle transfers (as opposed to slow diffusive processes) between two or more separate phases, i.e., the molecule may transfer from one bulk phase to another bulk phase without having to encounter an interface. Nevertheless, the GEMC technique can also be used to connect phases that exhibit different characteristic length scales (e.g., a nanoporous sorptive phase and a bulk solution phase) and may involve explicit interfaces. Such a set-up is used in our Monte Carlo simulation of RPLC systems. As illustrated in Fig. 1, these GEMC simulations use a three-box setup. The first simulation box contains the stationary phase and is elongated in the z -direction with $L_x = 20.0$, $L_y = 26.0$, and $L_z = 90.0$ Å. The center of this box contains a five layer slab of β -cristobalite with its two (1 1 1) surfaces exposed. To these surfaces, octadecylsilane chains (ODS) are grafted. In contact with this stationary phase, and connected through the periodic boundaries used in the simulations, is the mobile phase solvent sorbed in the pore. This box is periodically replicated in all three dimensions. Thus, the setup in this first box

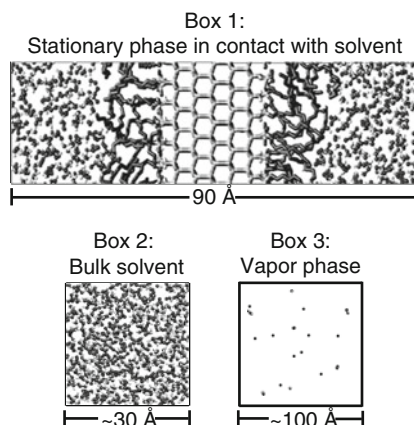


Fig. 1 Three-box Gibbs ensemble simulation setup

corresponds to a planar slit pore with a spacing of about 70 Å. The second box contains a bulk solvent reservoir and the third box a helium vapor phase (used as a reference phase to deconstruct free energies of transfer into stationary and mobile phase contributions). These two boxes are cubic and their volumes are allowed to fluctuate in response to an external pressure bath (whereas the volume of the slit pore is fixed).

The three simulation boxes are in thermodynamic contact through the exchange of solvent and analyte molecules via CBMC particle transfer [3, 4] and identity exchange moves [5, 6] (i.e., this move converts a molecule of type *A* into a molecule of type *B* in one simulation box and concurrently converts *B* to *A* in another box) that ensure that the chemical potential of these species is the same in all three boxes. Exchange of solvent molecules is an extremely important aspect of the simulation setup because there is no way to know, a priori, the correct amount of solvent molecules to be placed in the stationary-phase box with its hydrophobic surfaces in order to represent a given chromatographic condition (composition of the mobile phase, temperature, and pressure). In addition, these particle transfer and exchange moves allow for a much more efficient sampling of the spatial distribution of the solvent and analyte molecules than could be achieved by simple translational moves that mimic diffusive behavior. Precise information on the spatial distribution allows for the calculation of distribution coefficients and free energies of transfer between the mobile and stationary phase (the free energy of retention) from the ratio of average analyte number densities in each phase:

$$\Delta G_{a \rightarrow b} = -RT \ln K = -RT \ln \frac{\rho_b}{\rho_a}, \quad (1)$$

where ρ_i is the number density of the species in phase *i*. The distribution coefficients and free energies computed from the simulations can be directly compared with experimental retention data to validate simulation results.

The coupled-decoupled CBMC method [3, 4, 7] selects trial configurations for a stepwise growth of the trial molecule using a biased preselection process to find favorable regions of phase space and, thereby, greatly increases the acceptance rates for particle transfer and identity exchange moves. In addition, it allows for the sampling of the conformational degrees of freedom of articulated molecules via regrowing a part of the molecule that includes either one or multiple terminal segments [3] or only interior segments [8]. The CBMC technique can also be extended to utilize multiple energy scales. To enhance computational efficiency for simulations using molecular mechanics force fields, the biased preselection of configurations first generates a set of growth directions based on the bonded interactions (e.g., bond stretching and angle bending) for which the cost of the energy calculation depends only on the trial molecule and is independent of system size. This is followed by a selection of a specific growth direction based on a less expensive approximation of the nonbonded interactions (shorter cutoff and only direct-space part of the Ewald sum), while the final acceptance of the move is based on the full potential including a correction for the more expensive part of the

nonbonded interactions (longer cutoff, reciprocal-space part of the Ewald sum) [3, 9]. In another application, the CBMC technique (with preselection of moves based on an empirical potential) has also been used to enable the simulation of phase equilibria using a first principles description of the interaction system [10].

The main advantage of the GEMC/CBMC approach in the study of chromatography is the multiple time and length scales that can be accessed. In chromatography, observing the distribution of analyte molecules between the mobile and stationary phases is of utmost importance. The time scale of this event, which depends on the diffusion of molecules over large distances, is simply inaccessible in a traditional molecular dynamics simulation that is limited to a few nanoseconds. Furthermore, events such as the re-equilibration of the RPLC stationary phase after switching solvents, which has been shown to take tens of minutes in the laboratory [11–13], can be observed in a GEMC/CBMC simulation.

3 Gas Chromatography

In gas chromatography, the mobile phase is a low-density gas. It is further divided into gas–solid chromatography (GSC) and gas–liquid chromatography (GLC) according to the nature of the stationary phase. The former is sometimes referred to as gas adsorption chromatography and the latter as gas–liquid partition chromatography, indicating the thermodynamic processes that are the main driving forces for the retention processes. The liquid phase in GLC is only one part of the stationary phase which is coated (bonded) to a solid support material such as fused silica. Thus adsorption at the gas–liquid and liquid–solid interfaces (and, if the phase loading is low or if the liquid does not wet the support, also at the gas/solid interface) can contribute to retention [14].

The Kovats retention index [15] has proved to be one of the most useful concepts in GLC, allows for direct comparison of experimental and simulated retention data, and can be expressed either in terms of specific retention volumes or partition coefficients:

$$I_x = 100n + 100 \left[\frac{\log(V'_x/V'_n)}{\log(V'_{n+1}/V'_n)} \right] = 100n + 100 \left[\frac{\log(K_x/K_n)}{\log(K_{n+1}/K_n)} \right], \quad (2)$$

where the subscripts x , n , and $n + 1$ represent the analyte of interest, the highest normal alkane (having n carbon atoms) that elutes before and the lowest normal alkane that elutes after the analyte, respectively. Use of the Kovats retention index has many advantages: (1) if interfacial adsorption can be neglected then I should be independent of phase loading; (2) I is far less temperature dependent than V'_x or K_x ; (3) in any homologous series (at least for higher homologs) I should increase by 100 per methylene group added; and (4) experimentally measured I values are extremely reproducible.

Figure 2 illustrates the ability of GEMC/CBMC simulations using the TraPPE force field [3, 16–18] to predict (with some exceptions) rather reliable Kovats indices for various combinations of analytes and stationary phases. It should be noted here that the GEMC/CBMC methodology affords computations of Kovats indices of multiple analyte molecules in the same simulation box with a statistical uncertainty (standard error of the mean) of less than 10 Kovats units, an impressive achievement when one considers that the computation of a Kovats index requires partition constants (or transfer free energies) of three analytes. The retention order and Kovats indices of six branched alkanes (2-methylpentane, 3-methylpentane, 2,2-dimethylpentane, 3-ethylpentane, 2,5-dimethylhexane, and 3,4-dimethylhexane in order of increasing I) were correctly reproduced [19], i.e., the simulations are able to pick up differences caused by displacing a methyl branch along the backbone or by replacing two methyl branches with a single ethyl branch.

Benzene and toluene are more retained in a squalane phase than n -hexane and n -heptane, respectively. This separation involves a subtle balance of enthalpic contributions (larger polarizability of arenes, but lack of significant quadrupole–multipole interactions for arene/squalane pairs) and entropic contributions (smaller partial molar volume for benzene vs n -hexane). The data in Fig. 2 illustrate that the TraPPE force field with explicit quadrupole for arenes yields accurate predictions for the Kovats retention indices of benzene and toluene at 323 and 383 K where the temperature increase results in an increase of the I values by about 20 Kovats units [20]. Figure 2 also depicts the I values for benzene, toluene, and the three xylene isomers in a polyethylene oxide phase. In this case, the quadrupole–dipole interactions result in stronger retention for arenes with I values close to those of n -alkanes with four additional carbon atoms [21]. Although the simulations predict the correct retention order for alcohols (methanol, 2-propanol, ethanol, 1-propanol, and 1-butanol in order of increasing I values), it consistently overpredicts the I values by about 100 Kovats units; this is an indication that the TraPPE force field overestimates the strength of hydrogen bonds between alcohols and the polyether [21].

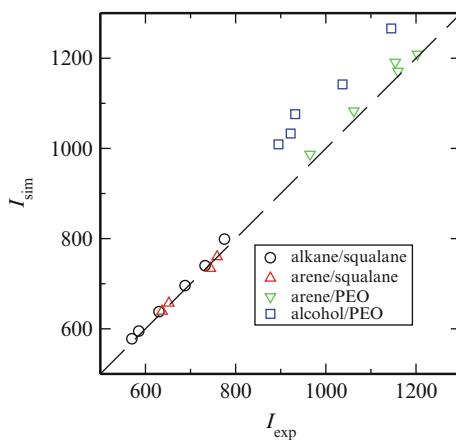


Fig. 2 Predicted Kovats retention indices for various analytes in gas–liquid chromatography with squalane and polyethylene oxide stationary phases. See text for description of the analyte molecules

GEMC/CBMC simulations were also successfully applied to investigate the contribution of adsorption at the mobile/stationary phase interface (here the stationary phase is represented as a two-dimensional film) [22] and the influence of analyte overloading (here the analyte concentration is sufficiently high to result in departures from Henry's law behavior) [23] in gas-liquid chromatography.

4 Reversed-Phase Liquid Chromatography

RPLC is similar in principle to GLC in that both techniques rely on the differing affinities of analyte molecules for the mobile and stationary phases to enact a separation. The distinct difference between RPLC and GC is complexity. In RPLC, the stationary phase typically consists of dimethyl octadecyl silane chains grafted to the surfaces of highly porous silica particles and the mobile phase is a binary solvent containing water and an organic modifier, most commonly methanol or acetonitrile. The complex interplay between the stationary phase, solvent, and analyte has made a molecular-level description of RPLC extremely difficult.

Before focusing on the application of Monte Carlo techniques to RPLC systems, we would like to highlight some important contributions that employed the molecular dynamics approach. In 1994, Schure investigated the bonded-chain conformation and solvent structure in an alkylsilane RPLC system [24]. A few years later, Klatte and Beck [25] and Slusher and Mountain [26] provided quantitative data on the transfer of a methane solute from the mobile phase to the stationary phase in model RPLC systems. More recently, Zhao and Cann [27] investigated chiral RPLC phases and obtained qualitative data on the retention of ten analytes, Fouqueau et al. [28] investigated the adsorption of acridine orange at the mobile/stationary-phase interface in an RPLC system, and Melnikov et al. [29] probed the influence of residual (protonated and deprotonated) silanol groups on solvent and ion distribution in an RPLC system.

Over the past 6 years we have used the GEMC/CBMC methodology to provide much needed molecular-level insight into the stationary-phase structure and the retention mechanism in various RPLC systems. For alkylsilane stationary phases we have investigated the effects of mobile phase composition for water/methanol and water/acetonitrile mixtures [30–34], of alkylsilane coverage [35, 36], of alkylsilane chain length [37, 38], and pressure and pore shape [37]. In addition, the effects of embedded polar groups (ether and amide) have been explored [39]. To illustrate some of the insights that can be gleaned from the GEMC/CBMC simulations, this review will focus on mobile-phase effects for water/acetonitrile mixture on bonded-phase structure and retention of small analyte molecules (a more complete discussion can be found in [34]), and briefly mention studies involving the retention of large polycyclic aromatic hydrocarbon (a more complete discussion can be found in [40]).

4.1 *Mobile Phase Solvent Effects and Small Molecule Retention Mechanisms*

For decades, RPLC has been one of the most widely used techniques for the separation and analysis of chemical mixtures, but there remains significant debate about the molecular level details of numerous aspects of the RPLC process. For example, it has been observed that a dramatic loss of retention can occur when the concentration of water in the mobile phase exceeds a certain threshold [12, 41]. One explanation for this phenomenon is that the alkyl chains of the stationary phase collapse in the presence of highly aqueous solvents [41–43]. A more recent, competing explanation for this retention loss is that a highly aqueous mobile phase does not enter a substantial fraction of the smaller pores in the silica particles due to its higher surface tension and hence is not able to bring the analytes in contact with the stationary phase chains [12]. With regard to solvent effects, another well known phenomenon in RPLC is the tendency of the organic component of the mobile phase to solvate preferentially the stationary phase. However, it is not fully resolved if this preferential solvation occurs solely through the formation of an organic layer atop the stationary phase [44, 45] or if penetration of the organic modifier into the stationary phase is also important [46, 47].

Perhaps the biggest and most longstanding debate in RPLC is on the retention mechanism. Here there are conflicting views as to whether analyte molecules adsorb at the stationary phase/mobile phase interface or fully partition into the stationary phase and to what extent various chromatographic parameters, such as mobile phase composition, affect this mechanism [48–51]. Even if partitioning is taken to be the dominant mechanism of retention, it is not clear if the process can be modeled accurately by bulk liquid–liquid (e.g., oil–water) partitioning [48, 52] or if partitioning into the constrained hydrocarbon chains of the RPLC stationary phase would involve a different molecular mechanism [53, 54]. Furthermore, it is disputed whether the thermodynamic driving forces for analyte retention (transfer from mobile to stationary phase) are primarily from solvophobic interactions with the mobile phase [50] or lipophilic interactions with the stationary phase [48, 52].

To illustrate the effects of mobile phase composition, results for four different water/acetonitrile mobile phase compositions are discussed here: (1) pure water, (2) 33% molfraction acetonitrile, (3) 67% molfraction acetonitrile, and (4) pure acetonitrile, (hereafter referred to as systems WAT, 33A, 67A, and ACN, respectively; the results for system WAT are taken from [32] and those for the other three compositions from [34]). Each system contained 1,200 solvent molecules and 16 analytes (2 each of C1 to C4 normal alkanes and alcohols) and utilized a stationary phase with a surface coverage of $2.9 \mu\text{mol}/\text{m}^2$ (9 ODS chains on each surface) which resulted in a residual silanol density of $4.8 \mu\text{mol}/\text{m}^2$ (15 silanols on each surface). The temperature and pressure used in this study were 323 K and 1 atm, respectively. For each solvent system, four independent simulations were carried out. Each simulation was equilibrated for 2×10^5 Monte Carlo (MC) cycles (one MC cycle corresponds to N MC moves, where N is the total number of molecules in

the system). Thereafter, the simulations proceeded with an additional 2×10^5 MC cycles during which averages were collected.

Snapshots from the simulations at each solvent composition are shown in Fig. 3. Although these snapshots represent only a single configuration of the millions generated during the simulation, they convey a wealth of information about these systems. First, it is observed that as acetonitrile concentration is increased there is extension of the alkyl chains and greater amount of solvent sorbed into the stationary phase. Furthermore, it appears that in the binary solvent systems most of the solvent within the chain region is the organic modifier and not water. Also appearing enriched in the organic modifier is the interfacial region between the alkyl chains and the solvent.

These effects mentioned above can be quantified by examining the ensemble averaged density profiles, also shown in Fig. 3. These profiles show the density of each component of the system as a function of z , or the distance from the silica surface. As indicated by the solvent densities in the region $z = 5\text{--}15$ Å, there is essentially no solvent within bonded chains for system WAT. However, as the acetonitrile concentration is increased there is a dramatic increase in solvent penetration. For the binary solvents, most solvent in the alkyl chain region appears to be the organic modifier and very little water is present.

In all systems examined, there are peaks in the solvent density around $z = 3\text{--}5$ Å. These peaks are the result of solvent hydrogen bonded to the residual silanols present on the silica surface [31, 34]. There is a large decrease in the peaks as the fraction of acetonitrile in the solvent is increased. It appears that water shows a strong

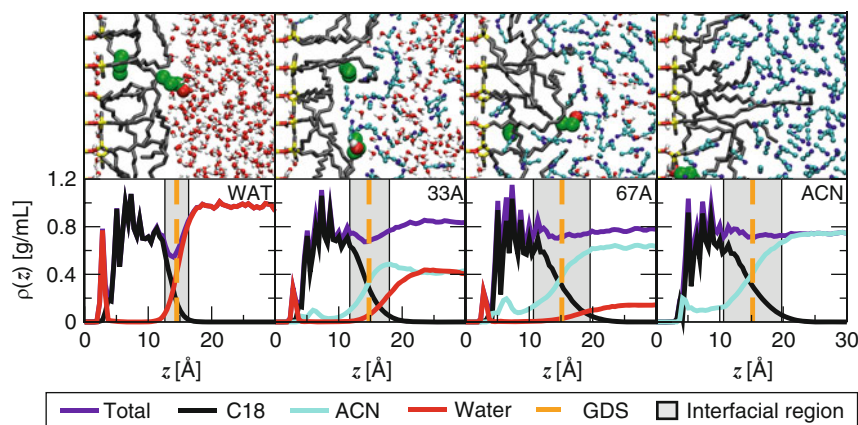


Fig. 3 Snapshots and density profiles of RPLC systems with varying water/acetonitrile ratios. In the snapshots the stationary phase is shown as tubes with carbon in *black*, silicon in *yellow*, oxygen in *red*, and hydrogen in *white*. The mobile phase is shown in the *ball and stick* representation with carbon in *cyan*, nitrogen in *blue*, oxygen in *red*, and hydrogen in *white*. The analytes are shown as *large spheres* with carbon in *green*, oxygen in *red*, and hydrogen in *white*. In the density profiles, the Gibbs dividing surface (GDS) is a plane that defines the boundary between the mobile and stationary phase [55, 56] and the *gray shaded area* represents the width of the 10–90 interfacial region [57] according to the total solvent density

interaction with the surface silanols and acetonitrile does not. This is also apparent when examining the fraction of silanols involved in a hydrogen bond with solvent. In system WAT, only $\approx 20\%$ of silanols are not involved in hydrogen bonds to solvent while in system ACN more than 80% of the silanols are free. This is an important effect because the degree of saturation of these surface silanols can have significant consequences on the retention of analytes that can interact with the silanols, as will be demonstrated later.

Moving into the interfacial region, other interesting effects are observed. First, there is a distinct minimum in the total system density for system WAT, but this minimum disappears as acetonitrile concentration is increased. This dewetting effect for water near extended hydrophobic surfaces has been predicted by Lum, Chandler, and Weeks and is attributed to a disruption of the solvent's hydrogen bonding network [58]. Also present in the interfacial region is an enrichment in the acetonitrile concentration for the binary solvent systems. This effect is most dramatic for system 33A, where a distinct density maximum for acetonitrile is observed at around $z = 17$ Å.

The preferential sorption of the organic component of the solvent is a well known effect in RPLC. Based on their interpretation of adsorption isotherm measurements with water/acetonitrile mixtures, Kazakevich and coworkers have inferred that the excess adsorption occurs purely at the surface of the alkyl chains and that acetonitrile forms up to five molecular layers atop the alkyl chains. Clearly, from the snapshots and the density profiles in Fig. 3, this layering does not appear. The simulations indicate that the excess adsorption occurs both within the chain region and at the chain surface. Furthermore, the adsorbed solvent at the surface is not pure acetonitrile; it is simply a local enhancement of the acetonitrile concentration.

From the solvent density profiles shown in Fig. 3, it is clear that the alkyl chains in the stationary phase are solvated to a larger extent when the concentration of acetonitrile in the mobile phase is increased. To ascertain what effect this solvation has on the structure of the C_{18} chains in the stationary phase, various structural properties for the alkyl chains can be examined (see Table 1 and Fig. 4). The first structural parameter assessed is the fraction of gauche defects along the carbon backbone of the chains. The fraction of gauche defects is around 26% in all solvent systems. This is in agreement with Raman spectroscopic measurements by

Table 1 Structural properties of alkylsilane chains in contact with different mobile phase solvents^a

System				
Property	WAT	33A	67A	ACN
f_{gauche}	0.25 ₁	0.27 ₁	0.26 ₁	0.27 ₁
$\cos \theta_{\text{ete}}$	0.25 ₁	0.46 ₃	0.52 ₂	0.56 ₂
${}^z\text{CH}_3$ (Å)	9.1 ₁	11.2 ₃	12.4 ₁	12.9 ₃
S	-0.14 ₂	-0.02 ₁	0.02 ₂	0.05 ₁

^aSubscripts indicate the statistical uncertainty in the final digit

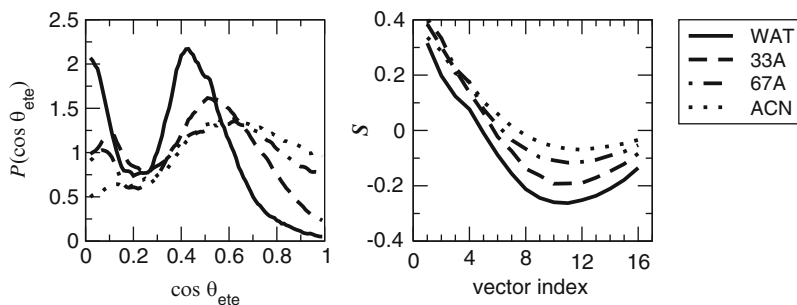


Fig. 4 Probability distribution for the angle between the C_{18} end-to-end vectors and the silica surface normal (*left*) and the order parameter along the chain backbone (*right*)

Pemberton and coworkers that indicated little dependence of the dihedral angle on solvent composition [46, 47].

Another important structural characteristic of the RPLC stationary phase is the degree of chain alignment. One parameter that measures this alignment is $\cos \theta_{ete}$, where θ_{ete} is the angle between the chain end-to-end vector and the normal to the silica surface, or the tilt angle. Unlike the fraction of gauche defects, the orientation of the end-to-end vector changes substantially with changes in solvent composition. In system WAT, the value of $\cos \theta_{ete}$ is 0.25 ($\theta_{ete} = 75^\circ$) and it increases to 0.56 (56°) in system ACN. A structural parameter complementary to the end-to-end orientation is z_{CH_3} , the height of the terminal methyl group above the silica surface. As indicated in Table 1, z_{CH_3} also steadily increases as the concentration of organic modifier is increased. Thus it appears that the chains become directed away from the silica surface and are more extended as acetonitrile concentration is increased.

In addition to the average value of $\cos \theta_{ete}$, it is useful to examine the distribution of this angle to ascertain whether a particular chain tilt angle is preferred (like the uniform tilt angle observed for alkyl monolayers on metal surfaces). These distributions are shown in Fig. 4. For system WAT, the distribution is clearly bimodal with peaks in $\cos \theta_{ete}$ near values 0 and 0.45, corresponding to chains nearly parallel to the surface and chains with $\theta_{ete} \approx 65^\circ$, respectively. This bimodal behavior also appears in the other solvent systems but, as the concentration of organic modifier is increased, the height of the peak corresponding to chains parallel to the surface decreases significantly. Furthermore, the peak corresponding to more extended chains becomes broadened and shifts to around $\cos \theta_{ete} \approx 0.6$. It is interesting to note that, despite the large differences in end-to-end orientation in the different solvents, there remains a somewhat broad distribution of chain alignments in all systems. There is no single conformation that dominates in any system and there remains a probability for both parallel and perpendicular chains.

Indeed, the chains are more extended in the mobile phases with higher organic concentrations. However, the simulation data does not indicate that the chains become collapsed in a purely aqueous solvent. This is clearly indicated by the broad $\cos \theta_{ete}$ distribution shown in Fig. 4 and the position of the Gibbs dividing

surface shown in Fig. 3, which shows little dependence of solvent composition. Therefore, retention loss when highly aqueous solvents are used is more likely due to pore dewetting as opposed to chain collapse.

The end-to-end vector discussed above provides a picture of the overall alignment of the chains, but more local information on individual segments within the chain can be gleaned from the orientational order parameter S_i along the chain backbone:

$$S_i = \frac{1}{2} \langle 3\cos^2\theta_i - 1 \rangle, \quad (3)$$

where θ_i is the angle between the i th 1–3 backbone vector in the C_{18} chain and the normal to the silica surface. This order parameter is equivalent to the experimentally observable NMR order parameter for deuterated alkyl chains [59, 60]. Figure 4 shows this order parameter for each 1–3 vector along the chain backbone and Table 1 gives S_n , the value of the order parameter averaged over all 16 1–3 backbone vectors. Looking at the order parameter along the chain backbone one sees a similar trend for all four solvent systems. The order parameter is large and positive for the first few backbone vectors and reaches a minimum somewhere near vector number 10, and then approaches zero beyond this minimum. Thus, the initial portion of the chain shows a significant preference to align itself away from the silica surface while the terminal portion of the chain is oriented more randomly. Despite the similarity in shape, the curves are shifted upward as acetonitrile concentration is increased, again indicating an increase in chain alignment.

Now that structural characteristics of the RPLC system have been discussed, a discussion of the retention mechanism can begin. To describe the mechanism of analyte retention in RPLC one needs to know, with high resolution, the preferred locations and orientations of the analyte molecule within the stationary phase. The simulations described here are able to yield this type of data directly. The preferred locations of the analytes are described through the z -dependent distribution coefficient profiles, or $K(z)$ plots, shown in Fig. 5 for n -butane and 1-propanol. These profiles are analogous to the (experimentally measurable) distribution coefficient for transfer from mobile to stationary phase but offer much more detailed information on where retention occurs within the stationary phase. Larger values of $K(z)$ correspond to more favorable (lower free energy) locations of the analyte within the stationary phase. In examining these profiles, one of the most striking features is the large dependence of the analyte distribution coefficient on z . From this, it is clearly evident that the stationary phase is not a homogeneous medium into which analytes partition nor a nonpolar surface to which analytes adsorb. Rather, the stationary phase is a microheterogeneous medium with multiple preferred regions for the analytes.

For n -butane the $K(z)$ profiles show a bimodal distribution in all solvent systems. There is one peak in the center of the bonded phase ($z \approx 8 \text{ \AA}$) and another in the interfacial region. The peak in the center of the bonded phase remains rather sharp regardless of solvent composition. However, the shape of the peak in the interfacial

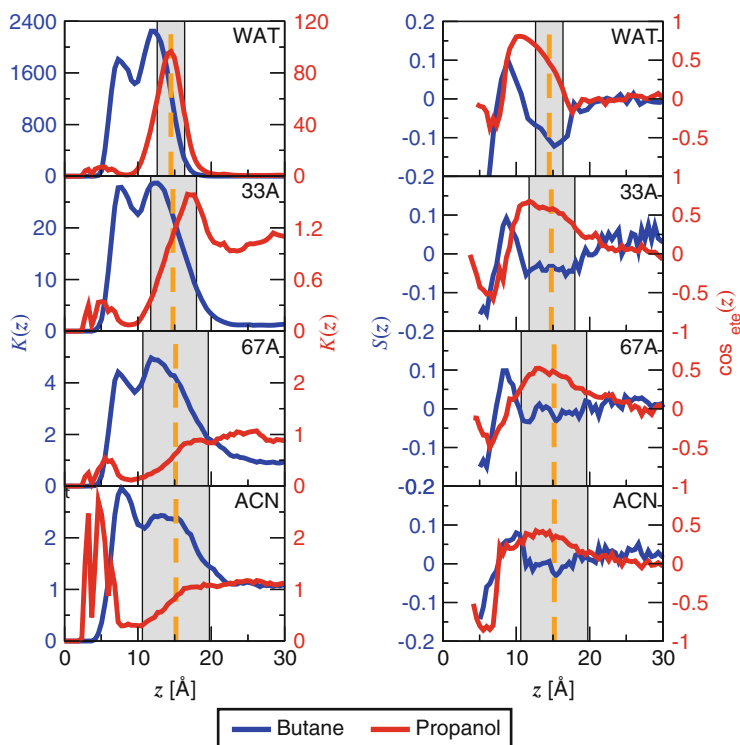


Fig. 5 Distribution coefficient (*left*) and orientational (*right*) profiles for *n*-butane and 1-propanol. The GDS for each system is indicated by a *dashed vertical* and the interfacial region is *shaded*

region broadens as acetonitrile concentration increases. The broadening of this interfacial peak coincides with the increasing width of the interfacial region. From this it is apparent that even a simple nonpolar analyte has multiple modes of sorption. It can either partition deep into the bonded phase or adsorb at the hydrocarbon surface.

The preference of *n*-butane to reside in the interfacial region is not entirely surprising since a density depletion is observed in this region (see Fig. 3). However, the peak deeper in the bonded phase is in a region where the overall system density is significantly higher. Analysis of the bonded-phase structure shows that it is much more ordered in this region, as indicated by the larger *S* values for the initial portion of the chain (Fig. 4). Thus, there may be more free volume of appropriate size and shape for the analyte in this region.

A very different retention mechanism is observed for the polar analyte, 1-propanol, and this retention mechanism appears to be somewhat dependent on the mobile phase composition. In system WAT, 1-propanol exhibits a distinct preference to reside in the interfacial region with a peak centered directly at the GDS. This preference to adsorb at the alkyl surface clearly diminishes as the fraction of organic modifier is increased and 1-propanol becomes more soluble in

the solvent. Additional but smaller peaks in the $K(z)$ profile of 1-propanol are observed in the $z = 3\text{--}7$ Å region of system WAT. These are due to hydrogen bonding of 1-propanol molecules directly to residual surface silanols and to solvent molecules which are bound to the substrate [32, 34]. Interestingly, these peaks become much stronger as acetonitrile concentration is increased. This is related to the availability of the surface silanols for hydrogen bonding. Interestingly, this effect is not observed for water/methanol mixtures where methanol is observed to have a higher affinity to interact with the surface silanols than acetonitrile [32, 34].

In addition to a comprehension of where an analyte molecule is retained within the stationary phase, a complete description of the retention mechanism would also require knowledge of how the analytes are oriented. For this reason, the z -dependent S profiles for n -butane and $\cos \theta_{\text{ete}}$ profiles for 1-propanol are shown in Fig. 5. For n -butane, the order parameter S is the same as that described for the alkyl chains in Table 1, except here is it plotted as a function of z . For 1-propanol, the end-to-end vector originates at the methyl group and terminates at the hydroxyl hydrogen. Thus, values of $\cos \theta_{\text{ete}}$ will be positive for hydroxyl groups pointing away from the silica surface, and negative for hydroxyl groups directed towards the surface.

The $S(z)$ profiles indicate that n -butane has some orientational preferences in the stationary phase and interfacial region, although these preferences are not very strong. In the interfacial region, the n -butane molecule prefers to lie parallel to the interface, consistent with interfacial adsorption. In the center of the bonded phase, n -butane changes its orientational preference to perpendicular with a maximum in $S(z)$ at around $z = 8$ Å. This maximum occurs at the same position as the peak in the $K(z)$ profile that was attributed to partitioning. Clearly, given the perpendicular orientational preference of n -butane and alignment of the chains in this region, this partitioning does not resemble bulk liquid–liquid partitioning for which one would see no orientational preference. Moving further into the bonded phase, $S(z)$ values become negative at around $z = 6$ Å, indicating a parallel preference as the n -butane analyte nears the silica surface.

1-Propanol exhibits much stronger orientational preferences than n -butane. In the interfacial region, the $\cos \theta_{\text{ete}}(z)$ profiles indicate that this analyte has a preference to direct its hydroxyl group towards the mobile phase and its alkyl tail towards the stationary phase. With this orientation, the analyte can hydrogen bond with the solvent while its nonpolar tail is solvated by the hydrocarbon stationary phase. The magnitude of the interfacial peak in $\cos \theta_{\text{ete}}(z)$ decreases as the molfraction of the organic modifier is increased. Moving through the bonded phase to near the silica surface, $\cos \theta_{\text{ete}}(z)$ shifts from positive to negative values, thus indicating that the polar hydroxyl group is directed towards the silica surface. In this manner, the 1-propanol analyte can form hydrogen bonds with the surface silanols. This preference is weakest in system WAT, where the silanols are mostly saturated, but very strong in system ACN, where the silanols are mostly unsaturated.

In order to validate the simulation data against experiment, and to assess the contribution of the mobile and stationary phases to the thermodynamics of

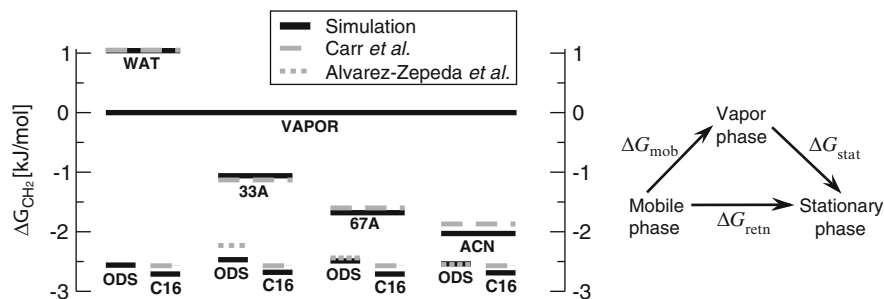


Fig. 6 Free energy level diagram for the methylene group (*left*) and thermodynamic cycle used to decompose the free energy of retention into mobile and stationary phase components (*right*). Simulation data are shown as *solid black lines* and experimental measurements are shown as *dashed gray lines*. The experimental data are from [52, 61]

retention, incremental free energies of transfer were computed for the methylene group, ΔG_{CH_2} . These free energies are computed from the $K(z)$ profiles for the series of alkane analytes used in the simulation. This data is presented in Fig. 6 in the form of a free energy level diagram. This diagram takes the vapor phase present in the simulation as the zero free energy reference state. Also present in the diagram is the free energy level of a liquid *n*-hexadecane phase to facilitate comparisons to bulk liquid–liquid partitioning. The simulation data for *n*-hexadecane comes from [33].

First, it should be noted that the calculated values of ΔG_{CH_2} are in excellent agreement with experiment (compare the solid black and dashed gray lines in Fig. 6). The largest deviation observed in the diagram is around 0.2 kJ/mol, a very small value in terms of free energy, and in most cases the deviation is much smaller. The fact that the simulations are able to reproduce the thermodynamics of the retention process, as measured experimentally, provides good confidence that the molecular details observed in simulations are indeed correct.

The free energy diagram also allows for a discussion of the driving forces for retention. Shown on the right in Fig. 6 is a thermodynamic cycle decomposing the free energy of retention (ΔG_{retn}) into mobile and stationary phase components (ΔG_{mob} and ΔG_{stat}) using the vapor phase as a reference state. As shown in Fig. 6, the mobile to vapor phase transfer (or mobile phase contribution to retention) for the methylene group is favorable only in system WAT. Thus, solvophobic forces [50, 62, 63] are not important for retention unless highly aqueous mobile phases are used. The free energy of transfer from the vapor phase to the ODS stationary phase (stationary phase contribution) is always favorable and greater in magnitude than the mobile phase contribution in all solvent systems examined. This is in agreement with the lipophilic view of Carr and coworkers [48, 49, 52], which argues that interactions with the stationary phase drives the retention process. The free energy level of the ODS phase changes little with changing solvent composition suggesting that solvent penetration into the stationary phase and its influence on chain alignment has little effect on nonpolar groups.

Comparing the ODS phase to the *n*-hexadecane phase, one sees that they are very similar in terms of free energy. This similarity has been used to suggest that the retention mechanism in RPLC resembles bulk liquid–liquid partitioning [49, 64]. The current work shows this conclusion is not valid and this comparison should not be made when assessing the molecular mechanism of retention. The profiles in Fig. 5 clearly indicate that nonpolar analytes can either partition into the stationary phase or adsorb at the surface and in either case have a clear orientational preference, which would not be present in a bulk liquid.

4.2 Shape Selectivity in RPLC

In most cases, the difference in retention times between two analytes is governed by differences in physical properties such as polarity. For example, a more polar analyte will have a higher affinity to reside in the mobile phase and, therefore, have a shorter retention time. However, many RPLC systems have the ability to separate analytes that have very similar chemical functionalities and physical properties, but differ only slightly in molecular shape. This ability, termed “shape selectivity,” is typically applied in the context of separating the geometric isomers of rigid molecules, such as polycyclic aromatic hydrocarbons (PAHs), among others. However, the separation of alkane isomers illustrated in Fig. 2 is also a form of shape selectivity.

The topic of shape selectivity has been examined in detail by Sander, Wise, and coworkers [65–70]. In general, it has been noted that, for PAHs of the same molecular formula (i.e., geometric isomers), ones with larger length to breadth (*L/B*) ratios are more retained. For example, in Fig. 7 the naphthacene molecule has the largest *L/B* ratio and the triphenylene molecule the smallest. It has also been observed that greater selectivity between different PAH isomers can be achieved by adjusting chromatographic parameters that lead to increased ordering of the stationary phase chains. These parameters include decreasing temperature, increased grafting density, increased chain length, and

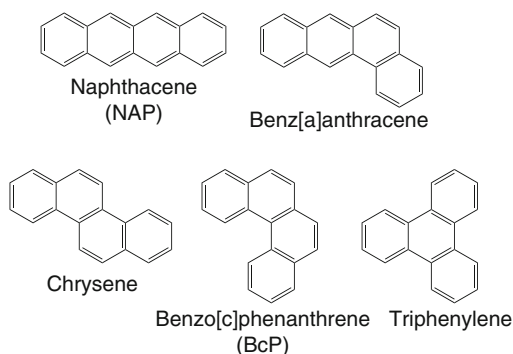


Fig. 7 Four-ring PAH molecules present in the GEMC/CBMC simulations

the use of polymeric, as opposed to monomeric, bonded phases. The effect of grafting density will be examined here.

To discern the molecular causes of shape selectivity in some traditional RPLC systems, GEMC/CBMC simulations were carried out with a water/acetonitrile mobile phase containing ≈ 67 mol% acetonitrile and a stationary phase with ODS chains at surface coverages of 1.60, 2.88, and $4.15 \mu\text{mol}/\text{m}^2$ at a temperature of 308 K [40]. In addition to the stationary and mobile phase entities, numerous PAH analytes were present ranging in size from benzene up to the four ring isomers shown in Fig. 7. Although the focus of this study was the larger PAH analytes since they exhibit the largest shape effects, the smaller PAH analytes are needed as intermediates for the identity exchange moves mentioned in Sect. 2. It would be extremely difficult to sample the spatial distribution of the large PAH analytes by traditional means. However, by “growing in” these large molecules starting from a small molecule like benzene, a much higher sampling efficiency can be achieved [40].

As mentioned above, an important aspect of simulation is validation against experimental results. For this purpose, the selectivities between the five different four-ring isomers (see Fig. 7) computed from simulation are compared to experiment [71] in Table 2. The selectivity α for each PAH is computed as the ratio of its capacity factor k' to the capacity factor for triphenylene, the least retained of the four-ring isomers:

$$\alpha_x = k'_x / k'_{\text{Triphenylene}}. \quad (4)$$

The capacity factor can be determined experimentally from retention data and is exactly equivalent to the average number of analyte molecules in the stationary phase divided by the average number in the mobile phase. Thus, this quantity is easily measured in a GEMC/CBMC simulation.

As can be seen in Table 2, the selectivities computed from simulation are in excellent agreement with experiment. This agreement is a testament to the precision of the simulation method and the accuracy of the force field. For example, the selectivity between chrysene and triphenylene at a coverage of $2.8 \mu\text{mol}/\text{m}^2$ is

Table 2 Simulated selectivities of $\text{C}_{18}\text{H}_{12}$ PAH isomers relative to triphenylene compared to experiment at different surface coverages^{a,b}

	Coverage ($\mu\text{mol}/\text{m}^2$)	Naphthacene	Chrysene	Benz[a]-anthracene	Benzo[c]pyrene
Simulation	4.15	1.56 ₁₆	1.14 ₀₈	1.16 ₁₀	1.02 ₀₄
Experiment	4.20	1.70	1.13	1.13	1.08
Simulation	2.88	1.27 ₁₄	1.12 ₀₈	1.07 ₀₉	1.06 ₀₅
Experiment	2.84	1.58	1.08	1.12	1.06
Simulation	1.60	1.11 ₁₀	1.12 ₀₉	1.01 ₁₁	1.06 ₁₀
Experiment	1.60	1.19	0.97	1.02	1.02

^aSubscripts indicate the standard error of the mean in the final two digits

^bExperimental data from Sentell and Dorsey [71]

around 1.1. This represents a difference in free energy of retention of 0.2 kJ/mol, an extremely small quantity to measure in either experiment or simulation. Additionally, this agreement with experiment allows for much greater confidence in the molecular details from simulation results.

Analysis of $K(z)$ profiles and of the orientational distribution for the four-ring PAHs shows that the analyte molecules penetrate deeply into the bonded-phase region (the maxima in $K(z)$ are typically found at $z \approx 10$ Å) where the analytes exhibit an increasing preference to align perpendicular to the surface with increasing ODS coverage [40]. The orientational preference is most pronounced for NAP, the most retained compound. Analysis of the lateral distribution of the analytes at the highest ODS coverage (where the largest selectivities are observed) indicates that NAP is preferentially found in locations that are relatively more crowded with ODS chains, whereas BcP exhibits a preference for relatively less crowded regions [40].

From a calculation of the density of ODS methylene segments for configurations without an analyte molecule being present, it is evident that the upper part of the bonded phase resembles a liquid phase without pre-existing cavities [40]. Furthermore, it can be shown that the chain structure is modified by the presence of the PAH analytes. For all coverages, values of $\cos \theta_{\text{ete}}$ and S increase by about 0.1 for ODS chains near to any of the four-ring PAH analytes compared ODS chains without any analyte being present, i.e., the conformational flexibility of the ODS chains allows them to respond to the presence of the analyte and the favorable regions do not correspond to static cavities.

5 Conclusions

The GEMC/CBMC simulation methodology applied here has proved very useful for studying structure and retention in complex GLC and RPLC systems. The methodology affords the computation of retention data with sufficient precision (and accuracy for the TraPPE force field) for complex analytes to validate the predictions against experimental data. This validation, in turn, builds confidence in the molecular-level details on structure and retention mechanism that can be obtained from the simulations.

In closing, we would like to mention some applications of the GEMC/CBMC approach and very much related combination of CBMC and the grand canonical Monte Carlo technique to other complex systems: prediction of structure and transfer free energies into dry and water-saturated 1-octanol [72], prediction of the solubility of polymers in supercritical carbon dioxide [73], prediction of the upper critical solution pressure for gas-expanded liquids [74], investigation of the formation of multiple hydrates for a pharmaceutical compound [75], exploration of multicomponent vapor-to-particle nucleation pathways [76], and investigations of the adsorption of articulated molecules in zeolites and metal organic frameworks [77, 78].

Acknowledgments Financial support from the National Science Foundation (CHE-0718383) and The Dow Chemical Company is gratefully acknowledged. Part of the computer resources was provided by the Minnesota Supercomputing Institute.

References

1. Panagiotopoulos AZ (1987) *Mol Phys* 61:813
2. Panagiotopoulos AZ, Quirke N, Stapleton M, Tildesley DJ (1988) *Mol Phys* 63:527
3. Martin MG, Siepmann JI (1999) *J Phys Chem B* 103:4508
4. Mooij GCAM, Frenkel D, Smit B (1992) *J Phys Condens Matter* 4:L255
5. Siepmann JI, McDonald IR (1992) *Mol Phys* 75:255
6. Martin MG, Siepmann JI (1997) *J Am Chem Soc* 119:8921
7. Siepmann JI, Frenkel D (1992) *Mol Phys* 75:59
8. Wick CD, Siepmann JI (2000) *Macromolecules* 33:7207
9. Vlucht TJH, Martin MG, Smit B, Siepmann JI, Krishna R (1998) *Mol Phys* 94:727
10. McGrath M, Siepmann J, Kuo IF, Mundy C, VandeVondele J, Hutter J, Mohamed F, Krack M (2006) *J Phys Chem A* 110:640
11. Braithwaite A, Smith FJ (1985) *Chromatographic methods*. Chapman and Hall, London
12. Walter TH, Iraneta P, Capparella M (2005) *J Chromatogr A* 1075:177
13. McCormick RM, Karger BL (1980) *Anal Chem* 52:2249
14. Berezkin VG (1991) *Gas-liquid-solid chromatography*. Marcel Dekker, New York, NY
15. Kovats E (1965) *Adv Chromatogr* 1:229
16. Martin MG, Siepmann JI (1998) *J Phys Chem B* 102:2569
17. Chen B, Potoff JJ, Siepmann JI (2001) *J Phys Chem B* 105:3093
18. Stubbs JM, Potoff JJ, Siepmann JI (2004) *J Phys Chem B* 108:17596
19. Martin MG, Siepmann JI, Schure MR (1999) *J Phys Chem B* 103:11191
20. Wick CD, Siepmann JI, Klotz WL, Schure MR (2002) *J Chromatogr A* 954:181
21. Sun L, Siepmann JI, Klotz WL, Schure MR (2006) *J Chromatogr A* 1126:373
22. Wick CD, Siepmann JI, Schure MR (2002) *Anal Chem* 74:3518
23. Wick CD, Siepmann JI, Schure MR (2002) *Anal Chem* 74:37
24. Schure MR (1994) In: Pesek JJ, Leigh IE (eds) *Chemically modified surfaces*. Royal Society of Chemistry, Cambridge, pp 181–189
25. Klatte SJ, Beck TL (1996) *J Phys Chem* 100:5931
26. Slusher JT, Mountain RD (1999) *J Phys Chem B* 103:1354
27. Zhao CF, Cann NM (2008) *Anal Chem* 80:2426
28. Fouqueau A, Meuwly M, Bemish RJ (2007) *J Phys Chem B* 111:10208
29. Melnikov SM, Hölzel A, Seidel-Morgenstern A, Tallarek U (2009) *J Phys Chem C* 113:9230
30. Zhang L, Sun L, Siepmann JI, Schure MR (2005) *J Chromatogr A* 1079:127
31. Zhang L, Rafferty JL, Siepmann JI, Chen B, Schure MR (2006) *J Chromatogr A* 1126:219
32. Rafferty JL, Zhang L, Siepmann JI, Schure MR (2007) *Anal Chem* 79:6551
33. Rafferty JL, Sun L, Siepmann JI, Schure MR (2010) *Fluid Phase Equilib* 290:25
34. Rafferty JL, Sun L, Siepmann JI, Schure MR (2011) *J Chromatogr A* 1218:2203
35. Rafferty J, Siepmann J, Schure M (2008) *J Chromatogr A* 1204:11
36. Rafferty J, Siepmann J, Schure M (2008) *J Chromatogr A* 1204:20
37. Rafferty JL, Siepmann JI, Schure MR (2008) *J Chromatogr A* 1216:2320
38. Rafferty JL, Siepmann JI, Schure MR (2010) In: Brown PR, Grushka E (eds) *Advances in chromatography*, vol. 48. Marcel Dekker, New York, pp 1–55
39. Rafferty JL, Siepmann JI, Schure MR (2008) *Anal Chem* 80:6214
40. Rafferty JL, Siepmann JI, Schure MR (2011) *J Chromatogr A* Submitted for publication
41. Li Z, Rutan SC, Dong S (1996) *Anal Chem* 68:124

42. Gilpin RK, Gangoda ME, Krishen AE (1982) *J Chromatogr Sci* 20:345
43. Yang SS, Gilpin RK (1987) *J Chromatogr* 394:295
44. Rustamov I, Farcas T, Chan F, LoBrutto R, McNair HM, Kazakevich YV (2001) *J Chromatogr A* 913:49
45. Gritti F, Kazakevich YV, Guiochon G (2007) *J Chromatogr A* 1169:111
46. Pemberton JE, Ho M, Orendorff CJ, Ducey MW (2001) *J Chromatogr A* 913:243
47. Ducey MW, Orendorff CJ, Pemberton JE, Sander LC (2002) *Anal Chem* 74:5585
48. Carr PW, Li J, Dallas AJ, Eikens DI, Tan LT (1993) *J Chromatogr A* 656:113
49. Carr PW, Tan LC, Park JH (1996) *J Chromatogr A* 724:1
50. Vailaya A, Horváth C (1998) *J Chromatogr A* 829:1
51. Gritti F, Guiochon G (2005) *J Chromatogr A* 1099:1
52. Ranatunga RPJ, Carr PW (2000) *Anal Chem* 72:5679
53. Sander LC, Wise SA (1993) *J Chromatogr A* 656:335
54. Devido DR, Dorsey JG, Chan HS, Dill KA (1998) *J Phys Chem B* 102:7272
55. Zangwill A (1988) *Physics at surfaces*. Cambridge University Press, Cambridge, UK
56. Adamson AW, Gast AP (1997) *Physical chemistry of surfaces*, 6th edn. Wiley-Interscience, New York
57. Cahn JW, Hilliard JE (1958) *J Chem Phys* 28:258
58. Lum K, Chandler D, Weeks JD (1999) *J Phys Chem B* 103:4570
59. Seelig A, Seelig J (1974) *Biochemistry* 13:4839
60. van der Ploeg P, Berendsen HJC (1982) *J Chem Phys* 76:3271
61. Alvarez-Zepeda A (1991) Part I. A thermodynamic study of acetonitrile + water mixtures in reversed-phase liquid chromatography. Part II. A gas chromatographic study of the effect of temperature and column pressure on solute retention on a bonded phase using helium and carbon dioxide as mobile phases. Ph.D. thesis, Georgetown University, Washington DC
62. Horváth C, Melander W, Molnár I (1976) *J Chromatogr* 125:129
63. Vailaya A, Horváth C (1997) *J Phys Chem B* 101:5875
64. Dorsey J, Dill KA (1989) *Chem Rev* 89:331
65. Wise SA, Bonnett WJ, Guenther FR, May WE (1981) *J Chromatogr Sci* 19:457
66. Wise SA, Sander LC (1985) *J High Resol Chromatogr* 8:248
67. Sander LC, Wise SA (1995) *Anal Chem* 67:3284
68. Sander LC, Pursch M, Wise SA (1999) *Anal Chem* 71:4821
69. Lippa KA, Sander LC, Wise SA (2004) *Anal Bioanal Chem* 378:365
70. Rimmer CA, Lippa KA, Sander LC (2008) *LC GC North America* 26:984
71. Sentell K, Dorsey JG (1989) *J Chromatogr* 461:193
72. Chen B, Siepmann JI (2006) *J Phys Chem B* 110:3555
73. Wick CD, Siepmann JI, Theodorou DN (2005) *J Am Chem Soc* 127:12338
74. Zhang L, Siepmann J (2005) *J Phys Chem B* 109:2911
75. Zhao XS, Siepmann JI, Xu W, Kiang YH, Sheth AR, Karaborni S (2009) *J Phys Chem B* 113:5929
76. Kathmann SM, Schenter GK, Garrett BC, Chen B, Siepmann J (2009) *J Phys Chem B* 113:10354
77. Smit B, Maesen TLM (2008) *Chem Rev* 108:4125
78. Dubbeldam D, Galvin CJ, Walton KS, Ellis DE, Snurr RQ (2008) *J Am Chem Soc* 130:10884

Thermodynamic Properties for Applications in Chemical Industry via Classical Force Fields

Gabriela Guevara-Carrion, Hans Hasse, and Jadran Vrabec

Abstract Thermodynamic properties of fluids are of key importance for the chemical industry. Presently, the fluid property models used in process design and optimization are mostly equations of state or G^E models, which are parameterized using experimental data. Molecular modeling and simulation based on classical force fields is a promising alternative route, which in many cases reasonably complements the well established methods. This chapter gives an introduction to the state-of-the-art in this field regarding molecular models, simulation methods, and tools. Attention is given to the way modeling and simulation on the scale of molecular force fields interact with other scales, which is mainly by parameter inheritance. Parameters for molecular force fields are determined both bottom-up from quantum chemistry and top-down from experimental data. Commonly used functional forms for describing the intra- and intermolecular interactions are presented. Several approaches for ab initio to empirical force field parameterization are discussed. Some transferable force field families, which are frequently used in chemical engineering applications, are described. Furthermore, some examples of force fields that were parameterized for specific molecules are given. Molecular dynamics and Monte Carlo methods for the calculation of transport properties and vapor-liquid equilibria are introduced. Two case studies are presented. First, using liquid ammonia as an example, the capabilities of semi-empirical force fields, parameterized on the basis of quantum chemical information and experimental data, are discussed with respect to thermodynamic properties that are relevant for the chemical industry. Second, the ability of molecular simulation methods to describe accurately vapor-liquid equilibrium properties of binary mixtures containing CO₂ is shown.

Keywords Force field · Molecular dynamics · Monte Carlo · Thermodynamics

G. Guevara-Carrion and H. Hasse

Laboratory for Engineering Thermodynamics, University of Kaiserslautern, Erwin Schrödinger-Str. 44, 67663 Kaiserslautern, Germany

J. Vrabec (✉)

Thermodynamics and Energy Technology, University of Paderborn, Warburger Str. 100, 33098 Paderborn, Germany

e-mail: jadran.vrabec@upb.de

Contents

1	Introduction	202
2	Force Fields	203
2.1	Intermolecular Interactions	204
2.2	Intramolecular Interactions	210
3	Force Field Parameterization	213
3.1	Ab Initio Parameterization	214
3.2	Empirical Parameterization	217
4	Force Field Families	218
4.1	Transferable Force Fields	218
4.2	Specific Force Fields	223
5	Molecular Simulation Methods	226
5.1	Molecular Dynamics	226
5.2	Monte Carlo	226
5.3	Methods for Determining Phase Equilibria	227
5.4	Methods for Determining Transport Properties	229
5.5	Simulation Tools	230
6	Case Study: Ammonia	231
6.1	Force Fields	231
6.2	Vapor–Liquid Equilibria of Ammonia	232
6.3	Properties of the Homogeneous State	233
7	Case Study: Binary Mixtures Containing CO ₂	235
7.1	Force Fields	235
7.2	Henry’s Law Constant of CO ₂ in Ethanol	236
7.3	Vapor–Liquid Equilibria of the Mixture CO ₂ + C ₂ H ₆	236
8	Concluding Remarks	238
	References	238

1 Introduction

The knowledge of thermodynamic properties plays a crucial role in the design and operation of chemical plants [1]. Therefore, the chemical industry requires reliable and accurate thermodynamic data for very different fluids, covering a wide range of temperature, pressure, and composition [1–6]. There is a great demand for data on vapor–liquid, liquid–liquid, and solid–liquid equilibria, as well as an increasing need for caloric and transport properties [1]. Classical approaches to predict these properties like equations of state and G^E models, as reviewed, e.g., by Poling et al. [7], do exist. However, the parameters of these models are determined based on experimental data. These are often not available and may be difficult to obtain, especially for extreme conditions or when hazardous substances are involved. Furthermore, as the amount of experimental data is always limited, usually extrapolations are necessary, but they are inherently uncertain. Therefore, an alternative route to determine fluid properties, independent of the established phenomenological approaches, is highly desirable. This would allow the carrying out of predictions in different ways and, if the results agree, give confidence or, alternatively, give a warning regarding the quality of the extrapolation.

Moreover, most processes in the chemical industry are governed by nanoscale phenomena. In many cases the nanoscale structure plays an important role, e.g., the local concentrations and not the overall concentrations govern reactions at active sites of catalysts. Phenomenological thermodynamics provides no route to obtain insight in these nanoscale structures and processes, whereas molecular simulations based on forces fields do. The key is to carry them out with models that are suitably developed and reasonably represent the compounds.

Molecular modeling and simulation comprises computational techniques derived from quantum chemistry and statistical mechanics to predict equilibrium and non-equilibrium properties of molecular ensembles based on intra- and inter-molecular interaction potentials. Because of the ongoing exponential increase in computing power and the development of new numerical methods, the range of molecules that can be covered and the accuracy of the results is growing rapidly [8]. Nowadays, molecular modeling and simulation is being actively applied in physical, chemical, and biological sciences, as well as in engineering research, and its importance will further increase [1, 9]. The development of new molecular theories as well as the prediction of material properties as a function of molecular structure and thermodynamic conditions are other examples of current applications of molecular methods. Moreover, molecular simulation can also provide insight into the molecular behavior and properties which are not experimentally accessible.

One of the central issues of the molecular approach is to devise adequate force fields that accurately describe the properties of real systems. Depending on the application field, different requirements need to be fulfilled. In biology, for instance, to study protein folding in aqueous environments, typically rather complex force fields are used to determine microscopic molecular structures. In the chemical industry, much more aggregated macroscopic properties are needed, but the quantitative correctness of the data is essential.

Simulation results are primarily determined by the employed force field. Fortunately, in the past two decades, the quality of force fields has greatly improved, mainly due to the inclusion of molecular parameters obtained from high-level *ab initio* calculations [10]. The aim of this chapter is to review the current status of force field development and application for the prediction of thermodynamic properties of fluids that are relevant for the chemical industry.

2 Force Fields

The development of force fields comprises a trade-off between computational feasibility and coverage of the molecular interactions details and is thus driven by the growth of computational resources [9].

Force fields are a set of mathematical functions and parameters that relate a potential energy to a configuration of the regarded molecular system. The potential energy is usually described by pair potentials. Three- and more-body interactions contribute significantly to the potential energy as well [11], but are usually not

explicitly included in engineering force fields because of their high computational cost. Rather, their contributions are incorporated into pairwise approximations by effective pair potentials [11].

In molecular force fields, the interaction energy between sites can be divided into contributions from intramolecular and intermolecular interactions. The significance of the different contributions to the force field varies depending on the required application. E.g., for industrial engineering applications, simple models with a low computational cost are required that are nonetheless able to predict accurately thermodynamic properties. Numerous force fields of varying complexity are currently available. The simplest force fields include only potentials that describe the intermolecular interactions and are frequently used for small molecules. More complex force fields include intramolecular interactions that are necessary for the simulation of larger molecules such as polymers.

All-atom force fields consider every atom as an individual interaction site, while united-atom force fields gather different atoms of a functional group into one interaction site, e.g., as is often done to model methyl or methylene groups. To describe chain-like polymers or proteins, coarse grained force fields are also employed, where the interaction sites usually represent a larger number of atoms.

2.1 Intermolecular Interactions

In modeling with classical force fields, the intermolecular interactions are usually divided into Van der Waals interactions (repulsion and dispersion) and electrostatic interactions. In this framework, the Van der Waals interactions take into account all interactions between sites that are not related to permanent electrostatics, such as dispersion, repulsion, and induction [12]. Hydrogen bonding is usually modeled by electrostatic sites. For a detailed discussion of the intermolecular interactions, the interested reader is referred to [13].

2.1.1 Van der Waals Interactions

The simplest potential to describe the Van der Waals interactions, neglecting attractive forces, is the hard-sphere (HS):

$$u^{\text{HS}}(r_{ij}) = \begin{cases} \infty & r_{ij} \leq \sigma, \\ 0 & r_{ij} > \sigma, \end{cases} \quad (1)$$

where σ is the site diameter and r_{ij} is the site–site distance. A slightly more detailed alternative to the HS potential is the soft-sphere (SS) potential:

$$u^{\text{SS}}(r_{ij}) = \begin{cases} \zeta(\sigma/r_{ij})^v & r_{ij} \leq \sigma, \\ 0 & r_{ij} > \sigma, \end{cases} \quad (2)$$

where v is a parameter usually chosen to be an integer number and ζ is a measure of the magnitude of the repulsive interaction. The square-well (SW) potential is the simplest model that considers both repulsion and attraction:

$$u^{\text{SW}}(r_{ij}) = \begin{cases} \infty & r_{ij} \leq \sigma, \\ -\varepsilon & \sigma < r_{ij} \leq \lambda\sigma, \\ 0 & r_{ij} > \lambda\sigma, \end{cases} \quad (3)$$

where ε is a measure of the attractive interaction and λ is some multiple of the hard-sphere diameter. Another simple potential that includes a physical description of dispersion is the Sutherland potential:

$$u^{\text{Su}}(r_{ij}) = \begin{cases} \infty & r_{ij} \leq \sigma, \\ -\varepsilon(\sigma/r_{ij})^6 & r_{ij} > \sigma. \end{cases} \quad (4)$$

The HS, SS, SW, and Sutherland potentials are highly idealized approximations that are nowadays rather used for the development of liquid state theories.

The most popular effective pair potential representing the Van der Waals interactions is the Lennard-Jones (LJ) potential, which was given in a general form by Mie [14]:

$$u^{\text{Mie}}(r_{ij}) = \frac{\varepsilon}{n-m} \left(\frac{n^n}{m^m} \right)^{\frac{1}{n-m}} \left[\left(\frac{\sigma}{r_{ij}} \right)^n - \left(\frac{\sigma}{r_{ij}} \right)^m \right], \quad (5)$$

where σ and ε are the size parameter and the energy well-depth, respectively. For the dispersive term, $m = 6$ is specified because of its physical significance. For the repulsive term, with little theoretical justification, $n = 9 \cdots 16$ is usually employed. The most common form is the LJ 12–6 potential ($n = 12$, $m = 6$):

$$u^{\text{LJ}}(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]. \quad (6)$$

The choice of the exponent $n = 12$ has rather computational than physical reasons, because it is simply the square of the dispersion term.

There are also many variations of the LJ 12–6 potential. One example is the computationally inexpensive truncated and shifted Lennard-Jones potential (TSLJ), which is commonly used for molecular simulation studies in which large molecular ensembles are regarded, e.g., for investigating condensation processes [15, 16]. Another version of the LJ potential is the Kihara potential [17], which is a non-spherical generalization of the LJ model.

One weakness of the LJ potential is the lack of a realistic description of repulsion, which originates from the Pauli exclusion principle. The Buckingham exponential-6 potential takes the actual exponential decay into account [18]:

$$u^{X6}(r_{ij}) = \begin{cases} \frac{\varepsilon}{1 - 6/\alpha} \left[\frac{6}{\alpha} \exp \left[\alpha \left(1 - \frac{r_{ij}}{R} \right) \right] - \left(\frac{R}{r_{ij}} \right)^6 \right] & r_{ij} > R_{\max}, \\ \infty & r_{ij} \leq R_{\max}, \end{cases} \quad (7)$$

where α is the repulsive steepness factor, ε is the well-depth located at the distance R , and $R_{\max}$ represents the distance of the potential false maximum. By definition, the Buckingham potential is set to infinity for $r_{ij} \leq R_{\max}$ in order to avoid an unphysical behavior of the potential at short distances which is only due to mathematical reasons.

Another potential function that describes the dispersive forces due to instantaneous polarities arising from fluctuations in the electron clouds is the Drude model series expansion [19]:

$$u^D(r_{ij}) = -\frac{C_{ij}^{(6)}}{r_{ij}^6} - \frac{C_{ij}^{(8)}}{r_{ij}^8} - \frac{C_{ij}^{(10)}}{r_{ij}^{10}} - \dots, \quad (8)$$

where all coefficients $C_{ij}^{(n)}$ are positive, implying an attractive interaction. The first term of the expansion considers instantaneous dipole–dipole interactions, while the higher order terms include instantaneous dipole–quadrupole, quadrupole–quadrupole, etc., interactions.

As computational resources improved, force fields were introduced which are parameterized exclusively based on ab initio calculations [20–24]. Different analytical site–site potential functions are employed, e.g., the Tang and Toennies potential [25]:

$$u^{TT}(r_{ij}) = A_{ij} \exp(-\alpha_{ij} r_{ij}) - f_6 \frac{C_{ij}^{(6)}}{r_{ij}^6} - f_8 \frac{C_{ij}^{(8)}}{r_{ij}^8}. \quad (9)$$

The damping functions f_6 and f_8 account for the influence of the charge overlap on the dispersion potential in the region of the potential well and are defined by [25]:

$$f_n(r_{ij}) = 1 - \exp(-b_{ij} r_{ij}) \sum_{k=0}^n \frac{(b_{ij} r_{ij})^k}{k!}, \quad (10)$$

where b_{ij} is a parameter that determines the effective damping length.

2.1.2 Combining Rules

The definition of different sites, particularly in the case of mixtures, implies that interaction parameter sets for unlike site pairs are required. If sufficient data – either experimental or theoretical (ab initio) – are available, these rules can be abandoned completely. The main reason for using combining rules is to avoid a huge number

of additional parameters for unlike atoms. Thus, many combining rules have been proposed, which are empirical or based on mathematical and physical intuition [26–31]. Most of the combining rules for the LJ potential parameters rely on pure component data alone. However, some combining rules incorporate additional molecular information like polarizability, ionization potential, or diamagnetic susceptibility. A review and detailed description of LJ combining rules can be found, e.g., in [32–34].

Some force fields use the geometric mean (Berthelot rule) for both LJ parameters:

$$\begin{aligned}\sigma_{ij} &= \sqrt{\sigma_i \sigma_j}, \\ \varepsilon_{ij} &= \sqrt{\varepsilon_i \varepsilon_j}.\end{aligned}\tag{11}$$

However, by far the most commonly employed LJ combining rule is that of Lorentz [35] and Berthelot [36]:

$$\begin{aligned}\sigma_{ij} &= \frac{1}{2}(\sigma_i + \sigma_j), \\ \varepsilon_{ij} &= \sqrt{\varepsilon_i \varepsilon_j}.\end{aligned}\tag{12}$$

The use of the arithmetic mean for the unlike size parameter was proposed by Lorentz motivated by the collision of hard spheres; on the other hand, the geometric mean for the unlike energy parameter was proposed with little physical argument by Berthelot. Therefore, it is not surprising that this combining rule often leads to inaccurate mixture properties [34, 37, 38].

An effective approach is to provide the Lorentz–Berthelot combining rule with at least one extra parameter that can be adjusted to some experimental data of the mixture. A modification that is adequate for the description for the unlike LJ parameters for vapor-liquid equilibria [34] is

$$\begin{aligned}\sigma_{ij} &= \frac{1}{2}(\sigma_i + \sigma_j), \\ \varepsilon_{ij} &= \zeta \sqrt{\varepsilon_i \varepsilon_j}.\end{aligned}\tag{13}$$

This modified rule was successfully applied to vapor-liquid equilibria of numerous mixtures [39–41]. Here, the binary parameter ζ was adjusted to one experimental data point for vapor pressure or Henry’s law constant of the studied binary mixture.

2.1.3 Electrostatic Interactions

The electrostatic interactions of ionic and polar molecules in the form of charges or multipoles contribute significantly to the potential energy. For an exact description of a typical charge distribution, a large set of electric moments is required [42]. However, often just dipoles and quadrupoles are taken into account, since they are usually the most significant [19]. Because the multipole expansion at one site converges slowly,

an alternative approach is to distribute fictitious point charges, dipoles, or quadrupoles throughout the molecule to cover the multipole moments [11, 43].

Electrostatic interactions can also be represented by a distribution of point charges, an approach used in numerous force fields. Both representations (point charges and atomic multipoles) are strictly valid only at long range. The electrostatic interaction which acts between a pair of point charges q_i and q_j is described by Coulomb's law:

$$u^{qq}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (14)$$

where $\epsilon_0 = 8.854187817 \times 10^{-12}$ F/m is the permittivity of the vacuum.

The interaction potential between two dipoles μ_i and μ_j is

$$u^{\mu\mu}(r_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_j) = \frac{1}{4\pi\epsilon_0} \frac{\mu_i \mu_j}{r_{ij}^3} \cdot f^{\mu\mu}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j), \quad (15)$$

and the one between two quadrupoles Q_i and Q_j is given by:

$$u^{QQ}(r_{ij}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_j) = \frac{1}{4\pi\epsilon_0} \frac{3Q_i Q_j}{4r_{ij}^5} \cdot f^{QQ}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_j). \quad (16)$$

f^{QQ} and $f^{\mu\mu}$ are expressions for the dependency of the electrostatic interactions on the orientations $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_j$ of the molecules i and j . Their definition and the potential functions acting between higher order multipoles, as well as the cross-interactions between different polarities like charge–dipole or dipole–quadrupole, can be found, e.g., in [44].

2.1.4 Polarization

Polarization is a response of the spatial electronic charge distribution of a molecule to an external field, e.g., induced by neighboring polar molecules [19]. The energy contribution due to polarization is accounted for by different methods; reviews on this topic can be found in [45–47]. A common approach is to include induced point dipoles for each atom [48]. An induced atomic dipole is given by [19]:

$$\boldsymbol{\mu}^{\text{ind}} = \alpha \mathbf{V}_f, \quad (17)$$

and the interaction energy is

$$u^{\text{ind}}(\alpha, V_f) = - \int_0^{V_f} d\mathbf{V}_f \alpha \mathbf{V}_f = -\frac{1}{2} \alpha V_f^2, \quad (18)$$

where α is the atomic polarizability, which is usually assumed to be isotropic. The electric field $\mathbf{V}_f$ is the sum of the fields due to permanent and induced dipoles acting on the atoms.

An induced dipole can also be modeled by a charge fixed to an atom to which an opposite massless movable charge is attached via a harmonic potential. This simple model is called Drude oscillator or charge-on-spring [47]. An alternative route to model polarizability is the fluctuating charge model [49]. This method uses the principle of electronegativity equalization, which ensures that atomic chemical potentials are constant throughout the molecule. Hence, the charges are fluctuating variables which respond to their environment. The charges flow between the atoms until the instantaneous electronegativities of the atoms are balanced [45]. In this context, the charges are replaced by dynamic charge distributions whose interactions are calculated by a Coulomb integral expression [19].

2.1.5 Hydrogen Bonding

In addition to the interactions mentioned above, other types of attractive interactions can also be observed, e.g., the widespread hydrogen bond. Hydrogen bonding occurs when a hydrogen atom, which is covalently bonded to an electronegative atom A, is able to approach closely to another electronegative atom B with a relatively strong attractive interaction. A hydrogen bond can be represented as $A-H\cdots B$. Therein, A and B are usually of the chemical type O, N, S, or halogens. However, weakly electronegative atoms like C can also be bonded to an H atom that acts as a proton donor, e.g., in formic acid [50].

One common approach to represent hydrogen bonding in force fields is based on point charges superimposed on LJ 12–6 sites, such as, e.g., in the SPC/E water model. The charges on the electronegative and hydrogen sites provide the electrostatic forces between molecules, while Lennard–Jones interactions between electronegative sites provide short range repulsion to balance the electrostatic attraction and determine the size of the molecule. In this type of model, hydrogen bonds are purely electrostatic and arise because the hydrogen sites are near the periphery of the molecule. Figure 1 illustrates a hydrogen bonding group composed of one LJ site and two point charges. The negative point charge (q_-) coincides with the LJ site, while the positive point charge (q_+) is positioned eccentrically with a distance h to the LJ site.

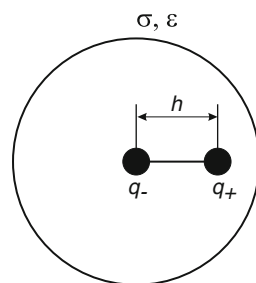


Fig. 1 Hydrogen bonding group composed of one Lennard-Jones site (σ, ϵ) and two point charges (q_-, q_+)

Hydrogen bonds can also be explicitly modeled, replacing the LJ 12–6 term between hydrogen bonding atoms by an empirical hydrogen bonding potential function that reproduces the hydrogen bonding distance and energy [19]. An example is a modification of the LJ 10–12 potential:

$$u^{\text{HB}}(r_{ij}, \theta_{\text{BHA}}) = \left(\frac{C_{ij}}{r_{ij}}\right)^{12} - \left(\frac{D_{ij}}{r_{ij}}\right)^{10} \cos^4(\theta_{\text{BHA}}), \quad (19)$$

where C_{ij} and D_{ij} are the repulsive and attractive parameters and θ_{BHA} is the angle between the atoms of the hydrogen bond (A–H...B).

2.2 Intramolecular Interactions

There are several types of intramolecular interactions, also called bonded or valence interactions, which contribute to the potential energy, i.e., bond stretching, bond angle bending, dihedral angle motion, improper angle bending, etc. These are not exclusively pair interactions, but include three- and four-body interactions as well. The parameters of the intramolecular potentials are typically fitted to reproduce geometries, vibrational frequencies, and energy profiles from ab initio calculations [9].

2.2.1 Bond Stretching

Bond stretching potentials describe the change in potential energy with the bond distance between two neighboring sites. Bond stretching is frequently represented by a harmonic potential. Thereby, analogously to Hook's law, the sites are connected by an ideal spring:

$$u_{\text{bond}}^{\text{har}}(r_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - r_0)^2, \quad (20)$$

where k_{ij} is the bond stretching force constant and r_0 is a reference bond length. This reference value is not the equilibrium length, but the bond length reached when all the other force field terms are set to zero [19]. However, physically, bond stretching does not exhibit a harmonic potential. Thus, anharmonic bond stretching potentials are also used, the simplest one adding a cubic term to (20). Other examples of anharmonic potentials are the Simon–Parr–Finland potential [51] or the Morse potential [52]:

$$u_{\text{bond}}^{\text{Mor}}(r_{ij}) = D[1 - \exp(-\alpha r_{ij})]^2, \quad (21)$$

where D is the energy well-depth and α is related to the stretching force constant of the bond. The Morse potential is more suitable to describe bond stretching than the simple harmonic potential [53]. Note that (20) is a good approximation of (21) in the case of small r_{ij} .

2.2.2 Angle Bending

Angle bending interactions occur when an angle formed by three consecutive sites is perturbed from its equilibrium value. Several potentials are employed to describe this energy contribution. However, the majority of force fields is based on the harmonic potential [54]

$$u_{\text{angle}}^{\text{har}}(\theta) = \frac{1}{2}k_{\theta}(\theta - \theta_0)^2 \quad (22)$$

or the trigonometric potential

$$u_{\text{angle}}^{\text{cos}}(\theta) = \frac{1}{2}k_{\theta}(\cos(\theta) - \cos(\theta_0))^2, \quad (23)$$

where θ is the angle formed by three consecutive sites; cf. Fig. 2, and θ_0 is its equilibrium value. The bending force constant k_{θ} is typically smaller than the bond force constant k_{ij} , because the energy required to distort an angle from its equilibrium value is lower than that required to stretch a bond [19].

2.2.3 Torsional Rotation Terms

Many of the major changes in molecular conformations are due to bond rotations. The torsion interactions account for the rotation around bonds of four adjacent sites or the motion of dihedral angles. The torsional potentials are 2π -periodic and symmetric at 0 and π . For alkanes, the Ryckaert and Belleman [55] torsional potential is often used:

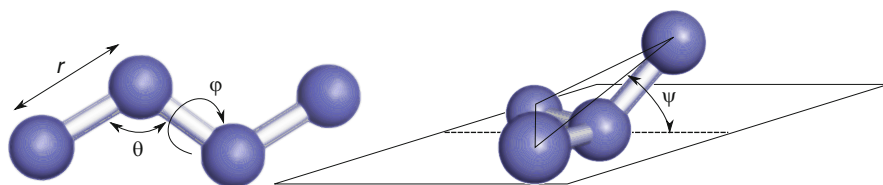


Fig. 2 Schematic representation of the intramolecular coordinates: bond length r , bending angle θ , torsional dihedral angle φ , and improper dihedral angle ψ

$$u_{\text{torsion}}^{\text{RB}}(\varphi) = \sum_n c_n \cos^n(\varphi), \quad (24)$$

where c_n are the dihedral force constants of order n . An equivalent torsional potential is based on the Fourier cosine series expansion:

$$u_{\text{torsion}}^{\text{F}}(\varphi) = \sum_n \frac{1}{2} V_n (1 + \cos(n\varphi - \delta_n)), \quad (25)$$

where φ is the dihedral angle as shown in Fig. 2. V_n are the torsional rotation force constants, δ_n the phase factors, and n the multiplicity or number of function minima upon a rotation of 2π . The specified number of terms in the series expansion varies for different force fields. Common choices are the first three terms of the expansion and terms with selected multiplicity from one to six [53].

2.2.4 Improper Torsion

A special type of torsional potential is employed to enforce geometrical constraints like planarity, e.g., in aromatic rings, or to prevent transitions between chiral structures. This potential is usually referred to as improper torsion or out-of-plane bending. Improper torsion acts between four atoms in a branched structure. There are several approaches to describe this potential. E.g., to maintain the improper dihedral at 0 or π , the torsional potential of the form

$$u_{\text{improper}}^{2\pi}(\varphi) = V_n (1 - \cos(2\varphi)) \quad (26)$$

can be used. Another route to incorporate the out-of-plane bending motion is to define an angle ψ between a bond from the central atom and the plane defined by the central atom and the other two atoms; cf. Fig. 2. With this definition, a harmonic potential can be constructed:

$$u_{\text{improper}}^{\text{har}}(\psi) = \frac{1}{2} k_\psi (\psi - \psi_0)^2, \quad (27)$$

where ψ is the improper angle and ψ_0 its equilibrium value. k_ψ is a constant that determines the stiffness of the potential.

2.2.5 Valence Coordinate Cross Terms

Some force fields include cross terms to account for the coupling between different intramolecular interactions. E.g., it has been found that, upon decrease of a bond

angle θ , adjacent bonds stretch to reduce the interactions between the atoms forming the bond. Only few force fields include such cross terms, because it was found that they are rarely important. Cross terms are usually a function of two interactions like bond–bond, bond–angle, bond–torsion, or angle–torsion, but terms containing more than two interactions can also be used. Cross terms are important to cover vibrational spectra, but do not significantly affect structural or thermodynamic properties [56]. Force fields can be classified depending on whether or not they include cross terms. Various forms of cross terms can be found in [19] that are not further discussed here.

2.2.6 1–4 Interactions

Van der Waals interactions were mentioned as intermolecular interactions. However, in many force fields Van der Waals and electrostatic interactions are also used to describe the intramolecular interactions between different sites of the same molecule that are separated by three (1–4 potential) or more bonds. Usually, the intramolecular 1–4 potential is scaled for both the LJ and coulombic contributions by an empirical factor, depending on the force field.

3 Force Field Parameterization

In the past, force fields were parameterized based only on experimental data; nowadays, most modern force fields include substantial quantum chemical information. According to the nature of the data used for parameterization, force fields can be classified as *ab initio*, semi-empirical, and empirical. Simple potentials, e.g., for argon, which require few parameters, can be fitted exclusively to macroscopic experimental data; however, more complex force fields have numerous parameters and thus depend heavily on *ab initio* data. This chapter gives an introduction to the present state-of-the-art in this field. Attention is given to the way modeling and simulation on the scale of molecular force fields interact with other scales, which is mainly by parameter inheritance. Parameters are determined both bottom-up from quantum chemistry and top-down from experimental data.

In principle, every quantity that can be predicted from force field calculations can also be used for its parameterization. The choice of the properties taken as optimization target may depend on the intended application. However, if the target properties are suitably chosen, force field models often show powerful predictive capabilities. This is due to the fact that they reasonably separate the different types of intermolecular interactions and are thus able to account for the interplay of interaction energy and structure of the fluid, which is generally a weak point in phenomenological approaches. Force fields for applications in the chemical industry should be developed, including data on the liquid density as well as on entropic properties, namely phase equilibria [57]. This is in line with more than 100 years

experience from phenomenological thermodynamics which shows that, for characterizing a pure fluid, its vapor pressure curve is of prime importance. It may, however, also be desirable to include other properties like transport coefficients in the parameter optimization. Mathematically, a multi-objective optimization problem has to be solved. However, because of parameter correlations [58], quite different parameter sets may reproduce a given set of target data with satisfactory accuracy.

There are several methods by which to perform a force field parameterization. In the trial and error approach, the parameters are gradually refined to fit the target data better. However, this is inefficient and difficult because of parameter coupling. It is more reasonable to use dedicated fitting algorithms to describe optimally the target data. Over the last few years, numerous algorithms have been developed to facilitate automated force field parameterization on the basis of thermodynamic target properties [59–62] and quantum chemical information such as energy surfaces [63–65].

3.1 *Ab Initio* Parameterization

Ab initio quantum mechanics (QM) can be used to calculate a wide range of molecular and structural properties and, additionally, to derive properties that depend on the electronic structure. Ab initio data from QM calculations can be employed in different stages of force field parameterization. They can be adopted directly into the force field, taken as target data or as initial values in optimization procedures, and may also be used for force field validation.

Bearing in mind that the quality of a force field depends on the quality of the data used for optimization, a benchmark analysis of the QM calculations should be performed, because ab initio data are only reliable when sufficiently high levels of theory and large basis sets are used. However, such QM calculations are computationally very demanding. Therefore, with current computer capabilities, system sizes that can be handled are limited to up to approximately 10^3 non-hydrogen atoms.

Ab initio calculations are mostly used for molecular geometries and intramolecular interaction parameters [66–68]. However, QM calculations can also be employed to determine parameters of the intermolecular potential, e.g., for the polar interactions.

3.1.1 Van der Waals Interactions

The Van der Waals interactions are not easily accessible with ab initio methods, because the characterization of short-range intermolecular interactions requires a much higher level of theory than molecular structure or conformational energies. Moreover, at least a reasonably complete two-body interaction energy landscape is

required. Thus, a large number of molecular separations and mutual orientations must be considered, which is computationally very demanding. Once the appropriate points of the energy landscape are obtained, they can be fitted to an analytical function [69]. Note that the liquid behavior is not well reproduced by *ab initio* calculations, since only small clusters can be handled [70]. A review of QM methods used for the calculation of interaction energies and potential energy sampling is given in [69].

3.1.2 Electrostatic Interactions

Electrostatic properties of molecules can be determined from the electron density distribution obtained by QM. Different methods have been proposed for this end. E.g., atomic charges can be estimated using different partitioning methods like Mulliken and Löwdin population analysis [71, 72], the charge model 2 (CM2) formalism [73], natural population analysis (NPA) [74], or the theory of atoms in molecules (AIM) [75]. A comparison of these methods for the calculation of atomic charges can be found, e.g., in [76]. Atomic charges calculated by population methods are often considered to be inappropriate for force field parameterization [19]. The most common approach is to derive the atomic charges from the electrostatic potential (ESP), applying either semi-empirical density functional theory (DFT), Hartree–Fock (HF), or post HF methods [77]. The ESP is a QM observable which can be determined from wave functions. In this method, atomic charges are fitted to the calculated ESP for a series of points in a three-dimensional spatial grid surrounding the molecule. The fitting procedure is performed with the constraint that the sum of the charges equals the net charge of the molecule. The positions where the potential is evaluated are often chosen just outside the atomic Van der Waals radii, because the accuracy of electrostatics is most important there. Different methods consider different sampling points where the ESP is calculated, i.e., the distance from the Van der Waals surface [19]. The CHELP [78] method considers spherical shells extended to 3 Å from the Van der Waals surface, whereas the CHELPG [79] method contemplates a cubic grid of points extended to 2.8 Å. A restrained electrostatic potential (RESP) [80] fit is often used to include restrictions to the obtained charges, e.g., to restrain charges in buried atoms. RESP can be employed to fit partial charges to the ESP of a single or multiple conformers [77]. There are various difficulties with the ESP fitting approach, like conformation, basis set dependency, and the presence of buried atoms. The inclusion of multiple conformations in the fitting procedure can be employed to overcome these problems [81]. A comparison of some commonly applied schemes can be found in [82].

The second order Møller–Plesset (MP2) perturbation theory is often adequate in terms of accuracy and efficiency for describing the ESP [10]. It is generally considered that the 6-31G* basis set gives reasonable results [19]. This basis set results in dipole moments that are 10–20% larger than expected in the gas phase, which is desirable for deriving charges for liquid phase simulations [80]. More

advanced ab initio methods, e.g., the coupled cluster method together with correlation consistent basis sets, can also be used for such calculations [83].

Electrostatic multipole moments of molecules, i.e., dipoles, quadrupoles, or octupoles, can also be obtained from QM wave functions. Methods like distributed multipole analysis (DMA) [84] or AIM [85] assign multipole moments to each atom or to specified sites of a molecule. The DMA method estimates multipole moments from QM wave functions and the highest obtained multipole moment depends on the basis set used. There are no limitations in this method on number or position of the multipoles; anisotropic effects due to lone pairs or π electrons can also be considered.

A simpler approach, typically employed for small symmetric molecules, is to estimate ideal point multipoles by integration over the orbitals resulting from the calculated electron density distribution. The accuracy of the calculated moments is highly dependent on the basis set, electron correlation, and molecular geometry [19]. The MP2 level of theory with the 6-31G* polarizable basis set is broadly applied in such calculations. In order to save computational effort, MP2 is often executed as a single point calculation for a geometry determined on the basis of a lower level of theory.

In condensed phases, the mutual polarization of solute and solvent molecules should be considered. This can be done by placing a single molecule into a cavity that is surrounded by a dielectric continuum and assigning the dielectric constant of the liquid to it [86]. Thus, the molecule in the cavity induces polarization in the surrounding dielectric continuum, which in turn interacts with the electron density of the molecule. There are numerous techniques of varying complexity; a review can be found e.g. in [87]. One of the pioneering techniques is the self consistent reaction field (SCRF) [88, 89] approach. Some variations of this method treat the continuum solvent as a conductor, such as in the conductor-like screening model (COSMO) [90] or the polarizable continuum model (PCM) [87]. Another rather simple approach to account for condensed phase polarization is the multipole scaling procedure [80, 91].

3.1.3 Intramolecular Interactions

The geometric parameters of force fields, i.e., reference bond lengths and bond angles, are commonly assigned according to equilibrium molecular geometries determined by QM, combined with an energy minimization algorithm. The agreement between ab initio and experimental equilibrium geometries increases with the size of the basis set and the level of theory. However, the HF level of theory with a relative small basis set, such as 6-31G, is sufficient to obtain good results [60, 86, 92]. Fortuitously, the STO-3G basis set often performs well with respect to molecular geometry, despite its deficiencies. In general, the bond lengths predicted by the STO-3G basis set are too long, while those obtained with the 6-31G basis set are too short [19]. As an alternative QM approach, DFT, using gradient corrected and hybrid methods, can be applied, since it is known to achieve excellent results

for equilibrium geometries [10, 93]. An important example is the Becke's three-parameter density functional hybrid method combined with the Lee, Yang, and Parr gradient-corrected correlation functional B3LYP [94].

QM is widely used to calculate relative energies of conformation sets and energy barriers between them. Hence, bond length, bond angle, and torsional potential terms can be fitted to reproduce intramolecular energy surfaces, the relative energy of conformational pairs, or rotational energy profiles. The variation of energy for different configurations can be calculated quite accurately with relatively small basis sets. The rotational energy profiles are often taken as a basis to determine the torsional interactions. For this purpose, the energy of a series of molecular structures generated by bond rotation is obtained from ab initio calculations. The torsional potential is fitted to the resulting energy profile together with the Van der Waals and electrostatic potentials [19]. Both HF and MP2, together with the 6-31G basis set, are often employed for such calculations [95]. It should be noted that DFT with the B3LYP functional performs rather poorly for intermolecular interactions and conformational energies [10].

3.2 Empirical Parameterization

Due to the difficulties of QM methods to describe correctly condensed phase behavior, Van der Waals parameters and atomic point charges of molecular models are often adjusted to reproduce experimental data on macroscopic properties of the liquid state. Usually, they are fitted to thermodynamic properties determined by means of molecular dynamics (MD) or Monte Carlo (MC) simulations.

3.2.1 Intermolecular Interactions

Intermolecular potential parameters can be optimized to different types of experimental data. For engineering applications, liquid density and liquid enthalpy are very often used. E.g., the liquid density strongly depends on the LJ size parameter σ , whereas the enthalpy of vaporization strongly depends on the LJ energy well depth ε [60]. Therefore, intermolecular parameters are frequently adjusted to experimental data on both of these quantities, as in the OPLS force field [96]. The vapor pressure is even more sensitive to the intermolecular potential parameters so that, particularly in recent years, it was chosen together with the saturated liquid density and the heat of vaporization to devise numerous generic force fields of interest to chemical engineers like TraPPE, AUA, and NERD. The latter strategy was also used for the development of a wide variety of specific molecular models for engineering applications [97–99].

Many other physical properties may also be taken as targets for parameter optimization of Van der Waals and electrostatic potentials: second virial coefficient

[100, 101], critical temperature [102], free energy of hydration [103], self-diffusion coefficient [104, 105], shear viscosity [105–107], radial distribution functions [57, 108], or multipole moments [109].

3.2.2 Intramolecular Interactions

Equilibrium geometries of molecules can be derived from gas-phase experiments, such as electron diffraction and microwave spectroscopy [60]. Raman and infrared vibrational frequencies can also be used to determine force constants for bond stretching and angle bending. If available, experimental data on relative configuration stabilities and barrier heights can be used to parameterize torsional and improper potential terms [19].

4 Force Field Families

Numerous force fields with different degrees of sophistication are in use today; however, none of them is universally accepted. A molecular force field is generally designed and parameterized to reproduce certain properties and should be able to predict a wide range of thermodynamic properties for different thermodynamic conditions. Force fields can be transferable (using the same set of parameters to model a variety of related compounds) or specific (using distinct sets of parameters for each molecule). Beside the choice of the potential functions of the force field, another important decision is whether or not to represent all atoms explicitly. In this context, force fields can be divided into all-atom, united-atom, and coarse grained types.

4.1 Transferable Force Fields

In transferable force fields, the parameters for a given functional group are deemed transferable between different molecules. There are numerous transferable force field families, which were developed for different applications. Below, some families relevant for chemical engineers are described: optimized potentials for liquid simulations (OPLS) [96, 110–112], transferable potential for phase equilibria (TraPPE) [113–122], optimized potential model for phase equilibria (OPPE) [59, 68, 107, 123–130], Nath, Escobedo, and de Pablo (NERD) force field [100, 131–133], and the GIBBS99 exponential-6 force field [18, 134], (cf. Table 1). Many force fields families are continuously being improved and extended to include new types of compounds; thus numerous versions are available.

Table 1 Some important characteristics of the different united-atom force field families for alkanes

Force field family	Van der Waals potential	Bond stretching	Angle bending	Torsional potential	Combining rule
OPLS-UA	LJ 12–6	No	No	Yes	Berthelot
OPLS-UA/AMBER	LJ 12–6	Yes	Yes	Yes	Berthelot
TraPPE	LJ 12–6	No	Yes	Yes	Lorentz–Berthelot
OPPE-AUA	LJ 12–6	No	Yes	Yes	Lorentz–Berthelot
NERD	LJ 12–6	Yes	Yes	Yes	Lorentz–Berthelot
GIBBS99	Buckingham	No	Yes	Yes	Lorentz–Berthelot

4.1.1 OPLS Force Field

The optimized potentials for liquid simulations (OPLS) force field can be divided into the OPLS-UA (united-atom) [96, 110–112] and the OPLS-AA (all-atom) [57, 67, 92, 103, 135–139] versions. Among the two, the OPLS-UA force field is predominantly used for engineering applications, mainly because it is computationally cheaper than the all-atom version. The OPLS-UA force field is available for hydrocarbons [96], amides [110], peptides [110], alcohols [111], or proteins [112]. The OPLS-AA force field was parameterized for small organic molecules and is intended for biochemical applications. The parameters of the OPLS-AA force field are available for a broader range of functional groups and molecules. Thus, besides hydrocarbons [103] and alcohols [57], parameters can be found for thiols [57], sulfides [57], ketones [57], amides [57], amines [139], pyrrole [138], furan [138], diazoles [138], oxazoles [138], proteins [67], and carbohydrates [92], among others.

The functional forms of both OPLS force field families are similar. In the original OPLS-UA force field, the only intramolecular degrees of freedom that were taken into account were torsions. Later on, the OPLS-UA force field was merged with the description of bond stretching and angle bending from the AMBER force field to yield the OPLS-UA/AMBER force field for peptides and proteins [57]. In OPLS-AA, bond stretching and angle bending were described by harmonic potentials [(20) and (22)]. The OPLS-UA and OPLS-AA force fields consider the energetic contribution of the torsional motion by a Fourier series truncated after the third term (25). The Van der Waals interactions are represented by the LJ 12–6 potential and electrostatics is represented by point charges. The LJ potential is not only used to describe the intermolecular interactions, but also the interactions between different sites of the same molecule that are separated by three (1–4 potential) or more bonds. The intramolecular 1–4 potential is scaled for both the LJ and coulombic contributions by an empirical factor. The OPLS-AA force field uses a scaling factor of 1/2 for both potentials, while the OPLS-UA/AMBER force field uses factors of 1/2 and 1/8, respectively. Also, the unlike LJ parameters are defined by the geometric mean for the size and energy parameter, cf. (11). Different OPLS force field versions were optimized applying different methods. Geometrical parameters, such as for bond stretching and angle bending, were taken from other force fields (AMBER94 [66], CHARMM [91, 95, 140]), fitted to experimental molecular structures or to ab initio calculations at the HF/6-31G* level of theory. The rotational terms of the OPLS-UA

force field were derived from rotational potentials obtained by molecular mechanics (MM) simulations, while in the most recent versions of the OPLS-AA force field, the torsional potentials were fitted to *ab initio* calculations at the MP2 level of theory with the 6-31G* or even the correlation consistent polarized triple zeta (cc-pVTZ) basis set. The parameterization of the intermolecular interactions was performed to reproduce saturated liquid density and enthalpy of vaporization. In some versions of the OPLS-AA force field [137, 138], the partial charges were fitted to the *ab initio* ESP with the CHELPG [79] procedure. Since the OPLS-UA force field for hydrocarbons was parameterized considering primarily short alkane chains, the deviations to experimental vapor-liquid equilibrium data become more significant for larger chain lengths [141].

4.1.2 TraPPE Force Field

The transferable potential for phase equilibria (TraPPE) was originally developed by devising a united-atom representation for the alkyl segments (TraPPE-UA). However, a TraPPE force field with explicitly considered hydrogens (TraPPE-EH) [119] was also formulated. The TraPPE force field is available for a large number of compound families, including linear and branched alkanes [117, 118], alcohols [114], ethers [120], ketones [120], glycols [120], amines [122], amides [122], thiols [115], aromatics [119, 121], and acrylates [116], among others. The TraPPE force field takes the intermolecular interactions into account by the LJ 12–6 potential (6) and coulombic terms (14). For the unlike LJ interactions, the standard Lorentz–Berthelot combining rule (12) is assumed. The intramolecular interactions covered by this force field are: angle bending on the basis of a harmonic potential (22) and torsional motion expressed as a set of cosine series (25) and a harmonic improper dihedral potential (27). The bond lengths are fixed and thus bond stretching is not taken into account. Usually, the intramolecular parameters for angle bending were transferred from the AMBER94 [66] force field and the dihedral parameters were taken from the OPLS-UA force field. The LJ and point charge parameters were fitted to reproduce experimental vapor-liquid coexistence data. The TraPPE force field reproduces the saturated liquid density of linear alkanes with a mean accuracy of approximately 1%, which has to be seen in the light of traditional predictive methods like Lee–Kesler that have an accuracy of 2–3% [9]. The TraPPE force field reproduces vapor pressure, saturated vapor density, and critical point with more significant deviations to the experiment [9]. It does not reproduce the second virial coefficient well and consistently underpredicts the shear viscosity of short chain paraffins [106].

4.1.3 OPPE-AUA4 Force Field

The anisotropic united-atom optimized potential model for phase equilibria (OPPE) force field is an elaboration of the anisotropic united-atom (AUA) force

field, initially proposed by Toxvaerd [142, 143] that was further developed by Ungerer and coworkers [130]. This force field is currently available for *n*-alkanes [130], olefins [59], alcohols [68], polyalcohols [144], amines [123], amides [123], nitriles [127], sulfides [126], thiols [126], ketones [128], aromatic hydrocarbons [124, 129], or polycyclic aromatics [125]. The major novelty of AUA force fields was that the force center is spatially located between the carbon and the hydrogen atoms of the represented molecular group. The intermolecular interactions were described by the LJ 12–6 potential and point charges [(6) and (14)]. The Lorentz–Berthelot combining rule (12) was used for the unlike LJ parameters. As in the TraPPE force field, the bond lengths were kept fixed. Angle bending was modeled by a trigonometric potential (23) and the torsional potential following Ryckaert and Belleman (24). Some angle parameters were taken from the AMBER94 [66] force field and the torsional potential parameters were taken from the OPLS-UA [96] force field. In other cases, molecular geometry and electrostatic charges were determined from ab initio calculations. Usually, geometries were optimized with the B3LYP functional and the 6-311G** basis set. The partial charges were parameterized according to the procedure of Lévy and Enescu [145] to reproduce the ESP around the isolated molecule for several representative conformations using RESP [80]. The ab initio ESP of the molecules was determined at the MP2 level of theory with a 6-31G* or a 6-311G** basis set. The LJ parameters were optimized to reproduce experimental values of saturated liquid density, enthalpy of vaporization, and vapor pressure. The OPPE force field provides a good representation of the vapor pressure and a very accurate representation of the liquid density over a wide temperature range for *n*-alkanes, branched alkanes, and cycloalkanes [56]. The vapor pressure of alkanes is predicted with an average deviation to experimental data of 15%, compared to 30% for the TraPPE force field and 35% for classical prediction methods based on boiling temperature and heat of vaporization [9]. Since transport properties are not well predicted by this force field, Nieto-Draghi et al. [107] proposed a modification of the OPPE model by adjusting the parameters of the torsional potential to reproduce experimental reorientation dynamics and shear viscosity.

4.1.4 NERD Force Field

The Nath, Escobedo, and de Pablo (NERD) force field [100, 131–133] was developed to provide accurate predictions of thermodynamic properties. It is currently available for linear [100] and branched alkanes [131, 133] as well as for alkenes [132]. It has a similar functional form as the TraPPE-UA force field, but bond stretching is included. This interaction and angle bending are represented by harmonic potentials [(20) and (22)]. The torsional potential is of the form of (25), neglecting cross terms. The LJ 12–6 potential (6) is used to describe the intermolecular and intramolecular interactions between sites that are separated by more than three bonds. The LJ parameters were obtained from fits to experimental values of liquid density and second virial coefficient. Saturated liquid densities from the NERD force field are in good

agreement with experimental data. However, the vapor pressure predictions are typically slightly above experimental data at low temperatures and below experimental data at high temperatures [56], while the critical temperature is overestimated, e.g., by 7 K for short-chain alkanes (ethane and pentane) [18].

4.1.5 GIBBS99 Force Field

The GIBBS99 exponential-6 force field [18] is a united atom representation that is available for linear alkanes, cyclohexane, or benzene [134]. It differs from the NERD and TraPPE force fields in the description of the Van der Waals interactions: The Buckingham exponential-6 potential (7) was used instead of the LJ 12–6 potential (6). Similarly to the TraPPE force field, bond stretching was neglected. However, the bond length between two methyl groups of the alkane chain was not fixed as in the TraPPE force field, but depends on the molecular groups that form the bond. Angle bending was represented by the harmonic potential (22) and the torsional motion by a third order Fourier series (25). The force field parameters were fitted to critical properties and saturated densities. The GIBBS99 force field represents the vapor pressure and saturated densities for the alkanes from ethane to *n*-dodecane with average deviations of around 2%. The experimental vapor pressures of benzene and cyclohexane are reproduced with average errors of 2.6 and 1.7%, respectively [134].

4.1.6 Other Force Fields

Transferable force field families intended for biological applications are sometimes applied in chemical engineering for the simulation of large molecules like polymers. Also, some ionic liquids were parameterized in that framework [105]. Some relevant force fields are chemistry at Harvard molecular mechanics (CHARMM) [91, 95, 140], assisted model building with energy refinement (AMBER) [65, 66, 146, 147], Groningen molecular simulation (GROMOS) [148–150], condensed-phase optimized molecular potentials for atomistic simulation studies (COMPASS) [70], and consistent force field (CFF) [151–153], among many others. These force fields best reproduce the data for which their parameters were optimized. AMBER, CHARMM, and OPLS-AA overestimate the free energy of hydration of protein functional groups [154]. Several works on the comparison of various of these force field families for the simulation of proteins [155–159], deoxyribonucleic acids [160], peptides [161], carbohydrates [162], or aqueous salt solutions [163] can be found in the literature.

Some examples for transferable polarizable force fields are Drude [83], TraPPE-pol, CHARMM-FQ [164], PIPF [165–167], and AMOEBA [168]. A review on polarizable force fields can be found, e.g., in [45]. Many of the mentioned force fields for biochemical applications as well as the polarizable force fields are being

continuously developed, improved, and refined. Therefore, numerous versions of each family can be found in the literature.

4.1.7 Force Fields Comparison

Martin [169] compared the AMBER, CHARMM, COMPASS, GROMOS, OPLS-AA, and TraPPE force fields with respect to their ability to predict vapor–liquid equilibrium properties and the liquid density of small alkanes and alcohols. He concluded that the force field families performing best for fluid phase simulations are TraPPE and CHARMM. CHARMM better predicts the vapor density, while TraPPE has a higher accuracy for liquid density predictions.

TraPPE and OPPE-UA are, in our opinion, the best transferable force fields developed to date for chemical engineering applications. However, they still have some deficiencies. The capabilities of these force field families are still less explored than group contribution methods like UNIFAC in phenomenological thermodynamics.

4.2 Specific Force Fields

A force field that is carefully parameterized for a specific substance is usually more accurate than a transferable force field. Therefore, when high levels of accuracy are required, specific force fields are preferred. Most of the newer specific force fields developed for engineering applications were parameterized to reproduce experimental data on saturated liquid density and enthalpy of vaporization. The use of *ab initio* calculations gained importance in the last decade and the majority of force field developers nowadays thus makes use of QM calculations to some extent. There are numerous parameterization strategies for the development of such force fields, which depend on the availability of experimental data and the complexity of the chosen functional form. There is an immense number of specific force fields; therefore it is impossible to give a comprehensive overview here. Only a small selection will be discussed in the following to exemplify different parameterization strategies. The re-parameterization of existing or transferable force fields using a different set of experimental or *ab initio* data as in [170] will not be treated in further detail.

4.2.1 Empirical Force Fields

All transferable force fields discussed in Sect. 4.1 employ point charges to account for the molecular charge distribution, although a more accurate description of the electrostatics with higher multipole moments may be used. Hasse, Vrabec, and co-workers [102, 109] proposed a set of simple united-atom force fields for more than 70 compounds of different classes that describe the intermolecular interactions using two LJ 12–6 sites plus a point dipole (15) or a point quadrupole (16). The potential

model parameters were optimized to experimental values of critical temperature, saturated liquid density, and vapor pressure; thus, no direct information on the multipole moments or the geometry was taken into account. These force fields allow the description of vapor–liquid equilibria with an average accuracy of 4% for the vapor pressure, 0.5% for the saturated liquid density, and 3% for the enthalpy of vaporization throughout the entire temperature range from the triple point to the critical point. Furthermore, the shear viscosity and the thermal conductivity are predicted within 10% accuracy [171, 172]. It can be argued that oversimplified molecular models can be adjusted to a few experimental pure substance properties, but major deficiencies should be visible when applied to mixtures. Recently, all systems for which experimental mixture data were available containing these simple models were studied by molecular simulation [39–41] using one experimental data point to obtain the adjustable combining rule parameter ξ (13). The results of this study were very satisfactory.

4.2.2 Semi-Empirical Force Fields

Fermeglia et al. [173] proposed flexible all-atom force fields for several hydrofluorocarbons to describe vapor–liquid equilibria. They used a more complex expression for the potential energy including intramolecular interactions, i.e., bond stretching (20), angle bending (23), and torsional motion (24). The intermolecular interactions were represented by the LJ 12–6 potential (6) and partial charges (14). The parameters of the intramolecular terms were determined by geometry optimization and potential energy surface sampling. For this purpose, DFT with the BPW91 functional was employed [173]. The partial charges were obtained from fits to the ESP, while the LJ parameters were optimized to experimental data on liquid density and cohesive energy. These force fields predict the phase behavior with an average error of about 2% for saturated densities in the temperature range from 200 to 380 K.

Hasse, Vrabec, and co-workers [86, 174] presented a set of semi-empirical rigid, united-atom force fields for hazardous fluids, such as cyanide, acetonitrile, nitromethane, or phosgene. They described the intermolecular interactions with LJ 12–6 sites and point charges, point dipoles, or point quadrupoles. The geometric parameters of these force fields, i.e., bond lengths, angles, and dihedrals, were determined by ab initio calculations at the HF/6-31G level of theory. The electrostatic multipole moments (dipoles and quadrupoles) were obtained by integration over the orbitals from the electron density distribution, using QM at the MP2 level of theory with a 6-31G(d,p) basis set. The LJ parameters and point charge magnitudes were optimized to experimental vapor–liquid equilibrium data. These specific force fields describe vapor–liquid equilibrium properties with a better accuracy than the available transferable force fields [86].

A united-atom force field based on GROMOS96 [148] was proposed by Micaelo et al. [105] for imidazolium-based ionic liquids. They used ab initio calculations at the HF/6-31G* level of theory to obtain partial charges based on the single-step

RESP method. Bond lengths and bond angles were constrained to reproduce the optimum geometry from QM calculations. The LJ parameters were optimized to experimental values of shear viscosity, self-diffusion coefficient, and liquid density. A review on force fields for the simulation of imidazolium-based ionic liquids can be found in [175].

Liu et al. [176] developed force fields for guanidinium-based ionic liquids following the AMBER force field approach. The intramolecular interactions of their force fields include harmonic bond stretching and angle bending, together with torsional motions. The equilibrium bond lengths and bond angles were taken from QM calculations at the HF/6-31+G(d) level. The force constants were adjusted to vibrational frequencies obtained by *ab initio* calculations or from experiment. Single point MP2/6-31+G(d) calculations were taken to parameterize the torsional potential and QM calculations at the B3LYP/6-31+G(d) level to obtain RESP charges. The LJ parameters were transferred from the AMBER99 force field.

4.2.3 Ab Initio Force Fields

Hellmann et al. [23, 177, 178] have proposed *ab initio* force fields for several small molecules, such as helium, neon, or methane, based on the Tang and Toennies potential (9) and coulombic terms (14). With these force fields, gas phase properties like second virial coefficient, shear viscosity, thermal conductivity, or self-diffusion coefficient can be predicted extremely accurately. Typically, the generated data are within the experimental uncertainty.

Domański et al. [179] developed an *ab initio* force field for liquid carbon dioxide by fitting the LJ parameters and the coulombic terms to the potential energy surface calculated with QM at the MP2 level of theory and the 6-31G* basis set. Unfortunately, their model does not reproduce the thermodynamic behavior of the liquid state so that an empirical scaling factor had to be adjusted to experimental data.

Hloucha et al. [24] developed force fields for methanol and acetonitrile from *ab initio* calculations for the prediction of macroscopic properties. These all-atom force fields include LJ 12-6 or modified Buckingham exponential 6 sites plus partial charges. Interaction energies for many hundreds of configurations calculated via symmetry adapted perturbation theory (SAPT) were employed for the parameterization of the LJ and Buckingham terms. To cover electrostatics, the charges were fitted to the ESP from quadratic configuration interaction with single and double substitution (QCISD) calculations and the augmented correlation-consistent polarized valence double-zeta basis set (aug-cc-pVD). Despite the fact that the force field for acetonitrile yields a reasonable agreement with the experiment for vapor-liquid equilibrium properties, for methanol the saturated liquid density was strongly underpredicted and the vapor pressure was overpredicted by one order of magnitude. Cabaleiro-Lago and Ríos [20] proposed a similar *ab initio* force field for acetonitrile optimized at the MP2/6-311+G* level of theory. However, their force field gives a poor prediction of the phase behavior [24]. Further examples of *ab initio* force fields can be found, e.g., in [69].

5 Molecular Simulation Methods

Given an adequate force field, molecular simulation is in principle capable of yielding predictions of thermodynamic properties for a broad range of thermodynamic conditions. To this end, different simulation techniques can be employed, which can be divided in MD and MC. Here, some simulations tools for predicting thermodynamic properties that are important for chemical engineering, i.e., vapor–liquid equilibrium and transport properties, will be addressed briefly.

5.1 *Molecular Dynamics*

MD is a technique in which the time evolution of the molecular motions is simulated following the laws of classical mechanics. Therefore, the physical variable time must be considered explicitly. In this way, the dynamic evolution of coordinates and moments, i.e., the trajectory of the system, is calculated by numerically solving Newton's equations of motion. This trajectory, together with the associated energies and forces, leads to the static and dynamic thermodynamic properties of the studied system via statistical analysis methods. MD is also a powerful tool to understand dynamic processes at the atomistic level that involve fluids or materials [9].

In MD, a set of second order differential equations is solved by finite difference techniques. This can be done with a variety of integration algorithms, such as Verlet, velocity Verlet, Leap-Frog, or Gear predictor-corrector. Although the microcanonical (*NVE*) ensemble is the most natural one for MD simulations, generally the canonic (*NVT*) or the isobaric-isothermal (*NpT*) ensembles are applied. Particularly in chemical engineering, physical properties are needed for specified thermodynamic conditions like temperature or pressure. Several methods exist to control temperature and pressure during simulation, e.g., velocity scaling, Anderson thermostat, Berendsen thermostat, Nosé–Hoover thermostat, Nosé–Hoover chains thermostat, or Berendsen barostat. A description of these algorithms can be found, e.g., in [11, 180].

An MD simulation yields a significant amount of useful information for chemical engineering applications [11]. E.g., it is employed to study dynamic processes, like diffusion, adsorption, or glass transition. A review of MD applications can be found, e.g., in [9].

5.2 *Monte Carlo*

MC is a stochastic method that samples the configuration space of a system with a specified Hamiltonian [181]. In MC simulations, the transition between states or configurations is achieved by a random generation of a new state, evaluating a probabilistic acceptance criterion, and accepting or rejecting the perturbation.

New configurations are usually generated by displacing, removing, or adding individual molecules. The acceptance of new states is performed most commonly according to the Metropolis criterion.

In the production phase of MC simulations, all configuration-dependent properties fluctuate around constant average values that correspond to the thermodynamic equilibrium. Each state is thereby sampled with a frequency proportional to its equilibrium probability density [182]. In the canonical ensemble the probability density ρ_m^{NVT} is given by [181]:

$$\rho_m^{NVT} = \frac{\exp(-E_m/(k_B T))}{\sum_{\text{all states}} \exp(-E_m/(k_B T))}, \quad (28)$$

where k_B is the Boltzmann constant and E_m is the potential energy of a state m .

An advantage of MC is that it can be readily adapted to any ensemble [11]. Therefore, many MC ensembles have been developed for the simulation of specific systems or properties. A wide variety of MC simulation techniques can thus be found in the literature. Reviews and detailed information about MC techniques are presented, e.g., in [11, 181–185].

5.3 Methods for Determining Phase Equilibria

The calculation of vapor–liquid equilibria by molecular simulation is a longstanding and important task. In the last two decades a variety of methods for this purpose have been presented. There are, among others, thermodynamic scaling [186], histogram reweighting [187, 188], Gibbs–Duhem integration [189], NpT plus test particle method [190], grand canonical ensemble [191], grand equilibrium method [192], or the Gibbs ensemble MC method [193]. Here, some of these simulation methods will be briefly addressed. A comprehensive discussion of the different approaches can be found, e.g., in [181, 182, 194, 195].

The Gibbs ensemble MC method (GEMC) [193] was developed to sample two homogeneous coexisting phases that are in thermodynamic equilibrium but not in physical contact with each other. The pressure and chemical potential of the phases are equated by allowing the volume and the number of molecules to fluctuate between the phases, while keeping the total volume and total number of molecules constant. This ensemble is widely employed to calculate phase equilibria [18], also in combination with Gibbs–Duhem integration [189, 196]. It is also used to simulate chemical reactions in phase equilibrium [197, 198]. In the literature, some advanced methods related to this ensemble can be found, e.g., the thermodynamic scaling Gibbs ensemble [199].

In the grand canonical (GC) ensemble, a system at constant temperature, volume, and chemical potential is considered. The number of molecules is therefore allowed to fluctuate. In such simulations, molecule displacement, insertion, and deletion are attempted. From a series of several GCMC simulations, the pressure

dependence of the chemical potential in the vapor and in the liquid phase can be obtained. The coexistence condition is then found at the intersection point [181]. The number of molecules is not constant for this ensemble and the coexisting phases are simulated independently. The semigrand canonical ensemble [200, 201] was introduced to overcome the low acceptance probability of molecule insertions and deletions for liquids in the GC ensemble. Furthermore, the GC ensemble can be combined with finite-size scaling methods, e.g., to evaluate the surface tension [202, 203].

Another technique to determine the vapor–liquid equilibrium of pure substances or mixtures, which has some similarities with what is described in [190, 204–206], is the grand equilibrium method [192]. It is a two-step procedure, where the coexisting phases are simulated independently and subsequently. In the first step, one NpT simulation of the liquid phase is performed to determine the chemical potentials μ_i^l and the partial molar volumes v_i^l of all components i . These entropic properties can be determined by Widom’s test molecule method [207] or more advanced techniques, such as gradual insertion [208–210] (see below). On the basis of the chemical potentials and partial molar volumes at a specified pressure p_0 , first order Taylor expansions can be made for the pressure dependence:

$$\mu_i^l(T, x, p) \approx \mu_i^{l, \text{id}}(T) + \mu_i^{l, \text{re}}(T, x, p_0) + v_i^l(T, x, p_0) \cdot (p - p_0), \quad (29)$$

where $\mu_i^{l, \text{id}}(T)$ is the solely temperature dependent ideal contribution and the residual chemical potential is $\mu_i^{l, \text{re}}(T, x, p_0)$. Note that $\mu_i^{l, \text{id}}(T)$ does not need to be evaluated for vapor–liquid equilibrium calculations, because it cancels out when (29) is equated to the corresponding expression for the vapor. In the second step, one pseudo- μVT simulation [192] is performed for the vapor phase on the basis of (29) that yields the saturated vapor state point of the vapor–liquid equilibrium. This simulation takes place in a pseudo-ensemble in the sense that the specified chemical potentials are not constant, but dependent on the actual pressure in the vapor phase. Thus the vapor simulation rapidly converges to the saturated vapor state point during equilibration so that effectively the equilibrium chemical potentials are specified via the attained vapor pressure. The grand equilibrium method has been extensively used for the determination of vapor–liquid equilibria of hundreds of systems [40, 41].

Several methods to obtain vapor–liquid equilibria or the Henry’s law constant [39] require the accurate calculation of the chemical potential. Widom’s test molecule method [207] is the most common approach for that task. Widom’s method randomly introduces test molecules of the component for which the chemical potential is evaluated into the simulation volume, to calculate their potential energy ψ_i . The test molecules are instantly removed after the calculation and do not influence the remaining molecules. Widom’s method is applicable to MC as well as to MD simulations [56]. In the NVT ensemble, the residual chemical potential μ_i^{re} of component i is calculated by

$$\mu_i^{\text{re}} = -k_B T \ln \langle \exp(-\psi_i / (k_B T)) \rangle. \quad (30)$$

Widom's method presents problems when dealing with very dense and strongly interacting fluids, because inserted test molecules almost always overlap with "real" molecules, which leads to extremely large values for the potential energy ψ_i . These insertions contribute little information, resulting in poor statistics [56]. Therefore, advanced methods have been proposed in the literature. An example is the gradual insertion method [208–210], where a fluctuating molecule is introduced into the simulation. The fluctuating molecule undergoes a stepwise transition between non-existence and existence, which allows the determination of the chemical potential. This method has been applied successfully to vapor–liquid equilibrium calculations of numerous binary and ternary mixtures [40, 41, 174]. Many other methods, such as configurational biased insertion [211] or minimum mapping [212], have been proposed in the literature. A detailed description and comparison thereof can be found, e.g., in [213].

The Henry's law constant can be obtained from molecular simulation using several approaches [214, 215]. It is related to the residual chemical potential of the solute i at infinite dilution μ_i^∞ by [216]:

$$H_i = \rho k_B T \exp(\mu_i^\infty / (k_B T)), \quad (31)$$

where ρ is the density of the solvent.

5.4 Methods for Determining Transport Properties

Transport properties, such as diffusion coefficients, shear viscosity, thermal or electrical conductivity, can be determined from the time evolution of the autocorrelation function of a particular microscopic flux in a system in equilibrium based on the Green–Kubo formalism [217, 218] or the Einstein equations [219]. Autocorrelation functions give an insight into the dynamics of a fluid and their Fourier transforms can be related to experimental spectra. The general Green–Kubo expression for an arbitrary transport coefficient γ is given by:

$$\gamma = \frac{1}{G} \int_0^\infty dt \langle \dot{\mathbf{A}}(t) \cdot \dot{\mathbf{A}}(0) \rangle, \quad (32)$$

and the general Einstein or square displacement formula can be written as

$$\gamma = \frac{1}{2Gt} \langle [\dot{\mathbf{A}}(t) - \dot{\mathbf{A}}(0)]^2 \rangle. \quad (33)$$

Therein, G is a transport property specific factor, $\mathbf{A}$ the related perturbation, and $\dot{\mathbf{A}}$ its time derivative. The brackets $\langle \dots \rangle$ denote the ensemble average. It was shown that (33) can be derived from (32); thus both methods are equivalent [220].

In case of the self-diffusion coefficient, $\mathbf{A}(t)$ is the position vector of a given molecule at some time t and $\dot{\mathbf{A}}(t)$ is its center of mass velocity vector. In this way,

the self-diffusion coefficient is related to the velocity autocorrelation function. On the other hand, the shear viscosity is associated with the time autocorrelation function of the off-diagonal elements of the stress tensor. The thermal conductivity and the electrical conductivity are related to the autocorrelation functions for the energy and electrical current, respectively.

Beside the Green–Kubo and the Einstein formulations, transport properties can be calculated by non-equilibrium MD (NEMD) methods. These involve an externally imposed field that drives the system out of the equilibrium. Similar to experimental approaches, the transport properties can be extracted from the long-time response to this imposed perturbation. E.g., shear flow and energy flux perturbations yield shear viscosity and thermal conductivity, respectively. Numerous NEMD algorithms can be found in the literature, e.g., the Dolls tensor [221], the Sllod algorithm [222], or the boundary-driven algorithm [223]. A detailed review of several NEMD approaches can be found, e.g., in [224].

The NEMD methods are based on the general expression [225]:

$$\gamma = \lim_{F_e \rightarrow 0} \lim_{t \rightarrow \infty} \frac{\langle J(t) \rangle}{F_e}, \quad (34)$$

where $\langle J(t) \rangle$ is the steady state average of the thermodynamic flux $J(t)$ perturbed by the external field F_e . Although a methodology for calculating diffusion coefficients with NEMD is available, such methods are predominantly employed to calculate the shear viscosity and the thermal conductivity [226, 227]. NEMD methods are favored when the signal-to-noise ratio is high for long times. There is an extensive ongoing discussion on whether or not NEMD methods should generally be preferred over equilibrium MD [11, 225, 228, 229].

5.5 Simulation Tools

There are numerous available open source and commercial molecular simulation codes. Examples for MD codes are CHARMM,¹ DL-POLY [230], GROMACS [231], LAMMPS [232], MACSIMUS,² Moldy [233], *ms2* [234], NAMD [235], Tinker [236], and YASP [237]. Some MC simulation codes are BIGMAC,³ BOSS [238], GCMC,⁴ Medea Gibbs,⁵ MCCCSTowhee⁶, and *ms2* [234]. These software packages have been developed for different applications and show large differences in terms of performance, parallelization paradigm, and handling. Most of them use

¹<http://www.charmm.org/>

²<http://www.vscht.cz/fch/software/macsimus/index.html>

³<http://molsim.chem.uva.nl/bigmac/bigmac.html>

⁴<http://kea.princeton.edu/jerring/gibbs>

⁵<http://www.materialsdesign.com/medea/medea-gibbs>

⁶<http://towhee.sourceforge.net/>

their own input and force field files as well as analysis programs to compute the desired properties from the simulation output. Many simulation tools are in constant development and have an increasing number of active users; thus their supported features are constantly changing.

6 Case Study: Ammonia

Ammonia is one of the most important industrial chemicals. Due to its relevance and its simple symmetric molecular structure, much work has been devoted to the development of a force field that is capable of accurately predicting a broad range of its thermodynamic properties. In the following, the capabilities of force fields fitted to QM and vapor–liquid equilibrium data to predict other pure component properties over a wide range of states are addressed.

6.1 Force Fields

Several semi-empirical and empirical force fields have been developed for ammonia [108, 139, 239–247]. In this work, some rigid, non-polarizable models optimized with different parameterization strategies will be addressed. Jorgensen and Ibrahim [239] used experimental geometric information, i.e., bond lengths and bond angles, together with *ab initio* information, to devise a force field based on one LJ 12–6 site and four point charges. They used the STO-3G minimal basis set to calculate the energy of 250 different ammonia dimer configurations. An empirical scaling factor was adopted to account for the polarizability in the liquid phase. Hinchliffe et al. [240] followed a similar parameterization strategy, but employed a Morse potential for repulsion and dispersion. The parameters of the Morse potential and the four point charges were fitted to the dimer energy surface calculated with the 6-31G* basis set for seven different dimer configurations. The geometric parameters were taken from experimental results. Impey and Klein [108] re-parameterized the model by Hinchliffe et al. [240] and replaced the Morse potential with one LJ 12–6 site located at the nitrogen nucleus to describe the dispersive and repulsive interactions. They kept the point charges at the hydrogen nucleus positions, but displaced the nitrogen partial charge towards the hydrogen atoms. The parameters of this five-site model were optimized to the radial distribution function of liquid ammonia.

Kristóf et al. [246] proposed an empirical force field, fitted to experimental molecular geometry and vapor–liquid equilibrium properties. This force field consists of one LJ 12–6 site plus four partial charges. Recently, Zhang and Siepmann [247] proposed a five-site ammonia force field based on the geometry of the Impey and Klein [108] model. This force field also consists of one LJ 12–6 site and four partial charges, three of them located at the hydrogen positions and one located at a

distance of 0.08 Å from the nitrogen nucleus. The LJ parameters, partial charge magnitudes, and the position of the displaced nitrogen charge were optimized to vapor–liquid equilibrium data.

Eckl et al. [97] introduced a semi-empirical force field for ammonia also based on one LJ 12–6 site and four partial charges that are located at the nitrogen and hydrogen positions. The geometry was calculated at the self-consistent field HF level of theory with a 6-31G basis set. The resulting geometry ($r_{\text{NH}} = 1.0136$ Å, $\angle_{\text{HNH}} = 105.99^\circ$) is very close to the experimental data ($r_{\text{NH}} = 1.0124$ Å, $\angle_{\text{HNH}} = 106.67^\circ$) [248]. Eckl et al. [97] adjusted the partial charge magnitudes to the results from a single point QM calculation at the MP2 level of theory with the polarizable basis set 6-311G(d,p) using the COSMO [90] method to account for the liquid polarizability. Only the two LJ parameters were adjusted to experimental data on saturated liquid density, vapor pressure, and enthalpy of vaporization.

6.2 Vapor–Liquid Equilibria of Ammonia

Both the GEMC and the grand equilibrium method have been applied to evaluate vapor–liquid equilibrium data for ammonia. Kristóf et al. [246] calculated the vapor pressure and saturated densities using the force field by Impey and Klein [108] and found systematic deviations from experimental data; cf. Fig. 3. Therefore, they proposed a new ammonia force field that was optimized to vapor–liquid equilibria [246], achieving a better accuracy. Simulated saturated densities and enthalpies based on this force field agree with the experimental data within 1 and 3%, respectively. However, it shows a mean deviation of 13% from experimental

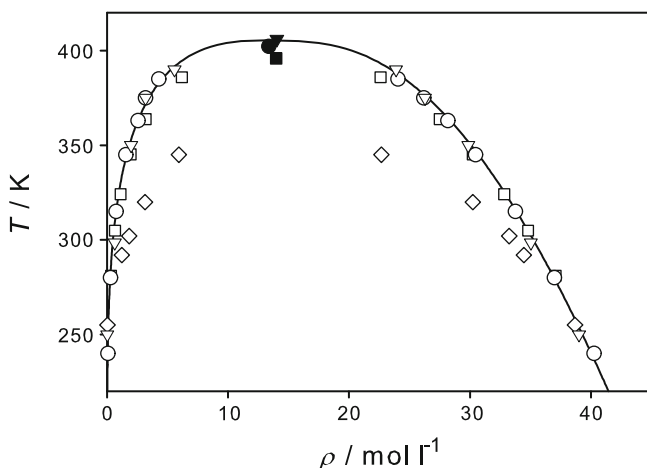


Fig. 3 Saturated densities of ammonia on the basis of different force fields by Impey and Klein (*open diamonds*) [108], Kristóf et al. (*open squares*) [246], Eckl et al. (*open circles*) [97], as well as Zhang and Siepmann (*open inverted triangles*) [247]. The simulation results are compared with a reference equation of state (*solid line*) [249]. The calculated critical points (*full symbols*) are also shown

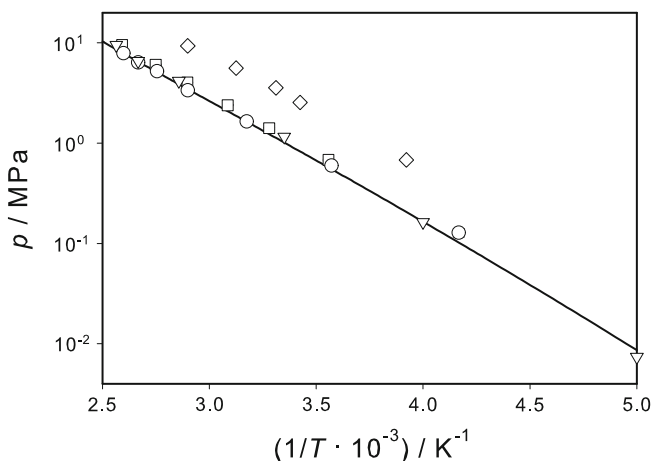


Fig. 4 Saturated vapor pressure of ammonia on the basis of different force fields by Impey and Klein (*open diamonds*) [108], Kristóf et al. (*open squares*) [246], Eckl et al. (*open circles*) [97], as well as Zhang and Siepmann (*open inverted triangles*) [247]. The simulation results are compared with a reference equation of state (*solid line*) [249]

vapor pressure data and the critical temperature is underestimated by 2.4% [97]. A further improvement was achieved by the model from Eckl et al. [97] with mean deviations from the critical temperature, saturated liquid density, vapor pressure, and enthalpy of vaporization of 0.8, 0.7, 1.6, and 2.7%, respectively. The recently introduced force field by Zhang and Siepmann [247] reproduces the saturated liquid densities up to 375 K with a similar accuracy to that of the model of Eckl et al. [97]. This force field predicts the critical density, critical pressure, and normal boiling point with deviations of 0.9, 2, and 0.5%, respectively.

Figures 3 and 4 show the saturated densities and the vapor pressure on the basis of the force fields by Impey and Klein [108], Kristóf et al. [246], Zhang and Siepmann [247], and Eckl et al. [97] for the whole temperature range from triple point to critical point together with a reference equation of state [249] for comparison.

6.3 Properties of the Homogeneous State

As discussed in Sect. 2, force fields should not only be able to represent the thermodynamic properties that were used for their parameterization, but should also be capable of predicting other properties at different thermodynamic conditions. The force field for ammonia by Eckl et al. [97] is an example of such a force field.

Eckl et al. [97] predicted the density and the enthalpy of liquid, gaseous, and supercritical ammonia at 70 different state points, covering a wide range of states for temperatures up to 700 K and pressures up to 700 MPa. They found typical deviations from experimental data below 3 and 5% for the density and the residual enthalpy, respectively. Figure 5 shows the density results on the basis of this force field compared with a reference equation of state [249].

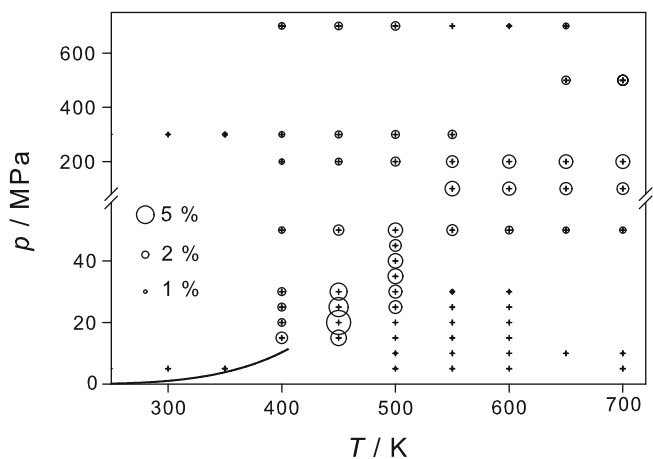


Fig. 5 Relative deviations of the density of ammonia as predicted from the force field by Eckl et al. (*plus signs*) [97] from a reference EOS [249]. The size of the bubbles denotes the relative deviations as indicated in the plot. The *solid line* is the vapor pressure curve

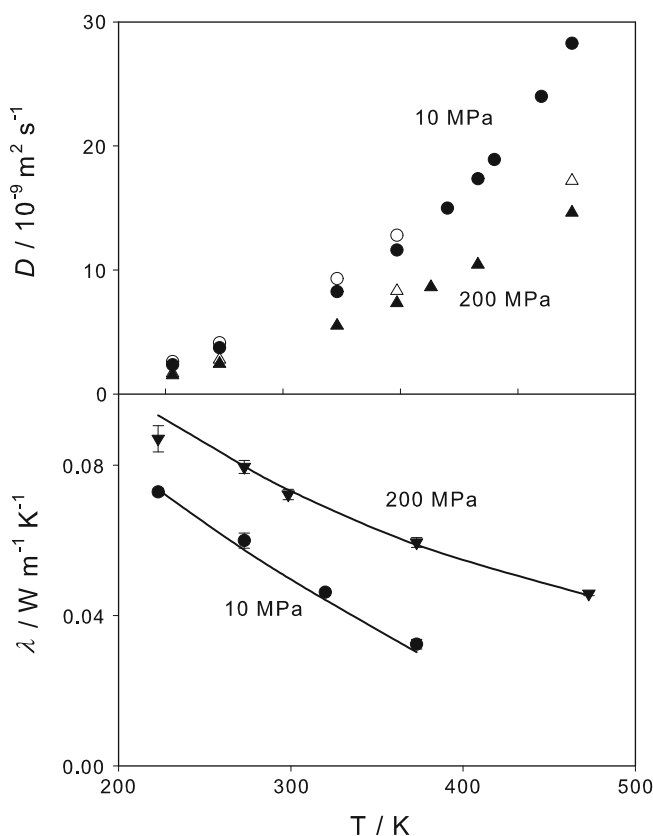


Fig. 6 Temperature dependence of the self-diffusion coefficient (*top*) and the thermal conductivity (*bottom*) of liquid ammonia on the basis of the force field by Eckl et al. [97]. Simulation results at 10 MPa (*filled circles*) and 200 MPa (*filled triangles*) are compared to experimental data (*open symbols*) [250] and to a correlation of experimental data (*solid line*) [251]

This model was extensively tested with respect to its ability to yield transport properties. E.g., the self-diffusion coefficient was predicted in the temperature range from 203 to 473 K for pressures between 10 and 200 MPa with a mean deviation of 15% over the whole range of studied conditions. As an example, Fig. 6 shows the temperature dependence of the self-diffusion coefficient at 10 and 200 MPa in comparison to experimental data [250].

The thermal conductivity and the shear viscosity of ammonia were also predicted with a good accuracy on the basis of the force field by Eckl et al. [97] in the same temperature and pressure range. The predictions of the thermal conductivity and the shear viscosity deviate on average by 3 and 14%, respectively, from the experimental data.

7 Case Study: Binary Mixtures Containing CO₂

CO₂ is an important substance which is present in many processes in the chemical industry. In the following, a case study on the prediction of the Henry's law constant for CO₂ in ethanol and the vapor–liquid equilibrium of the binary mixture CO₂ + C₂H₆ is discussed. The aim is to explore the capabilities of force fields to predict the temperature dependence of gas solubility and to predict azeotropic behavior.

7.1 Force Fields

The Van der Waals interactions of the force fields for CO₂ and C₂H₆ were described by two LJ 12–6 sites and one point quadrupole (16). Both force fields were empirically parameterized to experimental critical temperatures, saturated liquid densities, and vapor pressures by means of a nonlinear optimization algorithm. For both pure substances, the vapor–liquid equilibrium properties from simulation deviate by less than 1% from experimental saturated liquid density data and less than 3% from experimental vapor pressure and enthalpy of vaporization data.

The force field for ethanol [252] consists of three LJ 12–6 sites plus three point charges and was parameterized to ab initio and experimental data. The nucleus positions of all ethanol atoms were computed by QM at the HF level of theory with a 6-31G basis set. This force field is also based on the anisotropic approach of Ungerer et al. [130]. The LJ parameters and the anisotropic offset were fitted to experimental saturated liquid density, vapor pressure, and enthalpy of vaporization. The simulation results from this ethanol force field deviate on average from experimental values of vapor pressure, saturated liquid density, and heat of vaporization by 3.7, 0.3, and 0.9%, respectively.

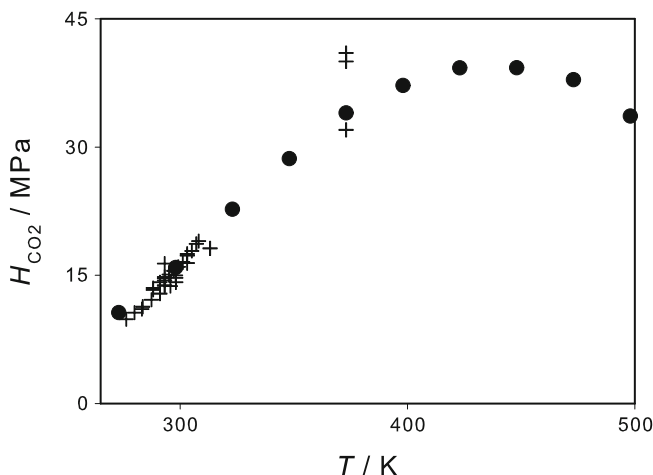


Fig. 7 Henry's law constant of CO₂ in ethanol. The simulation results by Schnabel et al. (*filled circles*) [252] are compared with experimental data (*plus signs*) [253–259]

7.2 Henry's Law Constant of CO₂ in Ethanol

Schnabel et al. [252] calculated the Henry's law constant of CO₂ in ethanol. They evaluated the chemical potential with Widom's test molecule method [207]; cf. (30). In this approach, by simulating the pure solvent, the mole fraction of the solute in the solvent is exactly zero, as required for infinite dilution, because the test molecules are instantly removed after the potential energy calculation.

The results from Schnabel et al. [252] are in excellent agreement with the experimental data; cf. Fig. 7. It has been shown for over 100 other mixtures [39, 252] that the Henry's law constant can reliably and accurately be obtained by molecular simulation using relatively simple force fields when the unlike LJ interaction is adjusted to a single binary data point from experiment.

7.3 Vapor–Liquid Equilibria of the Mixture CO₂ + C₂H₆

Particularly when polar groups are present in liquid mixtures, azeotropes are often formed. For the design of separation processes like distillation, the knowledge of the azeotropic composition at different thermodynamic conditions is of critical importance. In this context, molecular simulation offers a powerful route to predict azeotropic behavior in mixtures. The prediction of the vapor–liquid equilibrium of the mixture CO₂ + C₂H₆ is presented here as an example.

Vrabec et al. [41] predicted the vapor–liquid equilibrium of the mixture CO₂ + C₂H₆ for three different isotherms. The azeotropic behavior of this mixture was predicted using the Lorentz–Berthelot combining rule (12), i.e., relying exclusively on pure substance models without considering any experimental binary data. The quality of the predicted data is clearly superior to the Peng–Robinson EOS with the

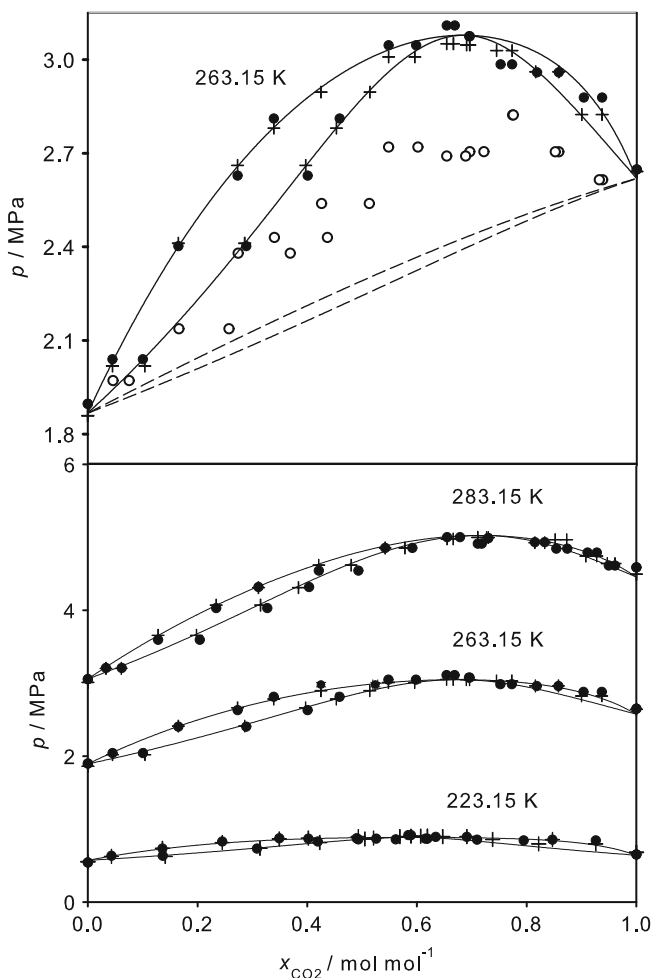


Fig. 8 Vapor-liquid equilibria of the mixture $\text{CO}_2 + \text{C}_2\text{H}_6$. The upper figure shows a magnified view of the simulation results at 263.15 K by Vrabec et al. [41] with $\xi = 1$ (open circles) and $\xi = 0.954$ (filled circles) compared with experimental data (plus signs) [260] and the Peng-Robinson equation of state with $k_{ij} = 0$ (dashed line) and $k_{ij} = 0.132$ (solid line). The figure at the bottom shows the simulation results by Vrabec et al. [41] for 223.15, 263.15, and 283.15 K with $\xi = 0.954$ (filled circles) and the Peng-Robinson EOS with $k_{ij} = 0.132$ (solid line) compared with experimental data (plus signs) [260]

binary interaction coefficient $k_{ij} = 0$, which shows no azeotrope; cf. Fig. 8. As discussed in Sect. 2.1.2, for simulations of binary mixtures, unlike LJ parameters are needed. In many cases, the Lorentz-Berthelot combining rule (12) is too crude to obtain accurate results [34]. Therefore, the modified version of the Lorentz-Berthelot rule (13) was preferred. When the binary parameter ξ is adjusted to one experimental binary data point, the simulation results are in excellent agreement with experimental data; cf. Fig. 8. The Peng-Robinson EOS, being a workhorse in

industrial applications, also shows very good agreement with the experiment when k_{ij} is adjusted.

8 Concluding Remarks

With the ongoing increase of computer performance, molecular modeling and simulation is gaining importance as a tool for predicting the thermodynamic properties for a wide variety of fluids in the chemical industry. One of the major issues of molecular simulation is the development of adequate force fields that are simple enough to be computationally efficient, but complex enough to consider the relevant inter- and intramolecular interactions. There are different approaches to force field development and parameterization. Parameters for molecular force fields can be determined both bottom-up from quantum chemistry and top-down from experimental data.

Transferable force fields have the benefit that they are ready to use and do not need to be fitted for each component individually, although at the expense of prediction accuracy. On the other hand, specific force fields, parameterized for a single molecule, are time-intensive in development and require experimental and/or QM data for optimization. Their main advantage is that they can yield excellent accuracies. The advances of the QM methods in recent years allow for the construction of force fields based on high quality *ab initio* data, i.e., nowadays force fields can be constructed even for new fluids whose properties have been poorly measured or not measured at all. Therefore, molecular modeling and simulation based on classical force fields is a promising alternative route, which in many cases complements the well established methods like classical equations of state or G^E models.

Acknowledgments The presented research was conducted under the auspices of the Boltzmann-Zuse Society of Computational Molecular Engineering (BZS). The simulations were performed on the national supercomputer NEC SX-8 at the High Performance Computing Center Stuttgart (HLRS) under the grant MMHBF and on the HP X6000 supercomputer at the Steinbuch Center for Computing, Karlsruhe under the grant LAMO. Furthermore, the authors are grateful to Ekaterina Elts and Gabor Rutkai for suggestions on improving the manuscript.

References

1. Hendriks E, Kontogeorgis GM, Dohrn R et al (2010) Industrial requirements for thermodynamics and transport properties. *Ind Eng Chem Res* 49:11131–11141
2. Dohrn R, Pfohl O (2002) Thermo-physical properties – industrial directions. *Fluid Phase Equilib* 194–197:15–29
3. Gupta S, Olson JD (2003) Industrial needs in physical properties. *Ind Eng Chem Res* 42:6359–6374

4. Rhodes CL (1996) The process simulation revolution: thermophysical property needs and concerns. *J Chem Eng Data* 41:947–950
5. Sandler SI (1994) Thermophysical properties: what have we learned recently, and what do we still need to know? *Int J Thermophys* 15:1013–1035
6. Zeck S, Wolf D (1993) Requirements of thermodynamic data in the chemical industry. *Fluid Phase Equilib* 82:27–38
7. Poling BE, Prausnitz JM, O'Connell JP (2000) The properties of gases and liquids, 5th edn. McGraw Hill, New York
8. Gubbins KE, Quirke N (1996) Introduction to molecular simulation and industrial applications: methods, examples and prospects. In: Gubbins KE, Quirke N (eds) *Molecular simulations and industrial applications*. Gordon and Breach Science Publishers, Amsterdam
9. Maginn EJ, Elliot JR (2010) Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Ind Eng Chem Res* 49:3059–3078
10. Friesner RA (2005) *Ab initio* quantum chemistry: methodology and applications. *Proc Natl Acad Sci USA* 102:6648–6653
11. Allen MP, Tildesley DJ (1997) *Computer simulation of liquids*. Clarendon Press, Oxford
12. Burkert U, Allinger NL (1982) *Molecular mechanics*, ACS monograph 177. American Chemical Society, Washington, DC
13. Stone AJ (1996) *The theory of intermolecular forces*. Clarendon Press, Oxford
14. Mie G (1903) Zur kinetischen Theorie der einatomigen Körper. *Ann Phys* 11:657–697
15. Horsch M, Vrabec J, Hasse H (2008) Modification of the classical nucleation theory based on molecular simulation data for surface tension, critical nucleus size, and nucleation rate. *Phys Rev E* 78:011603
16. Horsch M, Vrabec J (2009) Grand canonical steady-state simulation of nucleation. *J Chem Phys* 131:184104
17. Kihara T (1951) The second virial coefficient of non-spherical molecules. *J Phys Soc Jpn* 6:289–296
18. Errington JR, Panagiotopoulos AZ (1999) A new intermolecular potential model for the n-alkane homologous series. *J Phys Chem B* 103:6314–6322
19. Leach AR (2001) *Molecular modelling principles and applications*, 2nd edn. Pearson Education, Edinburgh
20. Cabaleiro-Lago EM, Rios MA (1997) A potential function for intermolecular interaction in the acetonitrile dimer constructed from *ab initio* data. *J Phys Chem A* 101:8327–8334
21. Eggenberger R, Gerber S, Huber H et al (1994) A new *ab initio* potential for the neon dimer and its application in molecular dynamics simulations of the condensed phase. *Mol Phys* 82:689–699
22. Grochola G, Russo S, Snook I (1998) An *ab initio* pair potential for Ne₂ and the equilibrium properties of neon. *Mol Phys* 95:471–475
23. Hellmann R, Bich E, Vogel E (2007) *Ab initio* potential energy curve for the helium atom pair and thermophysical properties of dilute helium gas. I. Helium-helium interatomic potential. *Mol Phys* 105:3013–3023
24. Hloucha M, Sum AK, Sandler SI (2000) Computer simulation of acetonitrile and methanol with *ab initio*-based pair potentials. *J Chem Phys* 113:5401–5406
25. Tang KT, Toennies JP (1980) An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J Chem Phys* 80:3726–3741
26. Al-Matar AK, Rockstraw D (2004) A generating equation for mixing rules and two new mixing rules for interatomic potential energy parameters. *J Comput Chem* 25:660–668
27. Fender BEF, Halsey GD (1962) Second virial coefficients of argon, krypton, and argon-krypton mixtures at low temperatures. *J Chem Phys* 36:1881–1888
28. Halgren TA (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J Am Chem Soc* 114:7827–7843

29. Kong CL (1973) Combining rules for intermolecular potential parameters. II. Rules for the Lennard-Jones (12–6) potential and the Morse potential. *J Chem Phys* 59:2464–2467
30. Sikora PT (1970) Combining rules for spherically symmetric intermolecular potentials. *J Phys B* 3:1475–1482
31. Waldman M, Hagler AT (1993) New combining rules for rare-gas van der Waals parameters. *J Comput Chem* 14:1077–1084
32. Peña MD, Pando C, Renuncio JAR (1982) Combination rules for intermolecular potential parameters. I. Rules based on approximations for the long-range dispersion energy. *J Chem Phys* 76:325–332
33. Peña MD, Pando C, Renuncio JAR (1982) Combination rules for intermolecular potential parameters. II. Rules based on approximations for the long-range dispersion energy and an atomic distortion model for the repulsive interactions. *J Chem Phys* 76:333–339
34. Schnabel T, Vrabec J, Hasse H (2007) Unlike Lennard-Jones parameters for vapor-liquid equilibria. *J Mol Liq* 135:170–178
35. Lorentz HA (1881) Über die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann Phys* 12:127–136
36. Berthelot D (1889) Sur le Mélange des Gaz. *C R Acad Sci* 126:1703–1706
37. Delhomme J, Millié P (2001) Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation. *Mol Phys* 99:619–625
38. Ungerer P, Wender A, Demoulin G et al (2004) Application of Gibbs ensemble and NPT Monte Carlo simulation to the development of improved processes for H₂S-rich gases. *Mol Simul* 30:631–648
39. Huang YL, Miroshnichenko S, Hasse H et al (2009) Henry's law constant from molecular simulation: a systematic study of 95 systems. *Int J Thermophys* 30:1791–1810
40. Huang YL, Vrabec J, Hasse H (2009) Prediction of ternary vapor-liquid equilibria for 33 systems by molecular simulation. *Fluid Phase Equilib* 287:62–69
41. Vrabec J, Huang YL, Hasse H (2009) Molecular models for 267 binary mixtures validated by vapor-liquid equilibria: a systematic approach. *Fluid Phase Equilib* 279:120–135
42. Stone AJ (2008) Intermolecular potentials. *Science* 321:787–789
43. Luthy CS, Singer K, Klein ML et al (1983) Electrostatic interactions in molecular crystals. Lattice dynamics of solid nitrogen and carbon dioxide. *Mol Phys* 50:531–541
44. Hirschfelder JO, Curtiss CF, Bird RB (1954) *Molecular theory of gases and liquids*. Wiley, New York
45. Halgren TA, Damm W (2001) Polarizable force fields. *Curr Opin Struct Biol* 11:236–242
46. Rick SW, Stuart SJ (2002) Potentials and algorithms for incorporating polarizability in computer simulations. In: Lipowitz DB, Boyd DB (eds) *Review in computational chemistry*. Wiley-VCH, New York
47. Yu H, van Gunsteren WF (2005) Accounting for polarization in molecular simulation. *Comput Phys Commun* 172:69–85
48. Dang LX, Rice JE, Caldwell J et al (1991) Ion solvation in polarizable water: molecular dynamics simulations. *J Am Chem Soc* 113:2481–2486
49. Rick SW, Stuart SJ, Berne BJ (1994) Dynamical fluctuating charge force fields: application to liquid water. *J Chem Phys* 101:6141–6157
50. Rigby M, Smith EB, Wakeham WA et al (1986) *The forces between molecules*. Clarendon Press, Oxford
51. Simons G, Parr RG, Finlan JM (1973) New alternative to the Dunham potential for diatomic molecules. *J Chem Phys* 59:3229–3234
52. Morse PM (1929) Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys Rev* 34:57–64
53. Hünenberger PH, van Gunsteren WF (1997) Empirical classical interaction functions for molecular simulation. In: van Gunsteren WF, Weiner PK, Wilkinson AJ (eds) *Computer simulation of biomolecular systems: theoretical and experimental applications*. Kluwer Academic, Dordrecht

54. van der Ploeg P, Berendsen HJC (1982) Molecular dynamics simulation of a bilayer membrane. *J Chem Phys* 76:3271–3276
55. Ryckaert JP, Bellemans A (1975) Molecular dynamics of liquid n-butane near its boiling point. *Chem Phys Lett* 30:123–125
56. Economou IG (2004) Molecular simulation of phase equilibria for industrial applications. In: Kontogeorgis GM, Gani R (eds) *Computer aided property estimation for process and product design*. Elsevier, Amsterdam
57. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 118:11225–11236
58. Foloppe N, MacKerell AD (2000) All-atom empirical force field for nucleic acids: parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* 21:86–104
59. Bourasseau E, Haboudou M, Boutin A et al (2003) New optimization method for intermolecular potentials: optimization of a new anisotropic united atoms potential for olefins. prediction of equilibrium properties. *J Chem Phys* 118:3020–3034
60. Faller R, Schmitz H, Biermann O et al (1999) Automatic parameterization of force fields for liquids by simplex optimization. *J Comput Chem* 20:1009–1017
61. Hülsmann M, Köddermann T, Vrabec J et al (2010) GROW: a gradient-based optimization workflow for the automated development of molecular models. *Comput Phys Commun* 181:499–513
62. Njo SL, van Gunsteren WF, Müller-Plathe F (1995) Determination of force field parameters for molecular simulation by molecular simulation: an application of the weak-coupling method. *J Chem Phys* 102:6199–6207
63. Waldher B, Kuta J, Chen S et al (2010) ForceFit: a code to fit classical force fields to quantum mechanical potential energy surfaces. *J Comput Chem* 31:2307–2316
64. Wang J, Kollman PA (2001) Automatic parameterization of force field by systematic search and genetic algorithms. *J Comput Chem* 22:1219–1228
65. Wang J, Wolf RM, Caldwell JW et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25:1157–1174
66. Cornell WD, Cieplak P, Bayly CI et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117: 5179–5197
67. Kaminski G, Friesner RA, Tirado-Rives J et al (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* 105:6474–6487
68. Pérez-Pellitero J, Bourasseau E, Demachy I et al (2008) Anisotropic united-atoms (AUA) potential for alcohols. *J Phys Chem B* 112:9853–9863
69. Sandler SI, Castier M (2007) Computational quantum mechanics: an underutilized tool in thermodynamics. *Pure Appl Chem* 79:1345–1359
70. Sun H (1998) COMPASS: an ab initio force-field optimized for condensed-phase applications. overview with details on alkane and benzene compounds. *J Phys Chem B* 102:7338–7364
71. Löwdin P (1970) On the nonorthogonality problem. *Adv Quantum Chem* 5:185–199
72. Mulliken RS (1962) Criteria for the construction of good self-consistent-field molecular orbital wave functions, and the significance of LCAOMO population analysis. *J Chem Phys* 36:3428–3440
73. Li J, Zhu T, Cramer CT et al (1998) New class IV charge model for extracting accurate partial charges from wave functions. *J Phys Chem A* 102:1820–1831
74. Reed AE, Weinstock RB, Weinhold FA (1985) Natural population analysis. *J Chem Phys* 83:735–747
75. Bader RFW (1985) *Atoms in molecules. A quantum theory*. Clarendon Press, Oxford
76. Wiberg KB, Rablen PR (1993) Comparison of atomic charges derived via different procedures. *J Comput Chem* 14:1504–1518

77. Mobley DL, Dumont E, Chodera JD (2007) Comparison of charge models for fixed-charge force fields: small molecule hydration free energies in explicit solvent. *J Phys Chem B* 111:2242–2254
78. Chirlian LE, Francel MM (1987) Atomic charges derived from electrostatic potentials: a detailed study. *J Comput Chem* 8:894–905
79. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11:361–373
80. Bayly CI, Cieplak P, Cornell W et al (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97:10269–10280
81. Stouch TR, Williams DE (1992) Conformational dependence of electrostatic potential derived charges of a lipid headgroup: glycerylphosphorylcholine. *J Comput Chem* 13:622–632
82. Sigfridsson E, Ryde U (1998) Comparison of methods for deriving atomic charges from the electrostatic potential and moments. *J Comput Chem* 19:377–395
83. Anisimov V, Vorobyov IV, Roux B et al (2007) Polarizable empirical force field for the primary and secondary alcohol series based on the classical Drude model. *J Chem Theory Comput* 3:1927–1946
84. Stone AJ (2005) Distributed multipole analysis: stability for large basis sets. *J Chem Theory Comput* 1:1128–1132
85. Joubert L, Popelier PLA (2002) Improved convergence of the atoms in molecules multipole expansion of electrostatic interaction. *Mol Phys* 100:3357–3365
86. Eckl B, Vrabec J, Hasse H (2008) Set of molecular models based on quantum mechanical ab initio calculations and thermodynamic data. *J Phys Chem B* 112:12710–12721
87. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. *Chem Rev* 105:2999–3093
88. Wong MA, Frisch MJ, Wiberg KB (1991) Solvent effects 1. The mediation of electrostatic effects by solvents. *J Am Chem Soc* 113:4776–4782
89. Wong MA, Frisch MJ, Wiberg KB (1992) Solvent effects 2. Medium effect on the structure, energy, charge density, and vibrational frequencies of sulfamic acid. *J Am Chem Soc* 114:523–529
90. Klamt A (1995) Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J Phys Chem* 99:2224–2235
91. MacKerell AD, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
92. Kony D, Damm W, Stoll S et al (2002) An improved OPLS-AA force field for carbohydrates. *J Comput Chem* 23:1416–1429
93. Koch W, Holthausen MC (2001) A chemist's guide to density functional theory, 2nd edn. Wiley, New York
94. Lee C, Yang W, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 37:785–789
95. Hatcher A, Guvench O, MacKerell AD (2009) CHARMM additive all-atom force field for acyclic polyalcohols, acyclic carbohydrates and inositol. *J Chem Theory Comput* 5:1315–1327
96. Jorgensen WL, Madura JD, Swenson CJ (1984) Optimized intermolecular potential functions for liquid hydrocarbons. *J Am Chem Soc* 106:6638–6646
97. Eckl B, Vrabec J, Hasse H (2008) An optimized molecular model for ammonia. *Mol Phys* 106:1039–1046
98. Schnabel T, Srivastava A, Vrabec J et al (2008) Hydrogen bonding of methanol in supercritical CO₂: comparison between 1 H-NMR spectroscopic data and molecular simulation results. *J Phys Chem B* 111:9871–9878

99. Schnabel T, Vrabec J, Hasse H (2008) Molecular simulation study of hydrogen bonding mixtures and new molecular models for mono- and dimethylamine. *Fluid Phase Equilib* 263:144–159
100. Nath SK, Escobedo FA, de Pablo JJ (1998) On the simulation of vapor-liquid equilibria for alkanes. *J Chem Phys* 108:9905–9911
101. Poncela A, Rubio AM, Freire JJ (1997) Determination of the potential parameters of a site model from calculations of second virial coefficients of linear and branched alkanes. *Mol Phys* 91:189–201
102. Vrabec J, Stoll J, Hasse H (2001) A set of molecular models for symmetric quadrupolar fluids. *J Phys Chem B* 105:12126–12133
103. Kaminski G, Duffy EM, Matsui T et al (1994) Free energies of hydration and pure liquid properties of hydrocarbons from the OPLS all-atom model. *J Phys Chem* 98:13077–13082
104. Fiorini M, Burger K, Mark A et al (2000) A new 2,2,2-trifluoroethanol model for molecular dynamics simulations. *J Phys Chem B* 104:12347–12354
105. Micaelo NM, Baptista AM, Soares CM (2006) Parametrization of 1-butyl-3-methylimidazolium hexafluorophosphate/nitrate ionic liquid for the GROMOS force field. *J Phys Chem B* 110:14444–14451
106. Gordon P (2006) Development of intermolecular potentials for predicting transport properties of hydrocarbons. *J Chem Phys* 125:014504
107. Nieto-Draghi C, Ungerer P, Rousseau B (2006) Optimization of the anisotropic united atoms intermolecular potential for n-alkanes: improvement of transport properties. *J Chem Phys* 125:044517
108. Impey RW, Klein ML (1984) A simple intermolecular potential for liquid ammonia. *Chem Phys Lett* 104:579–582
109. Stoll J, Vrabec J, Hasse H (2003) A set of molecular models for carbon monoxide and halogenated hydrocarbons. *J Chem Phys* 119:11396–11407
110. Jorgensen WL, Swenson CJ (1985) Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. *J Am Chem Soc* 107:569–578
111. Jorgensen WL (1986) Optimized intermolecular potential functions for liquid alcohols. *J Phys Chem* 90:1276–1284
112. Jorgensen WL, Tirado-Rives J (1988) The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 110:1657–1666
113. Chen B, Siepmann JI (1999) Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of n-alkanes. *J Phys Chem B* 103:5370–5379
114. Chen B, Potoff JJ, Siepmann JI (2001) Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols. *J Phys Chem B* 105:3093–3104
115. Lubna N, Kamath G, Potoff JJ et al (2005) Transferable potentials for phase equilibria. 8. United-atom description for thiols, sulfides, disulfides, and thiophene. *J Phys Chem B* 109:24100–24107
116. Maerzke KA, Schultz NE, Ross RB et al (2009) TraPPE-UA force field for acrylates and Monte Carlo simulations for their mixtures with alkanes and alcohols. *J Phys Chem B* 113:6415–6425
117. Martin MG, Siepmann JI (1998) Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J Phys Chem B* 102:2569–2577
118. Martin MG, Siepmann JI (1999) Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. 2. United-atom description of branched alkanes. *J Phys Chem B* 103:4508–4517
119. Rai N, Siepmann JI (2007) Transferable potentials for phase equilibria. 9. Explicit-hydrogen description of benzene and 5-membered and 6-membered heterocyclic aromatic compounds. *J Phys Chem B* 111:10790–10799
120. Stubbs JM, Potoff JJ, Siepmann JI (2004) Transferable potentials for phase equilibria. 6. United-atom description for ethers, glycols, ketones and aldehydes. *J Phys Chem B* 108:17596–17605

121. Wick CD, Martin MG, Siepmann JI (2000) Transferable potentials for phase equilibria. 4. United-atom description of linear and branched alkenes and alkylbenzenes. *J Phys Chem B* 104:8008–8016
122. Wick CD, Stubbs JM, Rai N et al (2005) Transferable potentials for phase equilibria. 7. United-atom description for nitrogen, amines, amides, nitriles, pyridine and pyrimidine. *J Phys Chem B* 109:18974–18982
123. Boutard Y, Ungerer P, Teuler JM et al (2005) Extension of the anisotropic united atoms intermolecular potential to amines, amides and alkanols. Application to the problems of the 2004 Fluid Simulation Challenge. *Fluid Phase Equilib* 236:25–41
124. Contreras-Camacho RO, Ungerer P, Boutin A et al (2004) Optimized intermolecular potential for aromatic hydrocarbons based on anisotropic united atoms. 1. Benzene. *J Phys Chem B* 108:14109–14114
125. Creton B, de Bruin T, Lachet V et al (2010) Extension of a charged anisotropic united atoms model to polycyclic aromatic compounds. *J Phys Chem B* 114:6522–6530
126. Delhommelle J, Tschirwitz C, Ungerer P et al (2000) Derivation of an optimized potential model for phase equilibria (OPPE) for sulfides and thiols. *J Phys Chem B* 104:4745–4753
127. Hadj-Kali MK, Gerbaud V, Joulia X et al (2008) Optimized intermolecular potential for nitriles based on anisotropic united atoms model. *J Mol Model* 14:571–580
128. Kranias S, Pattou D, Lévy B (2003) An optimized potential for phase equilibria calculation for ketone and aldehyde molecular fluids. *Phys Chem Chem Phys* 5:4175–4179
129. Nieto-Draghi C, Bonnaud P, Ungerer P (2007) Anisotropic united atom model including the electrostatic interactions of methylbenzenes. I. Thermodynamic and structural properties. *J Phys Chem C* 111:15686–15699
130. Ungerer P, le Beauvais C, Delhommelle J et al (2000) Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *J Chem Phys* 112:5499–5510
131. Nath SK, de Pablo JJ (2000) Simulation of vapor-liquid equilibria for branched alkanes. *Mol Phys* 98:231–238
132. Nath SK, Banaszak BJ, de Pablo JJ (2001) A new united atom force field for α -olefins. *J Chem Phys* 114:3612–3616
133. Nath SK, Khare R (2001) New force field parameters for branched hydrocarbons. *J Chem Phys* 115:10837–10844
134. Errington JR, Panagiotopoulos AZ (1999) New intermolecular potential models for benzene and cyclohexane. *J Chem Phys* 111:9731–9738
135. Damm W, Frontera A, Tirado-Rives J et al (1997) OPLS all-atom force field for carbohydrates. *J Comput Chem* 18:1955–1970
136. Jorgensen WL, McDonald NA (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes. *J Mol Struct THEOCHEM* 424:145–155
137. Kahn K, Bruice T (2001) Parameterization of OPLS-AA force field for the conformational analysis of macrocyclic polyketides. *J Comput Chem* 23:977–996
138. McDonald NA, Jorgensen WL (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. *J Phys Chem B* 102:8049–8059
139. Rizzo RC, Jorgensen WL (1999) OPLS all-atom model for amines: resolution of the amine hydration problem. *J Am Chem Soc* 121:4827–4836
140. MacKerell AD, Wiorkiewicz-Kuczera J, Karplus M (1995) An all-atom empirical energy function for the simulation of nucleic acids. *J Am Chem Soc* 117:11946–11975
141. Smit B, Karaborni S, Siepmann JI (1994) Computer simulations of vapor-liquid phase equilibria of n-alkanes. *J Chem Phys* 102:2126–2140
142. Toxvaerd S (1990) Molecular dynamics calculation of the equation of state of alkanes. *J Chem Phys* 93:4290–4296
143. Toxvaerd S (1997) Equation of state of alkanes II. *J Chem Phys* 107:5197–5204
144. Ferrando N, Lachet V, Teuler JM et al (2009) Transferable force field for alcohols and polyalcohols. *J Phys Chem B* 113:5985–5995

145. Lévy B, Enescu M (1998) Theoretical study of methylene blue: a new method to determine partial atomic charges; investigation of the interaction with guanine. *J Mol Struct Theochem* 432:235–245
146. Glennon TM, Merz KM (1997) A carbohydrate force field for AMBER and its application to the study of saccharides to surface adsorption. *J Mol Struct Theochem* 395–396:157–171
147. Yang L, Tan C, Hsieh MJ et al (2006) New-generation amber united-atom force field. *J Phys Chem B* 110:13166–13176
148. van Gunsteren WF, Billeter SR, Eising AA et al (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag, Zürich
149. Oostenbrink C, Villa A, Mark AE (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676
150. Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem* 22:1205–1218
151. Engelsen SB, Fabricius J, Rasmussen K (1994) The consistent force field. 1. Methods and strategies for optimization of empirical potential energy functions. *Acta Chem Scand Ser A* 48:548–552
152. Engelsen SB, Fabricius J, Rasmussen K (1994) The consistent force field. 2. An optimized set of potential energy functions for the alkanes. *Acta Chem Scand Ser A* 48:553–565
153. Jónsdóttir SO, Rasmussen K (2000) The consistent force field. Part 6: an optimized set of potential energy functions for primary amines. *New J Chem* 24:243–247
154. Shirts MR, Pitner JW, Swope WC et al (2003) Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. *J Chem Phys* 119:5740–5761
155. Guvench O, MacKerell AD (2008) Comparison of protein force fields for molecular dynamics simulations. In: Kukol A (ed) *Molecular modeling of proteins*. Humana Press, New York
156. Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* 25:1584–1604
157. Ponder JW, Case DA (2003) Force fields for protein simulations. *Protein Simul* 66:27–85
158. Price DJ, Brooks CL (2002) Modern protein force fields behave comparably in molecular dynamics simulations. *J Comput Chem* 23:1045–1057
159. Yoda T, Sugita Y, Okamoto Y (2004) Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem Phys Lett* 386:460–467
160. Hobza P, Hubálek F, Kabeláč M et al (1996) Ability of empirical potentials (AMBER, CHARMM, CVFF, OPLS, Poltev) and semi-empirical quantum chemical methods (AM1, MNDO/M, PM3) to describe H-bonding in DNA base pairs; comparison with ab initio results. *Chem Phys Lett* 257:31–35
161. Yeh IC, Hummer G (2002) Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J Am Chem Soc* 124:6563–6568
162. Stortz CA, Johnson GP, French AD et al (2009) Comparison of different force fields for the study of disaccharides. *Carbohydr Res* 344:2217–2228
163. Patra M, Karttunen M (2004) Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: diffusion, free energy of hydration and structural properties. *J Comput Chem* 25:678–689
164. Patel S, Brooks CL (2004) CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem* 25:1–15
165. Gao J, Habibollahzadeh D, Shao L (1995) A polarizable intermolecular potential function for simulation of liquid alcohols. *J Phys Chem* 99:16460–16467
166. Gao J, Pavelites J, Habibollahzadeh D (1996) Simulation of liquid amides using a polarizable intermolecular potential function. *J Phys Chem* 100:2689–2697
167. Xie W, Pu J, MacKerell AD et al (2007) Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *J Chem Theory Comput* 3:1878–1889

168. Ponder JW, Wu C, Ren P (2010) Current status of the AMOEBA polarizable force field. *J Phys Chem B* 114:2549–2564
169. Martin MG (2006) Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor-liquid coexistence curves and liquid densities. *Fluid Phase Equilib* 248:50–55
170. Dellis D, Samios J (2010) Molecular force field investigation for Sulfur Hexafluoride: a computer simulation study. *Fluid Phase Equilib* 291:81–89
171. Fernandez GA, Vrabec J, Hasse H (2005) Shear viscosity and thermal conductivity of quadrupolar real fluids from molecular simulation. *Mol Sim* 31:787–793
172. Fernandez GA, Vrabec J, Hasse H (2006) Shear viscosity and thermal conductivity of dipolar real fluids from molecular simulation. *Cryogenics* 46:711–717
173. Fermeglia M, Ferrone M, Priol S (2003) Development of an all-atoms force field from ab initio calculations for alternative refrigerants. *Fluid Phase Equilib* 210:105–116
174. Huang YL, Heilig M, Hasse H et al (2010) VLE of hydrogen chloride, phosgene, benzene, chlorobenzene, ortho-dichlorobenzene and toluene by molecular simulation. *AIChE J* 57:1043–1060
175. Hunt PA (2006) The simulation of imidazolium-based ionic liquids. *Mol Simul* 32:1–10
176. Liu X, Zhang S, Zhou G et al (2006) New force field for molecular simulation of guanidinium-based ionic liquids. *J Phys Chem B* 110:12062–12071
177. Hellmann R, Bich E, Vogel E (2008) *Ab initio* intermolecular potential energy surface and second pressure virial coefficients of methane. *J Chem Phys* 128:214303
178. Hellmann R, Bich E, Vogel E (2008) *Ab initio* potential energy curve for the neon atom pair and thermophysical properties of the dilute neon gas. I. Neon-neon interatomic potential and rovibrational spectra. *Mol Phys* 106:133–140
179. Domański K, Kitao O, Nakanishi K (1994) A new potential model for carbon dioxide from ab initio calculations. *Mol Simul* 12:343–353
180. Frenkel D, Smit B (2002) *Understanding molecular simulation: from algorithms to applications*, 2nd edn. Elsevier, San Diego
181. Panagiotopoulos AZ (1996) Current advances in Monte Carlo methods. *Fluid Phase Equilib* 116:257–266
182. Theodorou DN (2010) Progress and outlook in Monte Carlo simulations. *Ind Eng Chem Res* 49:3047–3058
183. Allen MP, Tildesley DJ (eds) (1993) *Computer simulation in chemical physics*. Kluwer Academic, Dordrecht
184. Amar JG (2006) The Monte Carlo method in science and engineering. *Comput Sci Eng* 8:9–19
185. Baus M, Rull LF, Ryckaert JP (eds) (1995) *Observation and prediction of phase transitions in complex fluids*. Kluwer Academic, Dordrecht
186. Valleau JP (1991) Density-scaling: a new Monte Carlo technique in statistical mechanics. *J Comput Phys* 96:193–216
187. Ferrenberg AM, Swendsen RH (1988) New Monte Carlo technique for studying phase transitions. *Phys Rev Lett* 61:2635–2638
188. Ferrenberg AM, Swendsen RH (1989) Optimized Monte Carlo data analysis. *Phys Rev Lett* 63:1195–1198
189. Kofke DA (1993) Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation. *Mol Phys* 78:1331–1336
190. Vrabec J, Hasse H (1995) Vapour liquid equilibria of mixtures from the NpT plus test particle method. *Mol Phys* 85:781–792
191. Boda D, Kristóf T, Liszi J et al (2001) A new simulation method for the determination of phase equilibria in mixtures in the grand canonical ensemble. *Mol Phys* 99:2011–2022
192. Vrabec J, Hasse H (2002) Grand equilibrium: vapour-liquid equilibria by a new molecular simulation method. *Mol Phys* 100:3375–3383

193. Panagiotopoulos AZ (1987) Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol Phys* 61:813–826
194. Panagiotopoulos AZ (2000) Monte Carlo methods for phase equilibria of fluids. *J Phys Condens Matter* 12:25–52
195. de Pablo JJ, Yan Q, Escobedo FA (1999) Simulation of phase transitions in fluids. *Annu Rev Phys Chem* 50:377–411
196. Lamm MJ, Hall CK (2001) Molecular simulation of complete phase diagrams for binary mixtures. *AIChE J* 47:1664–1675
197. Johnson JK, Panagiotopoulos AZ, Gubbins KE (1994) Reactive canonical Monte Carlo: a new simulation technique for reacting or associating fluids. *Mol Phys* 81:717–733
198. Smith WR, Triska B (1994) The reaction ensemble method for the computer simulation of chemical and phase equilibria I. Theory and basic examples. *J Chem Phys* 100:3019–3027
199. Kiyohara K, Spyriouni T, Gubbins KE et al (1996) Thermodynamic scaling Gibbs ensemble Monte Carlo: a new method for determination of phase coexistence properties of fluids. *Mol Phys* 89:965–974
200. Briano JG, Glandt ED (1984) Statistical thermodynamics of polydisperse fluids. *J Chem Phys* 80:3336–3343
201. Kofke DA, Glandt ED (1988) Monte Carlo simulation of multicomponent equilibria in a semigrand canonical ensemble. *Mol Phys* 64:1105–1131
202. Potoff JJ, Panagiotopoulos AZ (2000) Surface tension of the three-dimensional Lennard-Jones fluid from histogram-reweighting Monte Carlo simulations. *J Chem Phys* 112:6411–6416
203. Singh JK, Errington JR (2006) Calculation of phase coexistence properties and surface tensions of n-alkanes with grand-canonical transition-matrix Monte Carlo simulation and finite-size scaling. *J Phys Chem B* 110:1369–1376
204. Ungerer P, Boutin A, Fuchs AH (1999) Direct calculation of bubble points by Monte Carlo simulations. *Mol Phys* 97:523–539
205. Escobedo FA (1998) Novel pseudoensembles for simulation of multicomponent phase equilibria. *J Chem Phys* 108:8761–8772
206. van der Vegt NFA, Briels WJ, Wessling M et al (1999) The sorption induced glass transition in amorphous glassy polymers. *J Chem Phys* 110:11061–11069
207. Widom B (1963) Some topics in the theory of fluids. *J Chem Phys* 39:2808–2812
208. Lyubartsev AP, Martsinovski AA, Shevkunov SV et al (1992) New approach to Monte Carlo calculation of the free energy: method of expanded ensembles. *J Chem Phys* 96:1776–1785
209. Nezbeda I, Kolafa J (1991) A new version of the insertion particle method for determining the chemical potential by Monte Carlo simulation. *Mol Simul* 5:391–403
210. Vrabec J, Kettler M, Hasse H (2002) Chemical potential of quadrupolar two-centre Lennard-Jones fluids by gradual insertion. *Chem Phys Lett* 356:431–436
211. Maginn EJ, Bell AT, Theodorou DN (1995) Sorption thermodynamics, siting, and conformation of long n-alkanes in silicalite as predicted by configurational-bias Monte Carlo integration. *J Phys Chem* 99:2057–2079
212. Theodorou DN (2006) A reversible minimum-to-minimum mapping method for the calculation of free-energy differences. *J Chem Phys* 124:034109
213. Kofke DA, Cummings PT (1997) Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation. *Mol Phys* 92:973–996
214. Murad S, Gupta S (2000) A simple molecular dynamics simulation for calculating Henry's constant and solubility of gases in liquids. *Chem Phys Lett* 319:60–64
215. Sadus RJ (1997) Molecular simulation of Henry's constant at vapor-liquid and liquid-liquid phase boundaries. *J Phys Chem B* 101:3834–3838
216. Shing KS, Gubbins KE, Lucas K (1988) Henry constants in nonideal fluid mixtures. Computer simulation and theory. *Mol Phys* 65:123–1252
217. Green MS (1954) Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids. *J Chem Phys* 22:398–414

218. Kubo R (1957) Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J Phys Soc Jpn* 12:570–586
219. Ciccotti G, Frenkel D, McDonald IR (1987) Simulation of liquids and solids. North-Holland, Amsterdam
220. Rowley RL, Painter MM (1997) Diffusion and viscosity equations of state for a Lennard-Jones fluid obtained from molecular dynamics simulations. *Int J Thermophys* 18:1109–1121
221. Hoover WG, Evans DJ, Hickman RB et al (1980) Lennard-Jones triple-point bulk and shear viscosities. Green-Kubo theory, Hamiltonian mechanics, and nonequilibrium molecular dynamics. *Phys Rev A* 22:1690–1697
222. Ladd AJC (1984) Equations of motion for non-equilibrium molecular dynamics simulations of viscous flow in molecular fluids. *Mol Phys* 53:459–463
223. Lees AW, Edwards SF (1972) The computer study of transport processes under extreme conditions. *J Phys C* 5:1921–1929
224. Evans DJ, Morris GP (1990) Statistical mechanics of nonequilibrium liquids. Academic Press, London
225. Cummings PT, Evans DJ (1992) Nonequilibrium molecular dynamics properties and non-Newtonian fluid approaches to transport rheology. *Ind Eng Chem Res* 31:1237–1252
226. Müller-Plathe F (1997) A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity. *J Chem Phys* 106:6082–6085
227. Müller-Plathe F, Bordat P (2004) Reverse non-equilibrium molecular dynamics. *Lect Notes Phys* 640:310–326
228. Chen T, Smit B, Bell AT (2009) Are pressure fluctuation-based equilibrium methods really worse than nonequilibrium methods for calculating viscosities? *J Chem Phys* 131:246101
229. Sadus RJ (2002) Molecular simulation of fluids: theory, algorithms and object-orientation. Elsevier, Amsterdam
230. Smith W, Yong CW, Rodger PM (2002) DL-POLY: application to molecular simulation. *Mol Simul* 28:385–471
231. Hess B, Kutzner C, van der Spoel D et al (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447
232. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *J Comp Phys* 117:1–19
233. Refson K (2000) Moldy: a portable molecular dynamics simulation program for serial and parallel computers. *Comput Phys Commun* 126:309–328
234. Deublein S, Eckl B, Stoll J et al (2010) ms2: a molecular simulation tool for thermodynamic properties. *Comput Phys Commun*, in press
235. Phillips JC, Braun R, Wang W et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
236. Ponder JW (2009) TINKER: software tools for molecular design, 5.0. Washington University School of Medicine, Saint Louis
237. Müller-Plathe F (1993) YASP: a molecular simulation package. *Comput Phys Commun* 78:77–94
238. Jorgensen WL, Tirado-Rives J (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* 26:1689–1700
239. Jorgensen WL, Ibrahim M (1980) The structure and properties of liquid ammonia. *J Am Chem Soc* 102:3309–3315
240. Hinchliffe A, Bounds DG, Klein ML et al (1981) Intermolecular potentials for ammonia based on SCF-MO calculations. *J Chem Phys* 74:1211–1217
241. Sagarik KP, Ahlrichs R, Brode S (1986) Intermolecular potentials for ammonia based on the test particle mode and the coupled pair functional method. *Mol Phys* 57:1247–1264
242. Caillol JM, Levesque D, Weis JJ (1987) A theoretical study of a polar-polarizable model for liquid ammonia. *Mol Phys* 62:1225–1238

243. Mansour KA, Murad S (1987) A computer simulation study of fluid ammonia. *Fluid Phase Equilib* 37:305–325
244. Hannongbua SV, Ishida T, Spohr E et al (1988) Molecular dynamics study of a lithium ion in ammonia. *Z Naturforsch* 43a:572–582
245. Gao J, Xia X, George TF (1993) Importance of bimolecular interactions in developing empirical potential functions for liquid ammonia. *J Phys Chem* 97:9241–9247
246. Kristóf T, Vorholz J, Liszi J et al (1999) A simple effective pair potential for the molecular simulation of the thermodynamic properties of ammonia. *Mol Phys* 97:1129–1137
247. Zhang L, Siepmann JI (2010) Development of the TraPPE force field for ammonia. *Collect Czech Chem Commun* 75:577–591
248. Benedict WS, Plyler EK (1957) Vibration-rotation bands of ammonia: II. The molecular dimensions and harmonic frequencies of ammonia and deuterated ammonia. *Can J Phys* 35:1235–1241
249. Tillner-Roth R, Harms-Watzenberg F, Baehr HD (1993) Eine neue Fundamentalgleichung für Ammoniak. *DKV-Tagungsbericht* 20:167–181
250. Gross T, Buchhauser J, Price W et al (1997) The p, T-dependence of self-diffusion in fluid ammonia. *J Mol Liq* 73:433–444
251. Tufeu R, Ivanov DY, Garrabas Y et al (1984) Thermal conductivity of ammonia in a large temperature and pressure range including the critical region. *Ber Bunsenges Phys Chem* 88:422–427
252. Schnabel T, Vrabec J, Hasse H (2005) Henry's law constants of methane, nitrogen, oxygen and carbon dioxide in ethanol from 273 to 498 K: prediction from molecular simulation. *Fluid Phase Equilib* 233:134–143, 236:272, 239:125–126
253. Hayduk W, Cheng SC (1970) Solubilities of ethane and other gases in normal paraffin solvents. *Can J Chem Eng* 48:93–99
254. Kierzkowska-Pawlak H, Zarzycki R (2002) Solubility of carbon dioxide and nitrous oxide in water + methyldiethanolamine and ethanol + methyldiethanolamine solutions. *J Chem Eng Data* 47:1506–1509
255. Kunerth W (1922) Solubility of carbon dioxide and nitrous oxide in certain solvents. *Phys Rev* 19:512–524
256. Luehring P, Schumpe A (1989) Gas solubilities (hydrogen, helium, nitrogen, carbon monoxide, oxygen, argon, carbon dioxide) in organic liquids at 293.2 K. *J Chem Eng Data* 34:250–252
257. Postigo MA, Katz M (1987) Solubility and thermodynamics of carbon dioxide in aqueous ethanol solutions. *J Solution Chem* 16:1015–1024
258. Takahashi M, Kobayashi Y, Takeuchi H (1982) Diffusion coefficients and solubilities of carbon dioxide in binary mixed solvents. *J Chem Eng Data* 27:328–331
259. Tokunaga J, Nitta T, Katayama T (1969) Solubility of carbon dioxide in aqueous alcohol solutions. Methanol-water, ethanol-water systems. *Chem Eng Jpn* 33:775–779
260. Fredenslund A, Sather GA (1970) Gas-liquid equilibrium of the oxygen-carbon dioxide system. *J Chem Eng Data* 15:17–22

Multiscale Approaches and Perspectives to Modeling Aqueous Electrolytes and Polyelectrolytes

Luigi Delle Site, Christian Holm, and Nico F.A. van der Vegt

Abstract We review recent work on scale-bridging modeling approaches applied to aqueous electrolytes and polyelectrolytes, connecting the local quantum chemical details to classical statistical and thermodynamics properties. We discuss solvation and pairing of ions in water, ways to include solvent degrees of freedom in effective ion–ion interactions, and coarse-grained simulations of polyelectrolytes including dielectric boundary effects.

Keywords Molecular Simulation · Quantum · Atomistic · Coarse-Grained Methods

Contents

1	Introduction	252
2	Positively Charged Ions in Water: Quantum Calculations	253
2.1	Theoretical Framework	253
2.2	Results	254
2.3	Methodological Implications	258
2.4	Beyond the First Solvation Shell	258
3	Classical Models: Detailed-Atomistic and Coarse-Grained Force Fields for Multiscale Modeling of Electrolyte and Polyelectrolyte Solutions	260
3.1	Ion Pairing and Its Thermodynamic Implications	261

L. Delle Site (✉)

Max-Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany
e-mail: dellsite@mpip-mainz.mpg.de

C. Holm (✉)

Institute for Computational Physics, Universität Stuttgart, Pfaffenwaldring 27, 70569 Stuttgart, Germany
e-mail: holm@icp.uni-stuttgart.de

N.F.A. van der Vegt (✉)

Center of Smart Interfaces, Technische Universität Darmstadt, Petersenstrasse 32, 64287 Darmstadt, Germany
e-mail: vandervegt@csi.tu-darmstadt.de

3.2	Towards Coarse-Grained Models for Aqueous Electrolytes and Polyelectrolytes ..	267
4	The Generic Polyelectrolyte Coarse-Grained Model	276
4.1	Polyelectrolyte Multilayers	279
4.2	Polyelectrolyte Translocation Through a Nanopore	285
5	General Conclusions and Perspectives	287
	References	288

1 Introduction

Systematic bridging of length and time scales is one of the central themes of modern chemistry and physics to which computer simulations have provided important contributions in recent years. In particular in the fields of materials science and (soft) condensed matter, which also includes many areas of biophysics, most of the problems are inherently multiscale. This means that the interconnection among the different scales plays a key role in determining the relevant properties of the system. Some examples which may clarify this concept are edge dislocation in metals, where local chemistry affects large scale material properties [1], the cracking of materials where the rupture of a local interatomic bond is then propagated to the larger scale and again back to the next interatomic bond and so on [2, 3], or the adsorption of large molecules on metal surfaces out of a melt [4, 5] or out of solution [6, 7] to cite only a few. In general, an exhaustive description of such systems requires the simultaneous treatment of all relevant scales. However, despite the massive progress of computer technology this often remains a prohibitive task. A possible way to overcome this dilemma is to identify and use only those degrees of freedom that are essential for its physical characterization. This generates immediately the problem of how to know “a priori” which degrees of freedom are relevant and which are not. There is not, and probably never will be, a general method around which is able to separate unessential degrees of freedom from the essential ones. Therefore physical intuition (or experience) is needed for the initial development of the so-called coarse-grained model that can be verified only “a posteriori.” Another problem is that, even if one has a limit set of degrees of freedom, most often one still needs to develop sophisticated methods which combine a solid theoretical approach and an efficient use of computational resources. Such methods can have a sequential hierarchical structure which treats the results of a scale as an input for modeling the next (see, e.g., [8]) or treats all the scales at the same time but taking into account, as much as possible, only the strictly required degrees of freedom (see, e.g., [9–11]). Along the lines drawn above, in this review we report a scale bridging idea performed in a sequential way for aqueous electrolytes and polyelectrolytes. We start from the quantum study of small systems and show how this approach can provide precious information to model more complex systems at the classical atomistic level and from there to the coarse grained larger scales. The focus of this review is largely based on the authors’ own contributions to these fields.

2 Positively Charged Ions in Water: Quantum Calculations

Water as a natural solvent plays a central role in many biological, physical, and chemical processes. Owing to its unique capability of building a hydrogen bonded network, the liquid state properties remain difficult to describe in computer simulations or theoretical calculations. Although many of its important features have been uncovered over the past years by experimental and theoretical research efforts, several problems remain unsolved and their solution is crucial for addressing the current demand in biological, chemical and industrial applications. One problem of central importance is the behavior of water as a solvent of positive ions [12–19]. On a large scale, thermodynamic properties characterize ionic hydration, expressed, for example, by the free energy of solvation. These thermodynamical quantities are experimentally accessible and thus are of interest to theoretical studies whose methodology can reach these large scales, e.g., classical simulations. However, structural properties of the solvation shell, at the level of the single molecule, are characterized by a local microscopic scale (e.g., electronic) and go deep into the very specific chemistry of the ion–water interaction; at the same time the overall structure and stability of the first solvation shells are strongly influenced by the rest of the system (bulk). This interplay of local and global scales is not yet fully understood; this section concerns this latter aspect.

2.1 Theoretical Framework

Chemical intuition would suggest that the dominant ingredient in the solvation process is the direct ion–water interaction [20]. In fact one would expect that water molecules in the first solvation shell of a positive ion are not screened by the presence of other molecules and thus are highly polarized due to the direct interaction of the positive charge of the ion with the lone pair electrons of the oxygen. However, recent experimental [20, 21] and theoretical work [22–30] suggested that this effect is not crucial for the structure of liquid water around the ion. In reality the problem is rather complex and, as anticipated, a detailed understanding of the process requires, at theoretical level, a multiscale analysis that clarifies the interplay between the molecular scale in the vicinity of the solute (local scale) and the bulk scale (global scale) of the solvent. In the following we will present a scale analysis, in the sense mentioned above, of *ab initio* Density Functional (DFT), Car-Parrinello¹ simulations of several positive ions in water, namely Li^+ , Na^+ , K^+ , Mg^{2+} , and Ca^{2+} . This scale analysis is based on the calculation of the molecular dipole moment of water; such a quantity represents a physical indicator of

¹The results discussed here were produced with the version 3.9.2 of the CPMD code. CPMD, Copyright IBM Corp. 1990–2004, Copyright MPI für Festkörperforschung Stuttgart 1997–2001/
<http://www.cpmd.org>.

molecular charge displacement and local packing of the liquid, and can be used to derive an effective range of molecule–molecule correlation [30, 31]. First principles Density Functional (DFT) calculations have proven to reproduce experimental data well for aqueous systems of interest here [32–35]. This success is the reason why in the last decade this approach has also been used to investigate properties for which no experiments are performed [36, 37]. Here, within the DFT framework, molecular dipoles are determined using the Wannier localization method [38–40]. The centers of the maximally localized Wannier functions [38–40] are used to localize the electrons onto point charges as in [36, 37]. It must be underlined that the conclusions obtained from our studies reviewed below do not suffer from the fact that a unique way to localize the electron density in space does not exist [41–44], because our conclusions are based on comparing relative values. In order to identify local effects and somehow separate them from non-local ones, we adopt a procedure which consists of building stepwise the solvation shell of an ion [31]. This means that we study clusters starting with one water molecule (and the ion) up to the number of water molecules corresponding to the typical coordination number of the ion. For each system we then calculate the average molecular dipole moment of water; these results will tell us how any additional water molecule contributes to the change of the average molecular dipole. Next we compare such results with the results obtained for the molecules in the first solvation shell of the ion in bulk liquid water; this will tell us what is the influence that the molecules beyond the first solvation shell have on the polarization of those in the first shell, since the molecules of the bulk are not present in the cluster calculations. This is a practical way to link the local ion–molecule contribution to the more global water–water contribution. Technical details of the simulation set up can be found in [30, 31, 45].

2.2 Results

The left part of Fig. 1 shows the trend of the molecular dipole of water as a function of the clusters size. For one water molecule the chemical intuition is obviously correct; the molecule is strongly polarized and the divalent ions induce a larger polarization compared to the monovalent ones, as one would expect. But, interestingly, such a polarization rapidly decays as the number of water molecules increases. This is understandable from a general point of view because the presence of any additional molecule induces some electrostatic repulsion between the polarized electron clouds; however, here chemical intuition started to be less predictive. In fact the decay of the dipole value is much larger for the divalent ions than the monovalent ones; the ion–water interaction should still be very dominant but eventually, for large clusters, the difference between monovalent and divalent ions becomes roughly half that of a single molecule system. The maximum difference between monovalent and divalent ions, for comparable clusters (five molecules), is about 0.7–0.8 Debye compared to the 1.7 Debye of a single molecule system. The minimum difference is about 0.5 Debye for large clusters and 1.0 Debye for

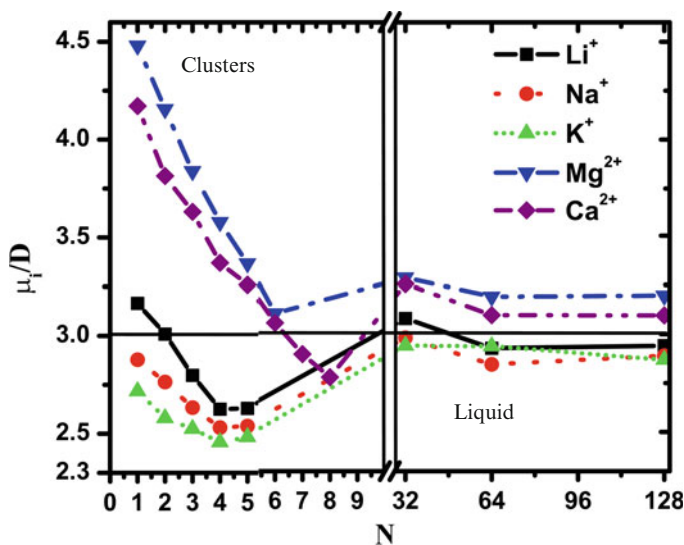


Fig. 1 Average molecular dipole of water. *Left part* for clusters of one molecules up to the typical coordination number of the specific ion. *Right part*, average molecular dipole of water molecules in the first solvation shell of the ion in a liquid system for three different system sizes of 32, 64, and 128 water molecules. Figure adapted from [31]

one molecule. These numbers clearly show that the direct ion–water interaction is not as dominant as expected. Furthermore, the comparison of the left side of Fig. 1 with the right side shows that, when the solvation shell is embedded into a liquid system, the average dipole moment of the molecule in this shell converges to the average value of bulk water, 3.0 Debye, independently from the size and the charge of the ion. The results of Fig. 1 lead to the conclusion that water–water interaction actually plays the principal role in the overall charge displacement of the molecules in the first solvation shell of an ion. There are two effects that contribute to this scenario: (1) the interaction among molecules in the solvation shell via the electrostatic repulsion between the electron clouds and (2) the interaction of those molecules with the rest of the system (bulk), which creates a hydrogen bond network and thus capture electron charge along these bonds. However, the presence of an ion in water should still have clear effects because it perturbs the typical structural order and density fluctuations of a pure liquid system. Although the absolute value of the molecular dipole in the first solvation shell is mainly determined by water–water interaction, it is expected that water molecule are oriented with their dipoles collinear with the ion–oxygen direction and the question is whether this ion-induced order extends to only the molecules in the first solvation shell or, at least in part, also goes beyond.

Figure 2 shows the orientation of the molecular dipole with respect to the oxygen–ion direction for different solvation shells. A clear signature of the orientation of the dipole along the ion–oxygen distance is represented by the peak at zero degrees of the distribution for divalent ions, while the same is less evident, though

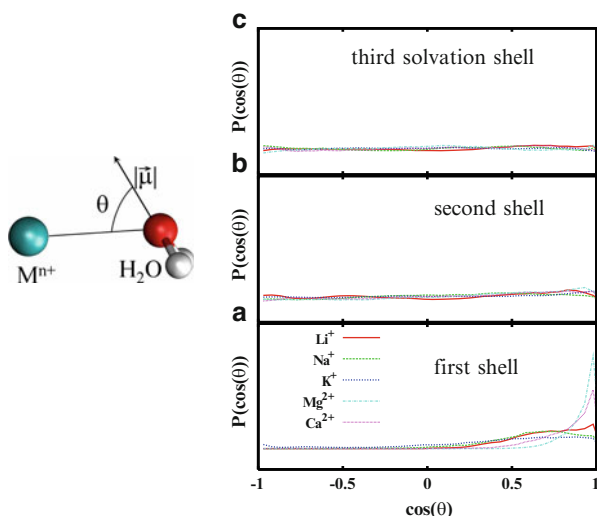


Fig. 2 Water dipole average orientation with respect to the oxygen-ion direction for different solvation shells. Some structural order can be seen only in the first shell of the divalent ions. Figure adapted from [31]

present, for monovalent ions and depends on the size of the ion. However, independently from the size and the charge of the ion, in the second shell there is no preferential structure. This means that water molecules beyond the first solvation shell do not feel the presence of the ion.

2.2.1 Electronic Analysis

An interesting question is how the results previously shown are linked to the specific electronic properties of the molecules [45]. In this section we perform an analysis of the results shown before in terms of electronic orbitals. With this aim, we decompose the various component of the molecular dipole moment, calculated by taking the oxygen as the origin of the Cartesian system: (1) the lone pair, negative charge, contribution of the valence electrons of the oxygen, (2) the bonding pair, negative charge, contribution of the electrons shared by the oxygen and the hydrogen along the O–H bond, and (3) the positive charge of the protons (see also Fig. 3a). Figure 3b shows the comparison for three representative ions, namely Mg^{2+} , Ca^{2+} , and Na^+ . The values are calculated taking the isolated water as a reference. In all three cases, on average, the positive charge of the protons does not contribute significantly (in absolute term this contribution is the same as in the gas phase). The main role is played by the balancing of the lone pair and bonding pair deformation. Interestingly, above all for divalent ions, the lone pair contribution rapidly decays as the number of molecules increases, in contrast to the bonding pair contribution. The negative values of the bonding pair contribution must be intended as the difference between absolute values (of such contribution) of the ion–water system and the isolated water (reference).

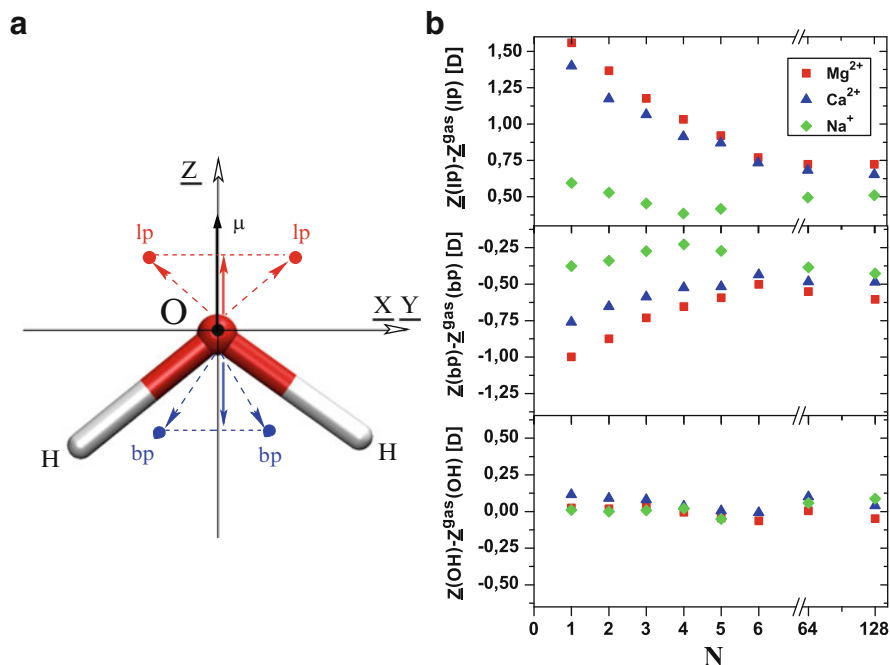


Fig. 3 (a) Pictorial representation of the projections of the charge vectors along the molecular axis. O-lp is the vector with origin at the oxygen which identifies the Wannier center for the electrons of the lone pair orbitals; this vector is indicated by lp. Analogously bp, for bonding pair, and OH, for the protons. μ represents the total dipole vector. (b) Projection of the oxygen-Wannier-center vectors (times the charge) along the molecular axis of water (in Debye) measured w.r.t. the isolated molecule. Figure adapted from [45]

These results basically show, at electronic level, what was already speculated on a larger scale in the previous section: as the number of water molecule increases, the repulsion of the electronic clouds of the lone pair electrons increases and thus the corresponding contribution decreases. This in turn leads to a partial charge redistribution of the electrons of the bonding pair. In fact, these electrons are also polarized by the presence of the ions and their orbitals are shifted from the hydrogens towards the oxygen in the oxygen–ion direction; however, since the repulsion increases as the number of molecules increases, the electrons are shifted back towards the hydrogens. Moreover, the rapid increase of the bonding pair contribution for larger clusters is due to the formation of hydrogen bonds that shift the bonding pair electron distribution further towards the hydrogen as in the case of bulk water. These conclusions can be pictorially described by Fig. 4 for the case of Mg^{2+} where the focus is on one specific molecule of the system, represented together with its corresponding electron density map. The latter is determined by subtracting the electron density of an isolated water molecule with the same geometry of the molecule in the water–ion system.

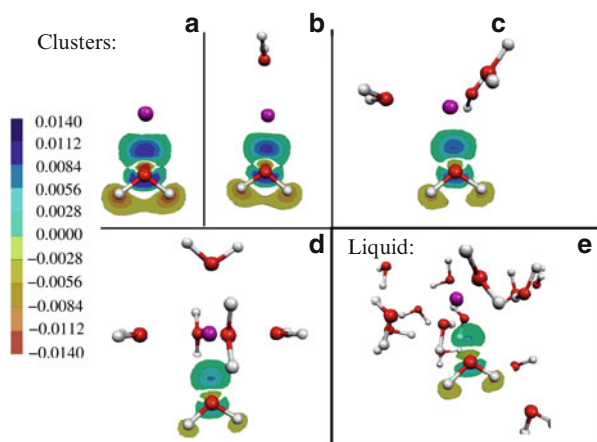


Fig. 4 A two-dimensional section of the density difference, with respect to the isolated water molecule with the same HOH geometry, projected onto the Wannier functions for the Mg^{2+} solvated by the various clusters and in the liquid. The density is measured in $\frac{e}{\text{au}^3}$. The *dark blue color* indicates an excess of charge compared to the isolated molecule while the *light brown* indicates depletion of electron charge. Figure adapted from [45]

2.3 Methodological Implications

Our main conclusions are that an ion produces some small structural effects only on a very local scale and that the molecular charge displacement is mostly due to water–water interaction. The average dipole in the first solvation shell of an ion, independent of the ion size and the charge, is the same as that of the bulk molecules. However, one must take into account that the conditions of our simulation correspond to the case of a dilute system; much higher ion concentrations would indeed change the situation. However, these results clarify why a single spherical positive ion (monovalent or divalent) does not produce the effects expected on the solvent; this conclusion may have far reaching consequences for molecular modeling. In fact a far lasting debate endorses the use of polarizable models of water for classical simulations where ions are present. The DFT calculations discussed here provide indications that non-polarizable water models, developed for bulk water, may perform particularly well in studies of ion solvation, as will be further discussed in the subsequent chapter of this review.

2.4 Beyond the First Solvation Shell

So far we have been focused on the behavior of the water molecules in the immediate neighborhood of the ion. We have concluded that water–water

interaction plays the major role in determining the physical properties of the solvation shell, while the presence of the ion plays almost (only) the role of “excluded volume” perturbation which acts very locally. This is clearly shown by study of the distribution of molecular dipoles and their orientation with respect to the oxygen–ion direction. However, one may perform a finer analysis based on the very microscopic (electronic) details and understand whether, on a finer scale, some physical effects, produced by the ions, have consequences regarding the properties of water, which in the dipole analysis are averaged out and thus not visible. Such an analysis was performed by Scipioni et al. [46], who have analyzed the distributions of the Wannier centers of the system. This is a finer scale analysis, compared to that of the dipoles (also based on Wannier Centers), because it considers each single localized electron center and its changes due to the presence of other molecules or ions; its contribution is not averaged out (with those of the other centers) as instead happens when the same quantity is employed to determine the molecular dipole. Scipioni et al. have shown that the electronic structure of water molecules in the bulk is modified due to the presence of the ion compared to the case of pure water with no ions. The hydrogen bonding network is slightly altered because of a lowering of the local tetrahedral order. These results in any case do not change the conclusions of the previous paragraph and actually strengthen the idea that, for low concentrations, positive ions play the role of perturbation to the pure water rather than an active physical role with effects on the whole system. For this reason our suggestions for the construction of classical models remain valid. In fact the hydrogen bond network distortion of the bulk water occurs on a very fine scale which does not induce relevant thermodynamic consequences on the scale covered by classical simulations. However, indirect effects of the hydrogen bonding distortion may be related to the statistical fluctuations of the solvation shell structure. Density fluctuations around the solute play an important role for both structural and thermodynamic properties [47]; for this reason, an analysis complementary to that performed by us should consider the exchange of molecules in the solvation shell. The typical theoretical tools used for this analysis is that of rare events type [48]. This tool requires a system size and time scale that are usually beyond the current quantum-based methods and apply only to classical simulations; however, here the multiscale idea may play an important role and the hope is that current techniques, employed so far for test systems only, can be extended to these problems as well [217]. Another point to put in perspective is the possibility of analyzing ions of non-spherical shape and defining general statistical quantities that define well the solvation shell, which in this case is no longer homogeneous. Recent ideas in this direction have been provided by Babiacyk et al. [49] who have proposed a combined distance–orientation distribution. This can give more details about the solvation structure and yet provide a general classification of the solvation properties. Combination of statistical methods and straight molecular dynamics simulation on different scales are the route to follow for the next few years in this field in order to build a closely connected exchange between theory and experiments and the multiscale idea based on quantum calculations will certainly play a major role.

3 Classical Models: Detailed-Atomistic and Coarse-Grained Force Fields for Multiscale Modeling of Electrolyte and Polyelectrolyte Solutions

In aqueous electrolyte and polyelectrolyte solutions, electrostatic interactions are strongly screened, owing to the large dielectric constant of the aqueous medium. In effect, the interaction between two ions in water at room temperature only exceeds $k_B T$ when the distance between them is smaller than 0.7 nm. At this distance the ions are separated by approximately three water molecules; hence the effective ionic interaction can be fully understood only when solvent molecules are explicitly considered. Monovalent and divalent ions polarize molecular dipole orientations of surrounding water molecules only within their first hydration shells, as indicated in Fig. 2. Hence, the electrostatic fields emanating from ions in water are to a large extent screened beyond the first hydration shell and, therefore, electrostatic interactions between ions in water are expected not to be strong enough to justify that water–water (hydrogen bonding) interactions can be left out in the description of inter-ionic interactions and interactions between ions and polyelectrolytes.

Global thermodynamic properties of complex systems are often affected by the local-scale chemical details. In recent years, experimental and computational studies of aqueous systems containing charged and uncharged polymers, polypeptides, and proteins have provided ample evidence that effects of ions are local, ion-specific, and involve direct interactions with macromolecules and their first hydration shells [50–54]. These systems, however, are usually too large to be described with theoretical models that include all those details. Local chemical processes in these systems may moreover depend on global system properties determined by structural organization or dynamical processes at large time and length scales, i.e., processes on local and global scales are interdependent. Computer simulations that use only one type of Hamiltonian (quantum-mechanical, classical atomistic, or classical coarse-grained) cannot provide a complete understanding of these complex systems and multiscale simulation approaches are needed instead. Here, we limit the discussion to hierarchical multiscale simulations. In hierarchical simulations, quantum-mechanical, classical-atomistic, or classical coarse-grained Hamiltonians are used consecutively. The different Hamiltonians each describe the same system with different resolution (different number of degrees of freedom), which is achieved by “parameter-inheritance” and exchange of configurations by means of forward and backward mapping procedures [55, 56]. Systematic coarse graining refers to deriving a coarse-grained (CG) model based on molecular simulations with a higher resolution, quantum-mechanical or classical atomistic model [56].

Linking chemistry and properties in computer simulations has made huge progress in recent years. However, significant challenges remain which are the subjects of many studies performed today, where systematic coarse-graining methods are being further explored. Some recent examples of our own research will be reviewed here. Section 3.1 presents a short review of ion pairing, a phenomenon which is determined by chemistry-specific, local details, but which potentially affects

large-scale properties of biological systems including proteins, lipid bilayers, and polyelectrolytes. In Sect. 3.2, coarse-graining methods are discussed that attempt to include chemical-specificity in effective potentials for aqueous electrolytes. The methods discussed in this section may also be used to derive effective potentials for ionic interactions in systems containing peptides, proteins, polyelectrolytes in solution or at (charged) interfaces, and so on. To arrive at a complete description of CG models for macromolecular systems in aqueous solution, we finally also briefly review CG models for polymers and hydrophobic moieties in water, emphasizing some aspects of model transferability.

3.1 Ion Pairing and Its Thermodynamic Implications

Formation of ion pairs in aqueous solution influences thermodynamic properties such as the salt activity and osmotic pressure, double layer interactions and surface forces [57], polyelectrolyte conformations [58], protein–protein interactions [59], and so forth. When ions pair up, their osmotic activity decreases, water activity (vapor pressure) increases, electrostatic double layer repulsions decrease, proteins precipitate, and so on. Despite their importance, the basic mechanisms of ion pairing are poorly understood, even in “simple” electrolyte solutions. In particular, aspects of solvent mediation at short length scales, and various types of the chemical specificity resulting from dispersion interactions, hydrogen bonding, and electronic polarization, cannot easily be described with a simple physical model. Early primitive models [60] used the electrostatic interactions between ions and the ionic radii as key physical ingredients, but replaced the solvent with a dielectric background that attenuates the ionic interactions. Effects of solvent degrees of freedom on short-range inter-ionic interactions are therefore not captured by these models. Primitive models have nevertheless been very successful in describing, for example, aqueous alkali halide solutions and polyelectrolyte solutions, but it was also realized early on that these models fail to describe the strength of ion–ion interactions correctly when solvent degrees of freedom play an important role [61]. Pairing between large ions serves as an example. Here, the hydrophobic contribution to the ion–ion attraction may become equally large or bigger than the electrostatic contribution [62]. The “law of matching water affinities” introduced by Collins [63, 64] emphasizes the role of the water molecules explicitly, and provides a qualitative explanation of contact ion pairing. This law states that contact ion pairs are formed between ions with comparable hydration enthalpies only. Within the validity of a simple electrostatic model, spherical ions with equal charge and size (radius) have equal hydration enthalpies, the magnitude of which increases with the ion charge density. Strongly hydrated ions (“kosmotropes”) are typically small in comparison with the size of a water molecule; weakly hydrated ions (“chaotropes”) are typically large in comparison with the size of a water molecule. It then follows that contact ion pairs are formed between equally-sized ions (small–small or large–large), while unlike pairs

(small–large) tend to stay solvent separated. Detailed-atomistic computer simulations performed by Vlachy, Dill, and coworkers have confirmed this qualitative picture [65].

3.1.1 Can Classical Force Field Models Predict Properties of Aqueous Electrolytes Quantitatively?

In biomolecular force fields the nonbonded interactions are usually described with Lennard–Jones and Coulomb potentials [66]. Several of the popular non-polarizable force fields use (partial) electronic charges that do not vary in response to changes of the local electric field at the positions of the interaction sites. While such simple force field models do realistically describe hydration thermodynamic properties of nonpolar and polar chemical compounds [66, 67], it is not a priori clear if this also holds for chemical groups that dissociate to form charged species in solution. Molecular simulations with non-polarizable force fields have indeed shown that short-range ion–ion attractions in water, and, as a result, solution osmotic coefficients, depend strongly on the ion parameters and water models used in the simulations [68]. Interestingly, quantum-mechanical calculations have indicated (cf. Sect. 2) that electronic polarization of water molecules in the first hydration shells of mono- and divalent ions is relatively weak compared to water self-polarization [31]. Therefore, non-polarizable classical models should in principle be suitable to describe ion hydration in bulk. In principle, efforts could be made to obtain classical non-polarizable force fields for electrolyte solutions by coarse-graining over electronic degrees of freedom in quantum-mechanical calculations (e.g., using data from CPMD simulations). This approach bears the advantage of not having to rely on experimental input in the development of the model. To this end, an inverse Monte Carlo method has been used in the past [33] in which it was illustrated that a simple single-exponential form describes the short-range lithium–water interaction in bulk water better than the Lennard–Jones interaction type. In view of this result [33], it perhaps is not surprising that recent attempts to parameterize classical force fields based on experimental data have pointed out limitations in applying standard mixing rules to describe the short range ion–water interaction with Lennard–Jones potentials [69, 70]. Successful parameterization could be achieved by applying scaling factors for the ion–water Lennard–Jones interaction, the effect of which reduces the short-range ion–water repulsion [69, 70]. The scaling is particularly substantial for the smaller lithium ion [70]. The standard mixing rules were shown to fail for the fluoride anion as well [71]. A number of recent force field parameterization studies have used experimental salt activity coefficients as model input [69–72]. There, it was shown that the resulting non-polarizable force fields are transferable over a fairly broad range of salt concentrations and provide quantitative accuracy. Similar observations were made by Kalcher and Dzubiella [73] who investigated aqueous LiCl, NaCl, and KCl solutions with an existing non-polarizable model. These authors found that the osmotic coefficients

calculated from the exact compressibility route from liquid state theory match experiments for salt concentrations up to 2 M.

3.1.2 Hofmeister Series of the Osmotic/Activity Coefficient

Ion-specific effects often give rise to changes in thermodynamic observables that follow a particular ion series [50, 51, 53]. These series are known as Hofmeister series or lyotropic series. Osmotic coefficients (or the related salt activity coefficients) depend on the ion types at finite salt concentration, while being independent of the ion type in the limit of infinite dilution, where simple electrostatic models (for example the Debye–Hückel limiting law) can be used to describe them. Because small ion–ion separation distances are rarely sampled at low salt concentrations, short-range ion–ion interactions are unimportant relative to the electrostatic forces operative at larger distances. Short-range interactions, however, start playing an important role at finite salt concentrations, leading to the formation of ion pairs which we examine here. It is well-known that osmotic coefficients follow a Hofmeister series as shown in Fig. 5 [74]. When activity coefficients are presented instead, this series remains qualitatively the same. The law of matching water affinities [63, 64] predicts that activity and osmotic coefficients of aqueous alkali bromide solutions decrease in the order $\text{Li}^+ > \text{Na}^+ > \text{K}^+ > \text{Rb}^+ > \text{Cs}^+$, as observed in Fig. 5a. In LiBr solution, a small (“hard”) cation (Li^+) combines with a large (“soft”) anion (Br^-); because small ions and large ions remain solvent separated, the osmotic coefficient of this system is high. In CsBr solution, on the other hand, two large ions are combined; these two ions favor forming contact ion pairs with a corresponding reduction of the osmotic (and salt activity) coefficient. Qualitatively similar ion series are obtained if the bromide anion is replaced by chloride or iodide. Interestingly, a reversed Hofmeister series is obtained upon replacing bromide with the molecular anion CH_3CO_2^- (acetate) as shown in Fig. 5b.

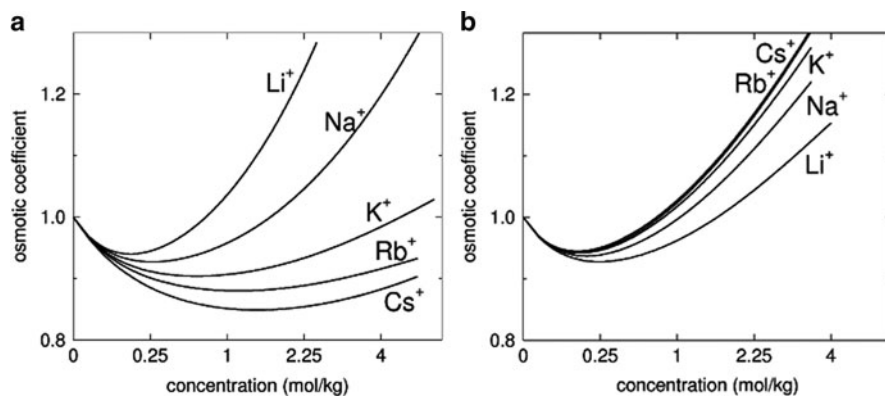


Fig. 5 (a) Experimental osmotic coefficients [74] of aqueous alkali bromides and (b) aqueous alkali acetates as a function of salt molality at 298 K

This particular anion is biologically relevant, because it is found in two amino acids (aspartic acid and glutamic acid) found abundantly on protein surfaces. The smaller osmotic coefficient of Li^+ in this system can be explained in terms of small–small pairing: Li^+ pairs with the negatively charged, small carboxyl oxygen of the acetate moiety. Despite these qualitative insights, the law of matching water affinities does not quantify the effects observed in Fig. 5 and moreover assumes formation of contact ions pairs, neglecting alternative pairing mechanisms.

Figure 6 shows the potential of mean force (PMF) between a sodium ion and a chloride ion in water, at infinite dilution of the two ions, obtained from classical atomistic simulations [75]. The first minimum of the potential corresponds to the contact ion pair (CIP) distance, the second minimum corresponds to the solvent-shared ion pair (SIP) distance, and the third minimum to the solvent-separated ion pair (2SIP) distance. Figure 7a shows an example of a SIP in aqueous NaCl [75]. The infinite dilute potential of mean force in Fig. 6 can be used as an effective pair potential in implicit solvent simulations. The osmotic coefficient $\phi(\rho_s) = \Pi/2\rho_s k_B T$ (with Π the osmotic pressure and ρ_s the salt number density) can be obtained through the virial route. For the case of a binary mixture of components i and j and pairwise additive, density-independent pair potentials, the virial equation can be expressed as

$$\phi(\rho_s) = 1 - \frac{\pi}{3} \rho_s \sum_{ij} \int_0^\infty g_{ij}(r; \rho_s) \frac{d\beta V_{ij}^{\text{eff}}(r)}{dr} r^3 dr. \quad (1)$$

This equation involves the infinite-dilute pair force dV_{ij}^{eff}/dr , while the pair correlation function $g_{ij}(r; \rho_s)$ has to be evaluated at the density considered.

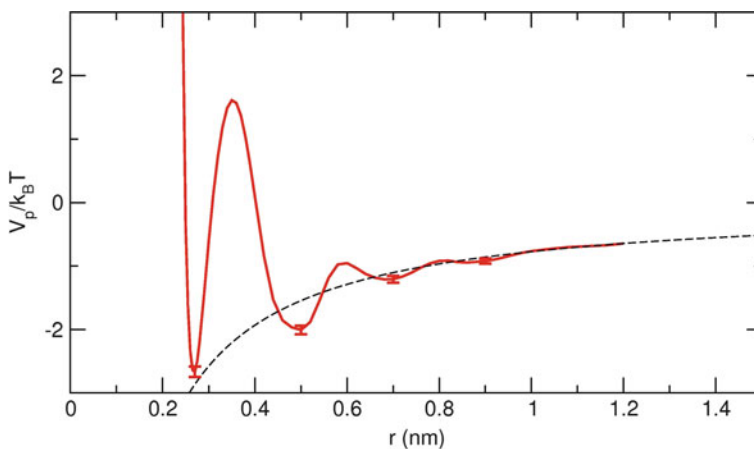


Fig. 6 Potential of mean force between a sodium ion and a chloride ion at infinite dilution in water at 298 K obtained from atomistic simulations with the SPC/E water model (*red curve*). (Adapted from [75]). The *dashed curve* shows the Coulomb potential $-e^2/4\pi\epsilon_0\epsilon r$ with $\epsilon = 71$ the static dielectric constant of SPC/E water

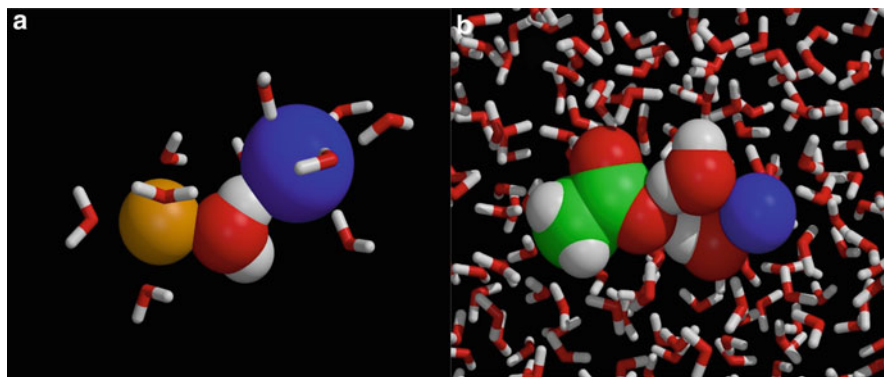


Fig. 7 (a) Solvent-shared ion pair (SIP) in aqueous sodium chloride solution [75]. The central water oxygen (red) coordinates the sodium ion (yellow) while this molecule is at the same time forming a hydrogen bond with the chloride ion (blue). (b) SIP in aqueous sodium (blue) acetate solution [70]. In this system, changes in the excess number of SIPs (the observed number of SIPs minus the number expected at the corresponding distance if all ions are statistically distributed) within the Li^+ , Na^+ , K^+ ion series are responsible for changes in the salt activity coefficient

Equation (1) is not exact, because it employs the infinite-dilute pair potential and multi-body contributions to the effective pair interaction at higher salt concentrations are not considered [75]. These multibody contributions are induced by the water (further described in Sect. 3.2.1), but also include ion multiplet interactions. Direct interpretation of the microscopic electrolyte structure with respect to the osmotic coefficients in Fig. 5 is subtle. Structure and pressure are linked by the virial expression, which is an integration over the pair structure and pair potential weighted by a factor r^3 . This weighting is the reason that the osmotic coefficient cannot be related one-to-one to peak heights of the pair correlation functions without also considering the minima and the whole oscillatory form of these functions [73]. To relate the contributions of CIPs, SIPs, and 2SIPs to the osmotic coefficient, different spatial regions corresponding to the peaks in $g_{ij}(r; \rho_s)$ can be integrated separately; however, this procedure remains approximate due to the assumption of pairwise additivity.

While linking structure and thermodynamics based on the virial expression is not straightforward, this link can in fact be established using an alternative description based on Kirkwood–Buff (KB) theory [76]. Whereas the virial route requires information on the effective potential, the KB description does not make any assumption on the nature of the potentials, is exact, and its central quantities can be interpreted in terms of local solution structure. To this end, we consider the derivatives of the salt activity with respect to the density at constant pressure p and temperature T . For the systems shown in Fig. 5 these derivatives show the same order as the osmotic coefficients/salt activities for the different ions [70]. Hence, the microscopic mechanism explaining the order among the derivatives of the salt activity for the different ions also explains the Hofmeister series for the activities obtained by integration of the derivatives. Based on this, the relation between

thermodynamic changes and formation of CIPs, SIPs, and 2SIPs can be examined in a relatively straightforward manner. To see this, we express these derivatives in terms of *excess coordination numbers* by means of the Kirkwood–Buff theory of solution [76, 77]:

$$\left(\frac{\partial \log a_s}{\partial \log \rho_s} \right)_{p,T} = \frac{1}{1 + \Delta N_{ss} - \Delta N_{ws}} \quad (2)$$

where $a_s = \gamma_s \rho_s$ the salt activity, γ_s denotes the molar scale salt activity coefficient, and ΔN_{ss} and ΔN_{ws} denote the salt–salt and water–salt excess coordination numbers defined as

$$\Delta N_{ij} = \rho_j 4\pi \int_0^\infty [g_{ij}(r) - 1] r^2 dr. \quad (3)$$

Differences in $\Delta N_{ss} - \Delta N_{ws}$ within a Hofmeister series result from short-range contributions to this integral. If we assume that contributions to the integral in equation (3) vanish beyond distance R , ΔN_{ij} can be interpreted as the change in the average number of particles of type j in a spherical region of radius R caused by placing a particle of type i at the center of the region. Hence, ΔN_{ij} is a measure of the affinity between particle types i and j . Molecular simulations of alkali chlorides and alkali acetates showed that (at 1 M salt) ion specificity, as expressed by the denominator on the right hand side of equation (2), arises from the interactions between oppositely charged ions. We therefore replace the denominator $1 + \Delta N_{ss} - \Delta N_{ws}$ with $1 + \Delta N_{\pm}$:

$$\left(\frac{\partial \log a_s}{\partial \log \rho_s} \right)_{p,T} \approx \frac{1}{1 + \Delta N_{\pm}}. \quad (4)$$

This equation relates thermodynamic changes to a measure of affinity, determined by the local electrolyte structure. We can furthermore write

$$\Delta N_{\pm} = \Delta N_{\text{CIP}} + \Delta N_{\text{SIP}} + \Delta N_{\text{2SIP}} + C \quad (5)$$

with ΔN_{CIP} denoting the excess number of CIPs, obtained by integration over the first peak of $g_{\pm}(r; \rho_s)$, ΔN_{SIP} denoting the excess number of SIPs, obtained by integration over the second peak, etc. Equations (4) and (5) provide a route to link thermodynamics to structure by means of integrals over peaks of the pair correlation function corresponding to CIPs, SIPs, and 2SIPs. All ion-specificity for the systems in Fig. 5 was observed in the three excess coordination numbers ΔN_{CIP} , ΔN_{SIP} , and ΔN_{2SIP} , while the constant C in equation (5), corresponding to the contribution from distances larger than approximately 0.8 nm, was the same for all cations within the alkali ion series investigated [70]. It was found [70] that the series shown in Fig. 5a (alkali bromides) can be explained with the observed changes in ΔN_{CIP} , in

agreement with the interpretation based on the law of matching water affinities. However, the behavior shown in Fig. 5b (alkali acetates) could not be explained with a contact pairing mechanism. Instead, this Hofmeister series was shown to result from changes in ΔN_{SIP} among the different cations. Hence, a solvent-mediated pairing mechanism, shown in Fig. 7b, explains the thermodynamic changes in this system. Hydrogen bonding within the hydration shells plays an important role in this mode of ion pairing. A similar solvent-mediated interaction mechanism was more recently observed in molecular simulations of aqueous solutions with dimethyl phosphate anions [78]. This system may serve as a model for describing interactions of alkali cations with phosphate groups on nucleic acids.

This example shows that, in contrast to the approximate virial route in equation (1), a link between thermodynamic changes and microscopic electrolyte structure can be made using an analysis based on Kirkwood–Buff theory. It illustrates, moreover, how different modes of ion pairing with a given cation may lead to *opposing* thermodynamic changes in systems with different anion types. Clearly, effective potentials used in coarse-grained models should ideally retain enough sophistication in order to distinguish CIP states from SIP states.

3.2 *Towards Coarse-Grained Models for Aqueous Electrolytes and Polyelectrolytes*

Systematic coarse graining aims at developing models which, despite improved computational efficiency, preserve chemical specificity and accuracy. Apart from these “specific” CG models, which have gained significant popularity during the past decade, “generic” CG models have been used much longer. In specific and generic CG models, the degrees of freedom are usually chosen with an eye to the required system size and the level of resolution, required to describe the physical and chemical phenomena of interest. The effective potentials for the “remaining” degrees of freedom of a specific CG model are determined by averaging over the “lost” degrees of freedom of the specific chemical system it is intended to describe. In generic models, there exists no such link to a detailed chemical system. Potentials for bonded and nonbonded interactions in generic models are usually kept simple, containing a small number of free constants needed to describe the generic physical behavior of various classes of materials systems.

Systematically CG models for specific systems are often obtained by sampling probability distributions of the “remaining” degrees of freedom with a high resolution model. The sampling method achieves statistical averaging over the “lost” degrees of freedom. The effective potentials obtained from the probability distribution functions are state-dependent, i.e., dependent on density (concentration) and temperature. The dependence on state limits the thermodynamic transferability of the effective potentials. Although a plethora of methods have been developed to obtain effective CG potentials [79], which in turn have been used to simulate

various systems under different conditions, little attention has been devoted to the question of transferability [75, 80–84]. In this section we provide examples of coarse graining approaches which led to CG models with improved thermodynamic transferability. We will focus on aqueous solutions and polymers, covering implicit solvent CG models for electrolytes and hydrophobic solutes, as well as conformational sampling of polymers.

3.2.1 Transferable Coarse-Grained Models for Aqueous Electrolytes

In this section we discuss recent progress made in developing implicit solvent coarse-grained models for aqueous electrolytes that, despite improved computational efficiency, preserve chemical specificity and quantitative accuracy. Implicit solvent CG models for ions may be obtained based on calculations of potentials of mean force (cf. Fig. 6) in atomistic simulations at high ion dilution [75]. Alternatively, they may be obtained by an inverse Monte Carlo or an iterative Boltzmann inversion (IBI) procedure that optimizes a set of effective two-body potentials that best reproduce the two-body correlations (i.e., the electrolyte radial distribution functions) at finite salt concentration [85]. Regardless of the procedure, the effective pair potentials obtained at one concentration are not transferable to another concentration, i.e., we cannot generally assume that the multibody potential of mean force of the ions at finite salt concentration can be expressed as a pairwise additive sum of effective two-body contributions determined at another concentration. Thus, no guarantee exists that one set of effective two-body potentials validated at one concentration also correctly describes the liquid structure and thermodynamic properties (i.e., osmotic coefficient/salt activity) at another concentration. A possible solution to this problem has been proposed by Hess et al. [75]. The PMF shown in Fig. 6 cannot be used as an effective two-body potential at finite salt concentration, because the two-body potential is affected by the additional ions through their effect on the dielectric behavior of the solvent. With increasing salt concentration, the solvent dielectric constant decreases due to electrostriction and the effective two-body attractions between oppositely charged ions become stronger. The authors [75] managed to obtain a *transferable* effective two-body potential V^{eff} , by correcting the potential of mean force V_p , obtained from atomistic simulations of only two ions in water, with a Coulomb term that includes a salt concentration dependent dielectric constant $\epsilon_E(\rho_s)$:

$$V^{\text{eff}}(r; \rho_s) = V_p(r) + \frac{q_1 q_2}{4\pi\epsilon_0} \left(\frac{1}{\epsilon_E(\rho_s)} - \frac{1}{\epsilon} \right) \frac{1}{r} = V_{\text{sr}}(r) + \frac{q_1 q_2}{4\pi\epsilon_0 \epsilon_E(\rho_s) r}. \quad (6)$$

In (6), ϵ is the dielectric constant of the atomistic water model and V_{sr} denotes the short range part of the potential which includes effects of dispersion interactions and ion hydration. It could be shown that the effective two-body potential in equation (6) reproduces the osmotic coefficients of aqueous sodium chloride solutions in satisfactory agreement with experiments up to almost 3 M salt [75]. The idea of

accounting for solvent-mediated multi-body correlations by means of a salt density dependent dielectric constant [75] has been adopted by Kalcher and Dzubiella [73] who compared the osmotic coefficients of alkali halide solutions obtained through the approximate virial route to osmotic coefficients obtained through the exact compressibility (Kirkwood–Buff) route. Although these studies [73, 75] have contributed to a better understanding of thermodynamic transferability of the CG implicit solvent pair potentials, it remained unclear whether – in addition – these potentials are transferable with respect to their ability also to reproduce the electrolyte pair correlations at different concentrations. In other words, are the CG implicit solvent pair potentials capable of reproducing the electrolyte structure and thermodynamics in agreement with the atomistic explicit solvent model? This question has been addressed in a recent study [86] that used aqueous sodium chloride as a model system. The osmotic pressures and pair correlation functions obtained with the CG and atomistic models were found to be in satisfactory agreement up to concentrations of approximately 1 M salt, providing evidence that thermodynamics and electrolyte structure (at the pair level) can be described with an effective pair potential which, moreover, is transferable over a fairly broad range of electrolyte concentrations. Discrepancies between the predictions of the atomistic explicit solvent model and the CG implicit solvent model were observed at larger salt concentration. At high salt concentration, the implicit solvent CG model failed to reproduce the sodium–sodium and chloride–chloride pair correlations obtained with the atomistic model, leading to an overestimation of the osmotic pressures, while the sodium–chloride pair correlations could still be reproduced.

3.2.2 Are Hydrophobic Interactions Between Coarse-Grained Moieties Pairwise Additive?

To develop an implicit solvent CG polyelectrolyte model by systematic coarse graining, several types of interactions need to be considered in addition to the electrostatic interactions discussed in the previous section. Large ions have a propensity for the water/vapor interface [87] as well as for nonpolar groups on protein surfaces [59] and are known to salt-in nonpolar molecules in water [88–91]. The physical origin of these effects can in part be explained by enhanced solvent density fluctuations vicinal to hydrophobic groups [91] and hydrophobic surfaces [92, 93]. Larger fluctuations imply greater ease of formation of cavities, favoring solvation of large hydrophobic ions. The magnitude of the corresponding, specific interactions can be quantified based on PMF calculations in simulations with atomistic, explicit-solvent models. These potentials of mean force have previously been used as effective pair potentials in implicit solvent CG models [57, 59, 68, 75, 94]. Hydrophobic interactions between nonpolar moieties can be treated in a similar spirit [95].

In view of our aim to develop *transferable* CG models, again the question should be asked as to whether the implicit solvent CG potentials – in this case involving CG interactions between nonpolar moieties – are transferable to thermodynamic

conditions with varying solution concentrations. This question has been addressed in a recent study of benzene–water mixtures [84]. There, the authors investigated the variation of the solute (benzene) activity coefficient with a detailed-atomistic and a CG model in the entire range of stable solution concentrations varying between high solute dilution up to the solubility limit where large density fluctuations and corresponding hydrophobic benzene clusters were observed. To this end, the authors developed a CG single-site benzene–benzene pair potential, based on the benzene–benzene pair potential of mean force obtained from detailed-atomistic simulations at high benzene dilution. The CG pair potential contained a hydrophobic contribution in addition to the contribution of direct benzene–benzene van der Waals interactions. It could be illustrated that the thermodynamic changes and benzene clustering obtained with the CG model matched almost perfectly with the predictions of the corresponding atomistic model in the entire range of stable solution concentrations [84]. This result indicates that the hydrophobic interactions in this system are pairwise additive, supporting the ideas of an earlier study by Wu and Prausnitz [96] who examined a larger set of hydrophobic solutes.

3.2.3 Coarse-Grained Polymer Models

Hierarchical modeling of polymers has made significant progress in recent years. The coarse-grained models discussed in this section consist of “united atoms,” which typically merge 5–15 real atoms [55, 56, 97, 98]. Figure 8 shows an example of a polystyrene (PS) fragment in which two coarse grained beads are used to represent the chemical repeat unit. Effective pair potentials are used to describe the non-bonded interactions between CG beads, the bonded interactions including potentials for bond stretching, angle bending, and torsional rotations along CG bonds [56]. Molecular dynamics simulations of polymer melts with these types of CG models are typically four orders of magnitude more efficient than detailed-atomistic models [99]. Hence, long enough time scales are achieved, required to

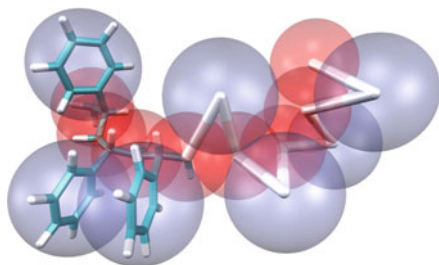


Fig. 8 Coarse-grained poly(styrene) model [98]. The *gray beads* represent the dangling phenyl groups, *red beads* represent the aliphatic (backbone) part of the chemical repeat unit. *Red and gray beads* are connected by coarse-grained bonds. Atomistic details can be reinserted in the coarse grained beads (“inverse mapping”), allowing for scale-hopping between the different resolution levels

equilibrate the chain conformations and melt packing of high molecular weight polymers and to investigate polymer dynamics [100–102] as well as the dynamics of slowly diffusing additives [99]. Today, these models are used to describe static and dynamics properties of soft matter systems at length scales ranging between 0.3 nm and 100 nm and time scales up to 10^{-3} s. Since the degree of coarse graining is moderate, atomistic chain conformations can easily be mapped onto the CG chain conformations, as illustrated in Fig. 8.

Switching resolution, from atomistic to coarse grained and back [56] offers new opportunities for scale-bridging approaches in multiscale modeling of complex soft matter systems. For example, recently developed hierarchical PS models have significantly contributed to computationally efficient, quantitative modeling of chain diffusion in entangled melts [102], liquid and vapor permeation data [99], and detailed-atomistic structures of free-standing polystyrene surfaces [103]. Most of these properties are not amenable to modeling with a detailed-atomistic model alone.

A coarse-grained force field for a selected macromolecule should ideally be developed such that it is transferable to conditions where the chemical environment or thermodynamic conditions are different. Different methodologies exist to obtain the CG potentials for polymer models. The IBI method [104] optimizes a set of CG potentials (bonded and nonbonded interactions) against selected target quantities. These target quantities are obtained from atomistic simulations of oligomer melts (or polymer solutions) and include pair correlation functions between the CG bead centers and probability distribution functions of the bonded degrees of freedom corresponding to the coarse-grained mapping points. The iterative procedure is usually performed under constant NVT conditions. The virial pressure of the atomistic system is often included as an addition target in the iterations. Because IBI is an automated procedure [104, 105], CG models can readily be obtained once an atomistic trajectory is available. However, it usually remains unclear whether the resulting CG models can be applied outside the thermodynamic state point where the IBI parameterization was carried through. Since IBI-derived models are fitted to the structure and pressure at a selected density and temperature, the models may fail to describe condensed phase properties at different temperatures and densities. In a recently developed, alternative coarse graining method [98], coarse-grained bonded and nonbonded potentials have been derived from sampling single oligomers and oligomer pairs in vacuum, respectively, with a detailed-atomistic model. Hence, fitting of the potentials on condensed phase structures or conformations is avoided in this method and the resulting potentials are not biased to represent any environment-dependent (melt, solution) property, such as chain conformations or bead packing characteristics. In order to obtain CG bonded potentials, single oligomers were sampled in vacuum, taking into account all atomistic interactions along the backbone up to a cutoff distance which is determined by the local interactions that influence the bond stretching, angle bending, and torsional degrees of freedom of the CG model. Longer-ranged nonbonded interactions along the backbone, that in the final CG model are accounted for with nonbonded interactions, were all switched off during the single-oligomer sampling stage in order to avoid “mixing”

of bonded and nonbonded contributions in the CG bonded potential. The potentials of mean force corresponding to the CG bonded degrees of freedom, obtained from this sampling procedure, are used as effective potentials to describe bonded degrees of freedom, including bond stretching, bond angle bending, and torsional rotation in the CG model. This procedure was performed for PS oligomers with different stereo-sequences, leading to a set of CG bonded potentials that were used to describe stereo-regular (isotactic and syndiotactic) and atactic PS conformations in condensed phase simulations. Effective pair potentials for nonbonded bead–bead interactions were obtained by computing the free energy of a coupling process that introduces the atomistic interactions between two groups of atoms corresponding to the CG beads on two oligomers in vacuum. Figure 9 illustrates this procedure. It should be noted that the effective nonbonded potentials obtained by this procedure include effects of multi-body correlations related to steric constraints mediated by chain backbone atoms surrounding the chemical moieties for which the CG potential is evaluated. The overall coarse graining procedure of [98] is very cheap (sampling in vacuum) and moreover ensures that bonded and nonbonded interactions are strictly decoupled.

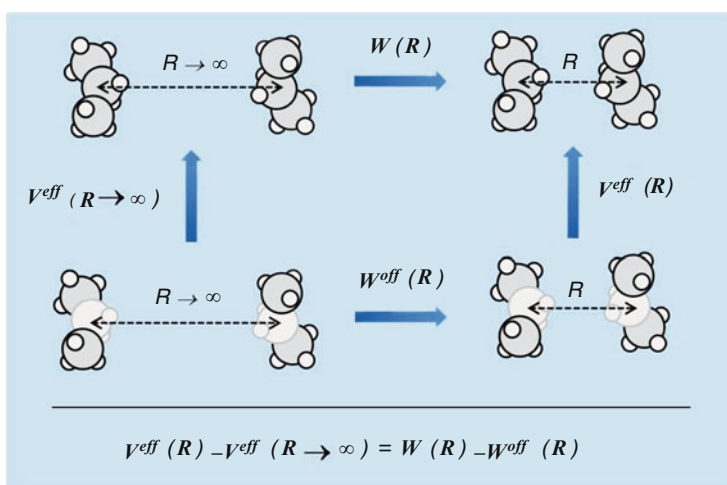


Fig. 9 Thermodynamic cycle used to obtain CG nonbonded interactions [98]. $W(R)$ denotes the reversible work to bring two trimer molecules from infinite distance to distance R computed with a detailed-atomistic force field. The distance coordinate R between the mass centers of the two central methylene units, for which a CG interaction potential is being sought, is used as reaction coordinate. $W^{\text{off}}(R)$ denotes the reversible work to bring the same two trimer molecules from infinite distance to distance R , while all direct nonbonded interactions between the atoms belonging to the two central methylene units are switched off. Propane is chosen here for illustration purposes. In the gas phase, $V^{\text{eff}}(R \rightarrow \infty) = 0$. The CG potential function $V^{\text{eff}}(R) = W(R) - W^{\text{off}}(R)$ represents the reversible work (free energy) expended to introduce nonbonded interactions between all atoms of the methylene moieties and can be used as a CG two-body potential for methylene united atoms in propane

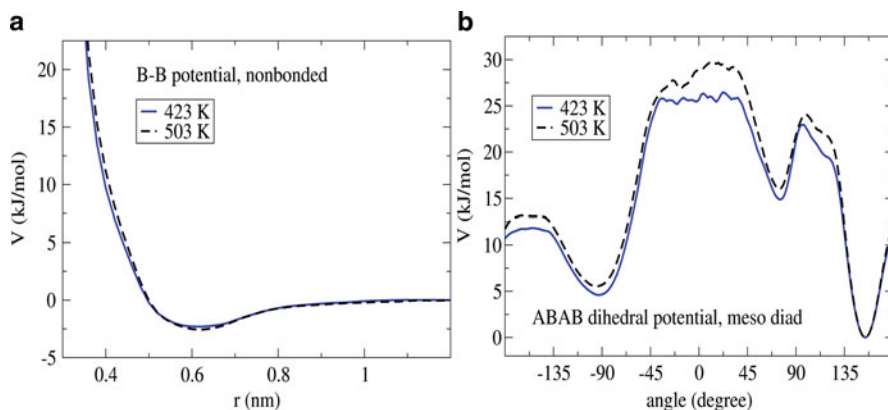


Fig. 10 (a) Nonbonded phenyl-phenyl (B-B) interaction potential and (b) the coarse-grained ABAB torsion potential (A = backbone bead; cf. Fig. 8). The potentials were sampled at two temperatures [106]

Figure 10 shows the CG nonbonded phenyl-phenyl potentials (“B-B”) at two temperatures obtained by applying the procedure depicted in Figure 9 to PS sequences which have been coarse grained according to the mapping scheme shown in Fig. 8 [106]. The attractive tail of the nonbonded potential can be fitted with an r^{-6} distance dependency, while the short-range repulsive part is typically “softer” ($n < 12$) than the r^{-n} ($n = 12$) dependency of the Lennard-Jones potential. The ABAB torsion potential of the meso diad is also shown in Fig. 10. The temperature dependencies of these bonded and nonbonded potentials are noticeable but weak; therefore the potentials obtained at the higher temperature may potentially also be used at lower temperatures. Although the parameterization procedure is cheap enough to be repeated at various temperatures, the potential obtained at 503 K was used in lower temperature simulations of PS melts in order to test the temperature transferability of the model.

Figure 11 shows the density of polystyrene 10-mers in the liquid phase (1 atm) obtained with the detailed-atomistic and CG models. Note that the density is a prediction of the CG model, since condensed phase data have not entered the parameterization. The CG model, parameterized at 503 K, is transferable to lower temperatures in a range as large as 100 K. Another IBI-derived PS model, with one CG bead per chemical repeat unit, was recently shown not to be transferable to other temperatures [107]. The temperature dependence of the density observed in that work was opposite to the dependence observed with the atomistic model (i.e., the CG model predicted a negative thermal expansion coefficient). The poor temperature transferability was suggested to be caused by the coarse description (one bead per repeat unit), which, for example, does not permit one to account for subtle changes in the liquid state packing of the phenyl groups [107].

The coarse-grained PS model in Fig. 8 furthermore describes the global chain conformations of stereo-regular PS melts in satisfactory agreement with

Fig. 11 Liquid density (1 atm.) of polystyrene 10-mers as a function of temperature obtained with the detailed-atomistic and coarse-grained models [98, 106]

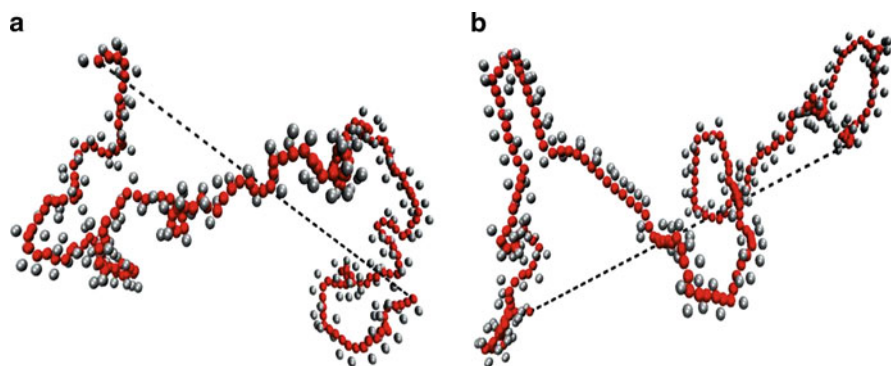
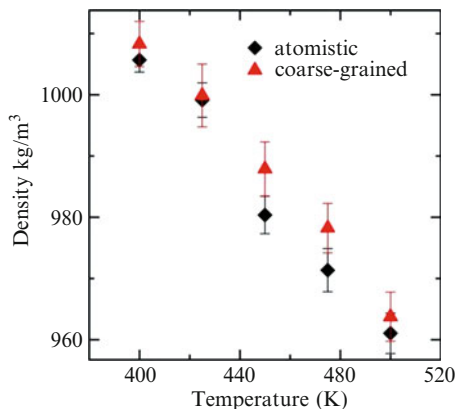


Fig. 12 (a) Snapshot of an atactic PS conformation in the melt [106]. (b) Snapshot of a syndiotactic PS conformation in the melt –*trans* configurations along the backbone leads to a larger radius of gyration compared to atactic PS and cause kinks that may serve as a crystallization nucleus at low temperatures

experiments [98]. At the local scale of a few chemical repeat units, chain conformations sampled with the CG and atomistic models were also in good agreement, hence properties related to the global *and* local conformational behavior can be realistically modeled with the CG model. Figure 12 shows two chain conformations taken from simulations of atactic and syndiotactic PS melts at 503 K [106]. Because neither the bonded potentials nor the nonbonded potentials have been biased to reproduce any structural property in the condensed phase, it can be justified using them not only in simulations of amorphous melts, but also in simulations of polystyrene in solution or at interfaces. Preliminary hierarchical simulations of free-standing PS surfaces have provided indications for surface-induced crystallization of syndiotactic PS [103], thereby significantly extending the scope and applicability of the isolated chain approach to molecular coarse graining [98].

3.2.4 Towards Systematic Coarse Graining of Polyelectrolytes in Solution

Polyelectrolytes in aqueous solution are frequently modeled with generic CG models rather than with systematically coarse-grained ones. The Bjerrum length in these systems is, however, small and contributions of the explicit water molecules to the effective forces between the ions cannot be neglected, in particular in systems where polyelectrolyte concentrations are large. Systematic coarse graining techniques can be used to account for these contributions in the short range part of an effective (implicit solvent) ion–ion potential. A so far non-existing implicit solvent CG model for polystyrene sulfonated (PSS) may be obtained using the isolated chain approach to molecular coarse graining [98]. To this end, the PS mapping scheme of Fig. 8 can be used (alternative mapping schemes may also be chosen), augmented with a CG bead that represents the sulfonate group connected to the phenyl bead. Conformational sampling with isolated atomistic PSS chains in vacuum provides the bonded potentials of the CG model. At this stage, the electrostatic interactions between the sulfonate groups may be set to zero (e.g., by using a SO_3H group instead of the charged SO_3^- group); hence the CG bonded potentials are not influenced by the electrostatic condition, which is described by nonbonded potentials determined independently. The nonbonded implicit solvent potentials between the various PSS beads can be obtained from pair potential of mean force calculations in explicit water, using chemical compounds resembling the CG beads, along the lines investigated previously for small peptides [94]. This approach, however, does not account for the bead being part of a polymer chain, nor does it take into account that the hydration properties of isolated chemical compounds are different from those of the corresponding chemical groups within a macromolecule. It is, for example, unclear to what extent the effective nonbonded interaction between two benzene molecules in water resembles the effective interaction between two phenyl beads on PSS chains in water. To account for both bead connectivity and differences in aqueous hydration, the approach in Fig. 9 can be invoked. Note that, in this case, not only the direct interactions between atoms contained by the central beads of the two solutes are switched off in the lower part of Fig. 9, but also the interactions between these atoms and the solvent. Hence, unlike in the gas phase, the effective potential $V^{\text{eff}}(R \rightarrow \infty) \neq 0$ and is determined by the conditional hydration free energy of the central bead, the condition being that interactions between the atoms in this bead and the solvent are introduced in an environment that has the solute atoms connected to the central bead already “dissolved.” Standard thermodynamic integration or thermodynamic perturbation techniques can be used to calculate the conditional hydration free energy. By means of the way described here, the effects of the solvent, counter ions, and salt on the conformations of the CG polyelectrolyte in solution are modeled through the nonbonded CG interactions, as required. Note that this will not necessarily be the case when alternative coarse graining methods are used. With the inverse Monte Carlo or IBI methods, the effects of the nonbonded interactions at the atomistic scale will be “mixed” between the CG bonded and CG nonbonded potentials, in effect limiting the transferability of the polyelectrolyte model thus

obtained. A concentration dependent dielectric constant [75] can furthermore be used to account for multi-body effects in the effective pair interaction between the charged beads. Although this systematic coarse graining approach has not been investigated so far, it offers a promising future route to studying chemistry-specific effects in polyelectrolyte systems.

Implicit solvent models are in particular efficient in modeling of dilute polyelectrolyte systems. The implicit solvent potentials may however lose validity in dense systems such as polyelectrolyte brushes, polyelectrolyte multilayers, and so on. In these systems, local compositional variations may give rise to strong, unscreened electrostatic interactions in regions where the water content is low as well as to weaker, screened electrostatic interactions in regions with greater water content. To describe such types of systems, CG models are needed that keep an explicit yet coarse-grained solvent description. These models have previously been developed for peptides and hydrophobic solutes in water [84, 108] where one water molecule has been represented by a single interaction site. Recently developed CG water models map four water molecules on a three-bead model that can account for the orientational polarizability of real water [109, 110].

4 The Generic Polyelectrolyte Coarse-Grained Model

The classical way of simulating the static properties of polyelectrolyte solutions with explicit electrostatic interactions and counterions dates back to the early 1990s [111–114]. The model is based on the restrictive primitive model of electrolytes that treats the counterions or salt ions in general as charged hard spheres moving in an implicit dielectric background of relative permittivity to that of water at the desired temperature (i.e., 78 for water at room temperature). The polymer model is that of a bead-spring model, i.e., a Kremer–Grest model [115] where parts of the beads are carrying an explicit charge (that is opposite to the counterions); see Fig. 13 for the mapping scheme for a simple PSS monomer.

Details of the model and some applications can be found in some reviews [58, 117–119]. The solvent is completely structureless, and therefore cannot account for ion specific effects in its simplest form and the other effects mentioned in the previous sections. However, since only coarse-grained polymer beads plus salt and counterions have to be taken into account, relatively large system sizes (meaning many and very long polyelectrolytes) can be modeled. In this way, for example, the structure of the solution, the origin of the polyelectrolyte-peak in the scattering patterns, and scaling properties of polyelectrolyte solutions or polyelectrolyte gels can be studied; see [114, 120–123] for a biased selection of examples. The polymers can each have up to about 300 beads, and 30–50 of them can be investigated. By means of a tailored attraction between the polymer beads, the polymer can be modeled to behave as being in a good, theta, or poor solvent. Polyelectrolytes in a poor solvent are a particular interesting class of PEs, since for those PEs pearl-necklace conformations [124] have been theoretically predicted,

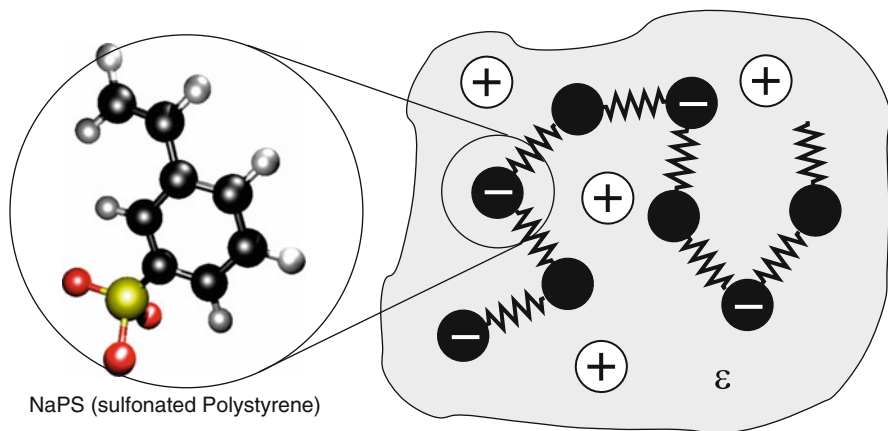


Fig. 13 Mapping scheme of a PSS polymer onto a generic charged bead-spring model with an implicit solvent representation and counterions. Figure adapted from [116]

found by simulations and experiments, and intensively studied during the last 10 years [120, 125–129]. Another interesting polyelectrolyte is, of course, DNA, which is a negatively charged PE. Since the persistence length of DNA is around 50 nm, on short length scales DNA can be considered as a charged rod. This rod can, to a certain extent, be modeled as an infinite charged cylinder within a cell model [130]. Furthermore, synthetic model systems have been synthesized that are strongly charged like DNA and rodlike, and can therefore serve as simple systems, where effects like counterion condensation or mean-field approaches can be studied. We will illustrate here as an example a comparison between osmotic coefficient data of short and stiff polyelectrolyte rods [131] and several simulation studies [132–134]. The target quantity here is the reduced thermodynamic activity of the counterions that can be expressed in terms of the osmotic coefficient:

$$\phi = \frac{\Pi}{\Pi_{\text{id}}}, \quad (7)$$

where $\Pi_{\text{id}} = c_c k_B T$ is the ideal gas pressure at given counterion concentration c_c and temperature T , and k_B is the Boltzmann constant. Experimental studies of strongly charged polyelectrolytes show that ϕ is a weak function of polymer concentration, and typically measured values are of the order of 0.2–0.3 for univalent counterions in the dilute regime; see references in [118]. In our first paper [132] we used the Poisson–Boltzmann (PB) solution of the infinite-rod cell model for strongly charged polyelectrolyte solutions that neglects electrostatic correlations and the finite sizes of the rods and compared it to explicit ion simulations. The correlation effects were found to decrease the osmotic pressure by approximately 7% [132] at low concentrations, but overall the main part of the decrease of the osmotic coefficient was captured by the model. However, in a later

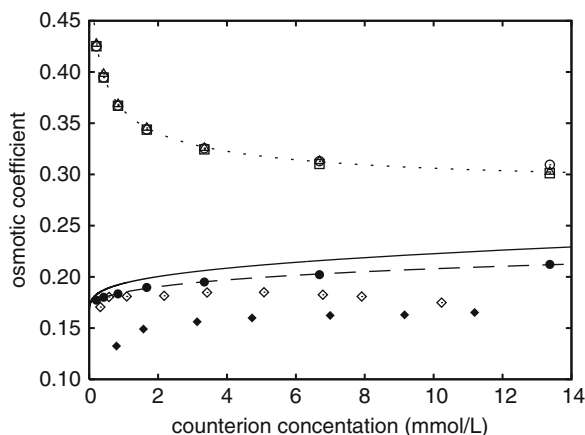


Fig. 14 The osmotic coefficient calculated at different counterion concentrations using various models. The results for a finite rod calculated using the spherical cell model (*empty circles*), cylindrical cell with adjustable geometry (*empty squares*), and bulk system (*empty triangles*) are found to be in very good agreement. The *dotted line* is a numerical fit to the bulk data drawn to guide the eye. The *solid line* is the prediction of the PB theory and the *dashed line* fits the results obtained from MMM1D [135] simulations for an infinitely long rod (*filled circles*). Experimentally measured osmotic coefficient for iodide (*open diamonds*) and chloride (*filled diamonds*) counterions [131] are also shown. Figure adapted from [133]

and more detailed study [133] we found that the finiteness of the chain can lead to even more drastic effects. Figure 14 summarizes our results. The osmotic coefficients obtained via the finite cell models are corrected by the $1/N$ term to obtain the bulk pressure. All finite rod methods yield similar results which are significantly higher than both the theoretical and the simulation data of the osmotic coefficient for an infinite rod. The prediction of the optimal cylindrical cell model in turn was always below the measured bulk pressure. This effect, found to be of the order of 1%, was attributed to the weak polymer–polymer interactions present in bulk but not in the cell model. The simulation of a cylindrical cell periodic in one dimension made it possible to assess directly the counterion–counterion electrostatic correlations. Their effect was found to decrease the osmotic pressure compared with the mean-field PB prediction. This is in good qualitative and quantitative agreement with previous theoretical studies [132] and simulations [136] of strong polyelectrolyte solutions. The correlation effects discussed above are all small in comparison to the large difference between the pressure obtained for a finite and an infinite polymer chain. The finiteness of the chain therefore plays the leading role in the studied concentration range. The $\Pi(c_c)$ curve obtained for the finite macroion not only decreases monotonically, as would indeed be expected at very low density, but also goes well above the curve obtained for an infinite rod. We finish this paragraph by stating that the difference between the simulation and the experimental data of [131] remains unclear up to now, although most relevant corrections to this simple model have been tested by us, but without any success [134].

We will conclude this part of the review by mentioning that if one is interested in solvent mediated interactions in polyelectrolytes, i.e., to study dynamical behavior, then one immediately faces a problem since we have used so far an “implicit” solvent. The obvious thought of including an explicit solvent is only viable for very small system sizes, since the amount of solvent particles needed quickly reaches the range of millions and more particles even for small simulation boxes, and the needed computational time quickly becomes prohibitive. A possible way out is to use a mesoscale solver for solvent dynamics such as the Lattice-Boltzmann (LB) method, stochastic rotational dynamics (SDR), dissipative particle dynamics (DPD), or Stokesian dynamics, and couple the coarse-grained PE model to the background fluid in such a way that the PE can exchange momentum with the fluid background and vice versa, such that correct hydrodynamical interactions are generated [137, 138]. This becomes necessary for electrophoretic applications [139] where the dynamics of the PE under applied fields is studied [140–143]. The usage of simple Langevin (or Brownian) dynamics destroys any hydrodynamic interactions, and, although several studies of dynamical aspects of PE behavior have appeared [144–146], they can be valid only as long as hydrodynamical interactions do not play any role.

4.1 Polyelectrolyte Multilayers

Polyelectrolyte multilayers (PEMs) denote an interesting class of materials that are formed by alternating layers of oppositely charged PEs. In the early 1990s Decher et al. [147, 148] demonstrated the feasibility of building such a type of multilayers using the so-called Layer-by-Layer (LbL) technique. In order to build up a PEM one sequentially exposes a normally negatively charged substrate to a cationic PE solution followed by a rinsing step in order to dispose the supernatant ions and PEs. Then the substrate is dipped into an aqueous solution of anionic PEs, always followed by a rinsing step. Repeating this simple procedure can result in a build-up of hundreds of alternating charged layers. Films made up by PEMs exhibit unique properties which make them suitable candidates for many different uses: matrix materials, bio-coatings, selective membranes, chromatography, optical materials and devices, micro- and nanocarriers, or biocides, just to name a few.

There have been only a few attempts to describe theoretically the electrostatic self-assembly of PEMs, and often they rely on several serious assumptions that are hard to test experimentally [149–151]. Recently simple models based on mean-field descriptions have been developed [152–155] that reproduced better some of the experimental observations regarding the stability and the different growth regimes for PEMs. The film thickness of PEMs normally grows linearly with the number of layers, but exponentially growing PEMs are also known. Strong correlations that exist between oppositely charged polyions provide a formidable challenge to the theoretical description of PEMs.

In order to go beyond mean-field theories, numerical simulations have been used to study PEM formation. The large number of particles involved in the simulation of such systems renders a fully atomistic description of the whole PEM and the solvent computationally too demanding. For this reason, so far only the adsorption of short chains onto a bare surface has been reported using atomistic models [156–161]. The more common approach has been the use of coarse-grained primitive models: the first simulational models aimed at understanding PEM was developed by Messina et al. [162] where the multilayering onto a charged sphere was investigated using Monte Carlo (MC) simulations. This study revealed an important point, namely that additional attractive short ranged forces of non-electrostatic origin were needed to produce stable PEMs, otherwise weakly bound PE complexes would form on the substrate and desorb from the surface again, diffusing into the bulk solution. The difficulty of obtaining a purely electrostatically bound PEM already suggested the possibility that PEM formation cannot be an equilibrium phenomenon. Later Messina [163, 164] studied the multilayering on charged rods and planar surfaces, confirming that the previous findings were not geometry specific, and also for these surface geometries additional non-electrostatic forces were needed to keep the PEMs stable. The main findings obtained up to 2004 were reviewed in [165]. Later Panchagnula et al. [166] used the same model with Langevin simulations to revisit PEM growth on spherical substrates. Their results suggested that the stratified multilayers found around adsorbing particles were only temporal structures whose intermixing could be slowed down via the formation of strong ionic-pairs between oppositely charged monomers. Subsequently, Messina [167] studied the effect of the image forces arising from the dielectric mismatch between surfaces and the surrounding aqueous solution on the adsorption of PE in planar surfaces. It was shown that image forces can considerably reduce the degree of adsorption of PE and can even inhibit the charge inversion process in the topmost layer. The effects of the length and fraction degree of charge of PE on PEM formation of planar substrates were studied by Patel et al. [168] using Langevin simulations, confirming the importance of short-range interactions for all fraction degrees of charge and the considerable intermixing between chains deposited on subsequent steps. Other works followed [166, 169, 170].

Despite the great progress gained during the last two decades of research in PEMs, many questions are still only partially solved. Some examples are: the source of the apparent stability of PEMs and the rationale behind the different types of growth modes (linear vs exponential growth; the role of non-electrostatic short-range interactions on the stability and properties of PEMs [171]; the possibility that PEMs are not true equilibrium structures, but kinetically trapped states [172]; the intermixing of PE over several layers [148, 173, 174] and the ion distribution inside PEMs [175, 176]; the influence of the substrate on the first layers [177, 178]; the influence of the charge density [179–181] and type of counter ions [182, 183] on PEM formation and properties; and the effects of pH, temperature and salt concentration on the properties of PEMs [184, 185], to name just some of the questions.

Numerical simulations can help to shed light on some of these issues. However, they also face some technical problems, namely, if the dynamics is slow in experiment, it will also be slow in the simulation, provided the *correct* interaction potentials are used. Therefore speeding up the dynamics, for example via a reduction of the number of degrees of freedom, is essential. Therefore, to obtain a realistic description of PEMs, current models and techniques need to be further improved.

Recently we have revisited the usage of the generic coarse-grained model to study PEM formation out of solution [186, 187]. There are several issues during the build-up of a PEM. First, one has to develop a suitable simulation protocol which allows one to mimic the dipping process closely in order to obtain realistic adsorbed chain conformations. Since in experiments the surface can be exposed to the solution in the order of minutes, one has to speed up the adsorption process in the simulation. In recent work on the adsorption of one layer of PEM, Carrillo et al. [188] have introduced a so-called “stirring” step where, after a certain time, chains that remain non-adsorbed are randomly repositioned in order to improve adsorption rates. They showed that, without stirring steps, the adsorption might not be in the real saturation regime within the time intervals used in their previous works [189]. We followed a somewhat different route and introduced several dipping-rinsing subcycles in which we “refilled” the bulk solution of the simulation box in order to keep the bulk polymer concentration constant. This is necessary since in a finite box under NVT conditions the adsorption process depletes the “bulk” from polymers. With this we were able to achieve excellent saturation conditions, and we could simulate larger systems with substrate areas A up to $A = 40 \times 40$ (in units of the square of the particle diameter).

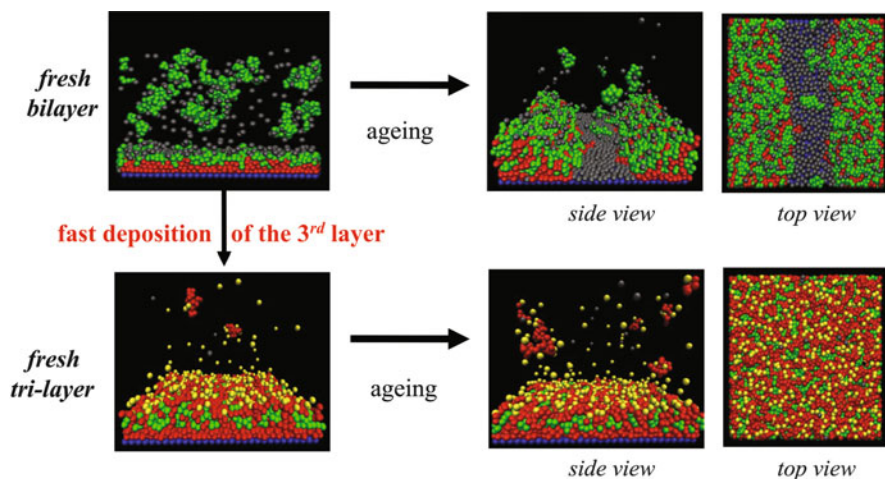


Fig. 15 Simulations showing the thermodynamic instability of a bilayer (*upper plot*). However, a fast deposition step of a third layer stabilizes the then tri-layer (*bottom figure*). The interaction parameters are $\epsilon_{mm} = 1.0$, $\epsilon_{ms} = 10$. Figure adapted from [187]

Our results on a bilayer showed that the standard coarse grained model can lead to a much more diverse behavior than previously reported [189]. Figure 15 shows the temporal evolution of a typical PE monolayer when the system is brought into contact with a solution of PE of opposite sign via successive dipping–rinsing sub-cycles with the same type of PE. Values for the main system parameters are roughly similar to previous works of Patel et al. [188, 189]: the fraction degree of charge is $f = 1/3$ (i.e., one in each three monomers is charged), the density surface charge is set to $\Sigma = Q_{\text{tot}}/A = 0.5625$, and both attractive short-range interactions (monomer–monomer ϵ_{m-m} and monomer–substrate ϵ_{m-s}) are of the order of $k_B T$. Our simulations revealed that the polyanions and the polycations can complex with each other, thus destroying the bilayer and the build-up of a PEM. We observed two different modes of complexation in our model: at low values of the monomer–monomer interaction $\epsilon_{m-m} \sim 0.2$, the bilayer expels just small complexes consisting of only a few pairs of chains. However, when the monomer–monomer attraction is of the order of the thermal energy, $\epsilon_{m-m} \sim 1$, the oppositely charged PEs do form a complex followed by a dewetting of the surface. Patel et al. [189] had already reported that for some longer runs the PEMs surface coverage also gradually decreased. They unfortunately only presented results for PEMs with six and eight layers, but not for bilayers. Thus it is plausible that the PEMs in the study of Patel et al. [189] corresponded to kinetically trapped states, which given enough time and a bulk reservoir large enough to accommodate the adsorbed matter might also redissolve over the course of time. However, we also found that if to the (unstable) bilayer we added sufficiently fast a third PE layer, then the trilayer seemed to be stable in the course of a very long MD run. Also the subsequent addition of a fourth layer did not change the apparent stability, at least within the long run time up to the order of tens of millions of MD steps we could afford. This dewetting process has indeed been observed in some types of PEMs, although it is poorly reported in the experimental literature because they focus mainly on reporting systems and pathways that are able to form PEMs. In principle our findings could also be related to the experimental fact that odd numbers of layers are more stable than even numbers of layers.

To investigate further the critical role played by dispersion forces for the PE adsorption process and for PEMs stability, we performed both CG and atomistic simulations. To investigate the role of dispersion forces in the CG models we took already formed PEMs and progressively reduced the value of the monomer charges q from 1 to 0. We observed that the PEM is able to sustain its inner structured layering almost without modifications even when electrostatic interactions are quite drastically lowered. Only for $q < 0.2$ could noticeable changes in the inner PEM structure be observed. Therefore, it seems that there is a minimum level of electrostatic interaction needed to sustain a PEM structure, but once such a level is attained, the inner structure is almost no further modified by increasing the strength of electrostatic interaction among PEs. This was verified by increasing the charge values up to three times the usual unit charge of monomers.

The most important facts we infer from our studies of the simple coarse-grained PEM model are that:

- Without any surface-monomer short range interaction there is no stable surface coverage.
- Using only electrostatic interactions and purely repulsive monomer–monomer interactions for the excluded volume, no multilayering occurs.
- A short range hydrophobic interaction between all polymeric monomers as well as a monomer–surface attraction is needed to achieve multilayering.
- The precise interaction values need to be fine tuned in order to obtain a stable layering.

This situation is, of course, far from satisfactory. We therefore turned to atomistic simulations of all-atom (AA) poly(styrene sulfonate) (PSS) and poly(diallyldimethylammonium) (PDADMA) systems [160]. The polymers had a low degree of polymerization which was 12 at most. Our main target was to investigate some properties such as the dielectric constant and water structure that cannot be addressed using CG models. Moreover, in a second recent effort, we looked at PSS monolayer simulations for hydrophilic/hydrophobic surfaces and different surface charge [161].

Our AA molecular dynamics simulations on PSS–PDADMA–water–salt (NaCl) mixtures have provided us with another significant insight: when NaCl is added to the PSS–PDADMA–water mixtures of composition similar to real PEMs, the salt ions tend to form clusters inside the mixture [187], which can explain why the addition of moderate amounts of salt has been observed not to modify significantly the structural properties of the PE complexes. In such PE complexes, intrinsic (polyions pair with polycations) and extrinsic (polyions pair with salt ions) charge compensations have been found to co-exist, although the intrinsic mechanism has been observed to be dominant [160]. The addition of NaCl decreases the weight of the intrinsic mechanisms in the PE complexes. But even at the highest simulated NaCl concentrations ($C_{NaCl} = 1$ mol/L), the intrinsic charge compensation mechanism is still the dominant one. Furthermore, our study of such mixtures has allowed us to determine the relative scale of the interaction energy of the ion pairs in the mixture: $Na - Cl > PSS - Na > PDADMA - Cl \approx PSS - PDADMA$ [160]. The relative scale of the interaction energy can be very useful to model properly the interactions between ion pairs in refined CG numerical simulations and theoretical approaches. The analysis of the water structure showed us that PE mixtures are percolated through leading to a homogeneous distribution of waters.

Even though the slow dynamics of water in PEMs has been experimentally studied [190], no theoretical or numerical works have been reported yet. Some recent simulation studies [68, 75, 191] have provided an important insight about the dielectric constant of water in concentrated electrolyte solutions where electrostatic interactions can be almost twice as strong as in pure water. In the work of Qiao et al. [160] the slow dynamics of water inside PSS/PDADMA systems was investigated via the static dielectric constant (i.e., static relative permittivity) and the water diffusion coefficient (see Table 1). Error estimates are determined by block

Table 1 Simulated diffusion coefficients and dielectric constant of water confined within a polyelectrolyte mixture that corresponds to an experimental PEM [160]. Three parallel simulations were performed for system (i), which has the slowest dynamics of water, and two for systems (ii) and (iii). In the parallel runs, only the random seeds in building the initial structures were different. Experimental data based on PSS/PDADMA PEMs

System	ϵ	Diffusion coefficient/ 10^{-7} cm ² /s
i	11.7 $\pm$ 1.4	3.61 $\pm$ 0.01
	11.1 $\pm$ 1.2	3.49 $\pm$ 0.03
	9.9 $\pm$ 0.6	3.65 $\pm$ 0.01
ii	14.0 $\pm$ 0.6	6.02 $\pm$ 0.01
	14.0 $\pm$ 0.4	6.51 $\pm$ 0.01
iii	16.2 $\pm$ 0.7	9.77 $\pm$ 0.03
	16.5 $\pm$ 0.4	9.00 $\pm$ 0.07
Exp. data	<19 [192, 193]	~2 [194]
SPC/E (pure)	71 $\pm$ 6 [195]	260 [160]

averages 20 ns in length. For the pure waters system, the experimental value ϵ is 78.3 [196], and for the pure SPC/E water system, the diffusion coefficient is obtained at ambient condition in an extra simulation of 4,000 SPC/E water molecules.

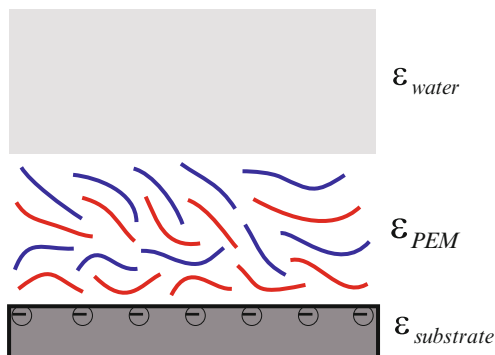
All the observed dielectric constants in the simulations fall within the range of 9–17, which is in good agreement with the experimentally measured values in PSS/PDADMA PEMs [192, 193]. Moreover, it demonstrates that the dielectric constant of water in PSS/PDADMA PEMs is greatly decreased when compared to the case of pure water.

The diffusion coefficients of water are in the range of $3.5 - 9.8 \times 10^{-7}$ cm²/s, which are of the same order of magnitude as the experimental value of about 2×10^{-7} cm²/s measured by pulsed field gradient diffusion NMR [194]. Comparing this with the diffusion coefficient of 2.6×10^{-5} cm²/s in a pure SPC/E water system, one can conclude that in the PSS/PDADMA PECs under study the water molecules diffuse about two orders of magnitude slower than in the pure water system. Therefore, the obtained results of the water diffusion lead to the same conclusion as our dielectric measurements: water molecules inside PE mixtures exhibit a slow dynamics when compared to water molecules in a bulk environment.

The AA simulation of a PSS monolayer that we performed showed that a narrow water-rich region is formed between the substrate and the first PSS adsorption layer. In this region, water molecules show several kinds of orientations, depending on the nature of the substrates. In systems with hydrophilic surfaces, a strong hydrogen-down orientation is observed which diminishes when the surface charge density is increased. The water orientation is strongly related to the formation of hydrogen-bondings of waters with the above PSS sulfonate groups (and the underlying hydrophilic surface groups on hydrophilic adsorbing surface, if present). The dielectric constant of waters very close to the surface is found to decrease almost linearly with the increase of the PSS adsorption from 40 to 15 [161].

These findings demonstrate convincingly that the electrostatic interactions within the PEMs are probably five to six times stronger than in bulk water. This might explain the necessity of ad hoc addition of extra short range “hydrophobic”

Fig. 16 Schematic plot showing the three-regions system suggested in the text as a first model to take into account the mismatch between the dielectric constant in the bulk, on the surface, and inside the PEM. Figure adapted from [186]



attractions to the coarse-grained PE model. However, there might be a valuable alternative to the extra short range attractions. In principle, one could take into account the change in the dielectric constant among the different regions within the system. Of course, an all atom simulation model with explicit solvent would be ideal, since these effects would be automatically included. However, as has been shown before, with that approach currently is not possible to tackle PEM systems due to their large size and long time scales. The simplest approach to try to circumvent the problem of the solvent could be to start with a very simple three-region approach of different homogeneous dielectrics, as shown in Fig. 16. In that approach, the substrate, the layer-region and the bulky-water region would have different dielectric constants. Due to the dielectric boundaries the calculation of the electrostatic interactions is necessarily more complex than in the case of just one homogeneous dielectric background. For planar geometries, the electrostatic boundary conditions can be fulfilled through the inclusion of virtual image charges. Based on this method, we have developed three very efficient and accurate algorithms, which can in one case even reach almost linear scaling [197–199], and which easily allow for the presence of three dielectric slabs. An even more refined model would be to allow the dielectric constant in the multilayer-region to be a variable function $\epsilon = \epsilon(x, y, z)$ of position. In this way one could also incorporate the local salt concentration, which would affect the dielectric constant. This case is much more difficult to handle efficiently such that it can be used within a standard MD or MC calculation. Some possible candidates for such algorithms could be extensions of standard multigrid algorithms [200, 201] and Maggswellian lattice algorithms [202–204].

4.2 Polyelectrolyte Translocation Through a Nanopore

The problem of polymer translocation through nanometer-sized pores has recently stimulated much experimental [205–208] as well as theoretical and simulation based research [209–214]. This is mainly since macro-molecular transport through

pores plays a major role in biological processes and also has a large potential for technological applications. Molecular transport is indeed one of the key functions fulfilled by the plasma and membranes of the cell, and a sizable amount of transport mechanisms which work in the cell are characterized by the same general design, namely by the presence of pores, mostly through membrane proteins. The controlled transport of single molecules through synthetic or biological nanopores is considered as a versatile tool of single molecule sensing and to be a most promising candidate for rapid DNA sequencing.

The complex interplay of interactions – electrostatic, hydrodynamic, and specific chemical ones – and the entropic properties of chain molecules make a full understanding of these systems very difficult. The presence of an interface between the highly polarizable aqueous solution ($\epsilon \approx 80$) and the membrane which is much less polarizable ($\epsilon \approx 2$) leads to repelling forces between charged objects and the pore wall. Since DNA is a highly charged molecule this effect is likely not to be negligible and potentially gives rise to an energetic barrier that the DNA has to overcome in order to transverse the pore. Its characteristics and dependence on the pore size, DNA length, or salt concentration are not known.

The role of the dielectric mismatch between solvent and pore can be investigated via a simple model DNA, consisting of a rigid charged DNA fragment, where the translocation free energy barrier can be easily computed; see Fig. 17 for a schematic plot of the situation.

All simulations were performed both with and without use of the ICC* algorithm [199] to investigate the influence of dielectric mismatch. In a recent article [215] we employed coarse-grained Molecular Dynamics (MD) simulations to compute the mean force acting on the DNA fragment, taking explicitly into account the combined effect of the DNA counterions, salt ions at different ionic strengths, and surface polarization charges generated by the presence of the dielectric mismatch.

It is straightforward then to calculate the free energy barrier by computing the PMF acting on the center of mass of the model DNA along the pore axis. For this reaction coordinate the Fixman potential [216] is constant, and the PMF can be obtained by numerical integration of the mean force. The obtained PMFs are shown in Fig. 18. The free energy barrier in the salt-free case is strikingly higher

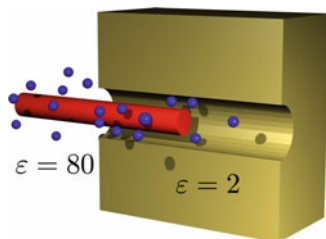
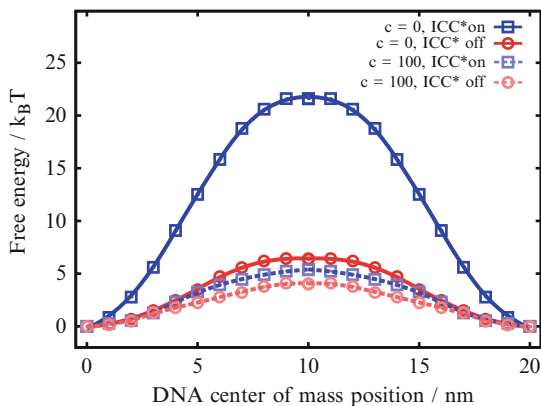


Fig. 17 The plot shows the schematic view of a charged rod confined to a finite nanopore. The environment has a dielectric constant of 80, whereas the material of the nanopore itself has a dielectric constant of 2. The inner part of the pore that is filled with water is also assumed to have a dielectric constant of 80

Fig. 18 Free energy profiles for a DNA segment under two different salt conditions (*solid lines*, $c = 0$ mmol/L; *dashed lines* $c = 100$ mmol/L). We neglected polarization effects (*circles*) and included the effects (*squares*) with ICC*. Statistical error bars are smaller than the symbols. Figure adapted from [215]



(increasing to approximately $20 k_B T$) when polarization charges are taken into account using ICC*, in comparison to the case when they are not considered. In contrast, at a salt concentration of 100 mmol/L, the barrier increase is less pronounced, and the curves obtained with or without taking ICC* into account show a comparable pattern with a barrier height of about $4 k_B T$. The presence of a barrier for the higher salt concentration case, as well as in the case when no polarization charges are considered, can be explained by the steric confinement of the counterion cloud in the nanopore. In contrast, at low salt concentration the Coulomb interaction is not screened, and the effect of polarization charges is maximized, leading to the observed higher free energy barrier.

5 General Conclusions and Perspectives

In this chapter we reviewed some of our recent works on scale-bridging modeling approaches for charged (macro)molecules. On the basis of quantum-chemical calculations, we illustrated that molecular dipoles of water molecules in the hydration shells of positive ions in liquid water are predominantly affected by the water–water interactions. The average molecular dipole moment in the first hydration shell is close to that of bulk water. Classical, non-polarizable water models are therefore suitable to study ion solvation in bulk water. We discussed molecular simulations of ion pairing and the resulting ion specific thermodynamics of aqueous electrolyte solutions containing carboxylate or phosphate anions, which may serve as models for charged groups on proteins and nucleic acids. It was shown that solvent-mediated interactions of these ions with alkali cations favor the formation of solvent-shared ion pairs over direct contact pairs. We then went on to discuss the important field of systematic coarse-graining strategies for electrolyte solutions and charged macro-molecules. We discussed how accurate atomistic and coarse-grained force fields can be obtained for aqueous electrolytes and polyelectrolytes,

and demonstrated in a few examples what one can expect from these models. In particular, we have discussed the thermodynamic transferability of effective potentials used in systematically coarse grained force field models and illustrated how to obtain nonbonded interaction potentials with improved transferability for implicit solvent electrolyte models and polymer melts. Poor transferability of effective potentials, however, remains a concern in several of the systematic coarse graining methods and requires further investigation in future. At the end of this contribution we presented the standard generic coarse-grained model for polyelectrolyte solutions. We argued why this might still be of interest for some large scale solution properties, but also gave some examples which showed the need for the development of more accurate models. Especially in the case of PEMs we still lack a predictive model. For the case of implicit solvent representation we showed that there are important cases like a nanopore where different dielectric regions are present. Here we need fast algorithms that can deal with such situations under various periodic boundary conditions and we gave some first results on the important case of a DNA fragment in aqueous solution passing through a nanopore.

The field of multiscale modeling is rapidly developing, but due to its complexity and wide area of applications, many more thorough investigations are still needed. Specifically we tried to demonstrate that our own attempts for connecting detailed information on the quantum level to coarse grained descriptions, and the subsequent developments of systematic coarse-graining strategies, as well as refined implicit solvent models for charged (macro-) molecules, might be worth following.

Acknowledgements The authors acknowledge fruitful collaborations with Dmytro Antypov, Juan J. Cerda, Dominik Fritz, Berk Hess, Stefan Kesselheim, Vagelis Harmandaris, Kurt Kremer, Hanjo Limbach, Christine Peter, Baofu Qiao, Marcello Sega, Sandeep Tyagi, and Alessandra Villa. Funding from a Volkswagen grant, the SFB 625 TP:C5, DFG grant Ho/1108-17, and DFG Clusters of Excellence 259 and 581 is gratefully acknowledged.

References

1. Lu G, Tadmor EB, Kaxiras E (2006) *Phys Rev B* 73:024108
2. Rottler J, Barsky S, Robbins MO (2002) *Phys Rev Lett* 89:148304
3. Csanyi G, Albaret T, Payne MC, Vita AD (2004) *Phys Rev Lett* 93:175503
4. Delle Site L, Abrams CF, Alavi A, Kremer K (2002) *Phys Rev Lett* 89:156103
5. Delle Site L, Leon S, Kremer K (2004) *J Am Chem Soc* 126:2944
6. Schravendijk P, van der Vegt N, Delle Site L, Kremer K (2005) *ChemPhysChem* 6:1866
7. Ghiringhelli LM, Hess B, van der Vegt NFA, Delle Site L (2008) *J Am Chem Soc* 130:13460
8. Schravendijk P, Ghiringhelli LM, Delle Site L, van der Vegt NFA (2007) *J Phys Chem C* 111:2631
9. Praprotnik M, Delle Site L, Kremer K (2005) *J Chem Phys* 123:224106
10. Ensing B, Nielsen SO, Moore PB, Klein ML, Parrinello M (2007) *J Chem Theory Comput* 3:1100
11. Heyden A, Truhlar DG (2008) *J Chem Theory Comput* 4:217
12. Frausto da Silva JJR, Williams RJP (1976) *Nature* 263:237
13. Seinfeld JH, Pandis SN (1998) *Atmospheric chemistry and physics*. Wiley, New York

14. Blaney DL, McCord TB (1995) *J Geophys Res* 100:14433
15. Wersin P, Curti E, Appelo CAJ (2004) *Appl Clay Sci* 26:249
16. Mickiewicz RA, Mayes AM, Knaack D (2002) *J Biomed Mat Res* 61:581
17. Adams CS, Mansfield K, Perlot RL, Shapiro IM (2001) *J Bio Chem* 276:20316
18. Südhof TC (2002) *J Bio Chem* 277:7629
19. Suzuki K, Sorimachi H (1998) *FEBS Lett* 1–4:433
20. Omta AW, Kropman MF, Woutersen S, Bakker HJ (2003) *Science* 301:347
21. Kropman MF, Bakker HJ (2004) *J Am Chem Soc* 126:9135
22. Lightstone FC, Schwegler E, Hood RQ, Gygi F, Galli G (2001) *Chem Phys Lett* 343:549
23. Bakó I, Hutter J, Pálinkás G (2002) *J Chem Phys* 117:9838
24. Lightstone FC, Schwegler E, Allesch M, Gygi F, Galli G (2005) *Chem Phys Chem* 6:1
25. White JA, Schwegler E, Galli G (2000) *J Chem Phys* 113:4668
26. Ikeda T, Boero M, Terakura K (2007) *J Chem Phys* 126:034501
27. Jalilvand F, Spångberg D, Lindqvist-Reis P, Hermansson K, Persson I, Sandström M (2001) *J Am Chem Soc* 123:431
28. Varma S, Rempe SB (2006) *Biophys Chem* 124:192
29. McLain S, Imberti S, Soper AK, Botti A, Bruni F, Ricci MA (2006) *Phys Rev B* 74:094201
30. Krekeler C, Hess B, Delle Site L (2006) *J Chem Phys* 125:054305
31. Krekeler C, Delle Site L (2007) *J Phys Condens Matter* 19:192101
32. Naor MM, Van Nostrand K, Dellago C (2003) *Chem Phys Lett* 369:159
33. Lyubartsev AP, Laasonen K, Laaksonen A (2001) *J Chem Phys* 114:3120
34. Ramaniah LM, Bernasconi M, Parrinello M (1999) *J Chem Phys* 111:1587
35. Asthagiri D, Pratt LR, Paulaitis ME, Rempe SB (2004) *J Am Chem Soc* 126:1285
36. Silvestrelli PL, Parrinello M (1999) *Phys Rev Lett* 82:3308
37. Silvestrelli PL, Parrinello M (1999) *J Chem Phys* 111:3572
38. Marzari N, Vanderbilt D (1997) *Phys Rev B* 56:12847
39. Souza I, Marzari N, Vanderbilt D (1997) *Phys Rev B* 65:035109
40. Resta R (1998) *Phys Rev Lett* 80:1800
41. Delle Site L, Alavi A, Lynden-Bell RM (1999) *Mol Phys* 96:1683
42. Delle Site L (2001) *Mol Sim* 26:217
43. Delle Site L (2001) *Mol Sim* 26:353
44. Delle Site L, Lynden-Bell RM, Alavi A (2002) *J Mol Liq* 98:79
45. Krekeler C, Delle Site L (2008) *J Chem Phys* 128:134515
46. Scipioni R, Schmidt DA, Boero M (2009) *J Chem Phys* 130:024502
47. Lambeth BP, Junghans C, Kremer K, Clementi C, Delle Site L (2010) *J Chem Phys* 133:221101
48. Kerisit S, Rosso KM (2009) *J Chem Phys* 131:114512
49. Babiacyk WI, Bonella S, Guidoni L, Ciccotti G (2010) *J Phys Chem B* 114:15018
50. Kunz W, Lo Nostro P, Ninham BW (2004) The present state of affairs with Hofmeister series. *Curr Opin Colloid Interface Sci* 9:1–18
51. Zhang YJ, Cremer PS (2006) Interactions between macromolecules and ions. *Curr Opin Chem Biol* 10:658–663
52. Ball P (2008) Water as an active constituent in cell biology. *Chem Rev* 108:74–108
53. Kunz W (2010) Specific ion effects. World Scientific, Singapore
54. Zhang YJ, Cremer PS (2010) Chemistry of Hofmeister anions and osmolytes. *Annu Rev Phys Chem* 61:63–83
55. Harmandaris VA, Adhikari NP, van der Vegt NFA, Kremer K (2006) Hierarchical modeling of polystyrene: from atomistic to coarse grained simulations. *Macromolecules* 39:6708–6719
56. van der Vegt NFA, Peter C, Kremer K (2008) Structure based coarse- and fine-graining in soft matter simulations. Chapter 25 in: *Coarse graining of condensed phase and biomolecular systems*. Voth GA (ed), Chapman and Hall/CRC Press, Taylor and Francis Group
57. Schwierz N, Horinek D, Netz RR (2010) Reversed anionic Hofmeister series: the interplay of surface charge and surface polarity. *Langmuir* 26:7370–7379

58. Holm C, Joanny JF, Kremer K, Netz RR, Reineker P, Seidel C, Vilgis TA, Winkler RG (2004) Polyelectrolyte theory. *Adv Polym Sci* 166:67–111
59. Lund M, Jungwirth P, Woodward CE (2008) Ion specific protein assembly and hydrophobic surface forces. *Phys Rev Lett* 100:258105
60. Rasaiah JC, Friedman HL (1968) Integral equation methods in computation of equilibrium properties of ionic solutions. *J Chem Phys* 48:2742–2752
61. Diamond RM (1963) Aqueous solution behavior of large univalent ions - a new type of ion pairing. *J Phys Chem* 67:2513–2517
62. Hess B, van der Vegt NFA (2007) Solvent-averaged potentials for alkali-, earth alkali-, and alkylammonium halide aqueous solutions. *J Chem Phys* 127:234508
63. Collins KD (1997) Charge density-dependent strength of hydration and biological structure. *Biophys J* 72:65–76
64. Collins KD, Neilson GW, Enderby JE (2007) Ions in water: characterizing the forces that control chemical processes and biological structure. *Biophys Chem* 128:95–104
65. Fennell CJ, Bizjak A, Vlatchy V, Dill KA (2009) Ion pairing in molecular simulations of aqueous alkali halide solutions. *J Phys Chem B* 113:6782–6791
66. van Gunsteren WF, Bakowies D et al (2006) Biomolecular modeling: goals, problems, perspectives. *Angew Chem Int Ed* 45:4064–4092
67. Hess B, van der Vegt NFA (2006) Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. *J Phys Chem B* 110:17616–17626
68. Hess B, Holm C, van der Vegt NFA (2006) Osmotic coefficients of NaCl(aq) force fields. *J Chem Phys* 124:164509
69. Weerasinghe S, Smith PE (2003) A Kirkwood-Buff derived force field for sodium chloride in water. *J Chem Phys* 119:11342–11349
70. Hess B, van der Vegt NFA (2009) Cation specific binding with protein surface charges. *Proc Natl Acad Sci USA* 106:13296–13300
71. Fyta M, Kalcher I, Dzubiella J, Vrbka L, Netz RR (2010) Ionic force field optimization based on single-ion and ion-pair solvation properties. *J Chem Phys* 132:024911
72. Klasczyk B, Knecht V (2010) Kirkwood-Buff derived force field for alkali chlorides in simple point charge water. *J Chem Phys* 132:024109
73. Kalcher I, Dzubiella J (2009) Structure-thermodynamics relation of electrolyte solutions. *J Chem Phys* 130:134507
74. Pitzer KS, Mayorga G (1973) Thermodynamics of electrolytes. II. Activity and osmotic coefficients for strong electrolytes with one or both ions univalent. *J Phys Chem* 77:2300–2308
75. Hess B, Holm C, van der Vegt N (2006) Modeling multibody effects in ionic solutions with a concentration dependent dielectric constant. *Phys Rev Lett* 96:147801
76. Kirkwood JG, Buff FP (1951) The statistical mechanical theory of solutions I. *J Chem Phys* 19:774–777
77. Ben-Naim A (2006) Molecular theory of solutions. Oxford University Press, New York
78. Ganguly P, Schravendijk P, Hess B, van der Vegt NFA (2011) Ion pairing in aqueous electrolyte solutions with biologically relevant anions. *J. Phys. Chem. B* 115: 3734–3739
79. Voth GA (ed) (2009) Coarse-graining of condensed phase and biomolecular systems. CRC Press, Boca Raton
80. Mullinax JW, Noid WGJ (2009) *Chem Phys* 131:104110
81. Allen EC, Rutledge GJ (2009) *Chem Phys* 130:034904
82. Mognetti BM, Virnau P, Yelash L, Paul W, Binder K, Mueller M, Mac-Dowell LGJ (2009) *Chem Phys* 130:044101
83. Murtola T, Karttunen M, Vattulainen IJ (2009) *Chem Phys* 131:055101
84. Villa A, Peter C, van der Vegt NFA (2010) Transferability of nonbonded interaction potentials for coarse grained simulations: benzene in water. *J Chem Theory Comput* 6:2434–2444

85. Lyubartsev AP, Laaksonen A (1997) Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys Rev E* 55:5689–5696
86. Wei J, Li C, van der Vegt NFA, Peter C (2011) Transferability of coarse grained potentials: implicit solvent models for hydrated ions. *J Chem Theory Comput*, Doi:[10.1021/ct2001396](https://doi.org/10.1021/ct2001396)
87. Jungwirth P, Tobias DJ (2006) Specific ion effects at the air/water interface. *Chem Rev* 106:1259–1281
88. Smith PE (1999) Computer simulation of cosolvent effects on hydrophobic hydration. *J Phys Chem B* 103:525–534
89. Kalra A, Tugcu N, Cramer SM, Garde S (2001) Salting-in and salting-out of hydrophobic solutes in aqueous salt solutions. *J Phys Chem B* 105:6380–6386
90. van der Vegt NFA, van Gunsteren WF (2004) Entropic contributions in cosolvent binding to hydrophobic solutes in water. *J Phys Chem B* 108:1056–1064
91. van der Vegt NFA, Trzesniak D, Kasumaj B, van Gunsteren WF (2004) Energy-entropy compensation in the transfer of nonpolar solutes from water to cosolvent/water mixtures. *ChemPhysChem* 5:144–1477
92. Godawat R, Jamadagni SN, Garde S (2009) Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proc Natl Acad Sci USA* 106:15119–15124
93. Jamadagni SN, Godawat R, Garde S (2009) How surface wettability affects the binding, folding, and dynamics of hydrophobic polymers at interfaces. *Langmuir* 25:13092–13099
94. Villa A, Peter C, van der Vegt NFA (2009) Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Phys Chem Chem Phys* 11:2077–2086
95. Lee ME, van der Vegt NFA (2006) Does urea denature hydrophobic interactions? *J Am Chem Soc* 128:4948–4949
96. Wu JZ, Prausnitz JM (2008) Pairwise hydrophobic effect for alkanes in water. *Proc Natl Acad Sci USA* 105:9512–9515
97. Harmandaris VA, Reith D, van der Vegt NFA, Kremer K (2007) Comparison between coarse-graining models for polymer systems: two mapping schemes for polystyrene. *Macromol Chem Phys* 208:2109–2120
98. Fritz D, Harmandaris VA, Kremer K, van der Vegt NFA (2009) Coarse-grained polymer melts based on isolated atomistic chains: simulation of polystyrene of different tacticities. *Macromolecules* 42:7579–7588
99. Fritz D, Herbers CR, Kremer K, van der Vegt NFA (2009) Hierarchical modeling of polymer permeation. *Soft Matter* 5:4556–4563
100. Leon S, van der Vegt N, Delle Site L, Kremer K (2005) Bisphenol-A-polycarbonate: entanglement analysis from coarse-grained MD simulations. *Macromolecules* 38:8078–8092
101. Hess B, Leon S, van der Vegt NFA, Kremer K (2006) Long time polymer trajectories from coarse grained simulations: bisphenol-A-polycarbonate. *Soft Matter* 2:409–414
102. Harmandaris VA, Kremer K (2009) Dynamics of polystyrene melts through hierarchical multiscale simulations. *Macromolecules* 42:791–802
103. Marcon V, Fritz D, van der Vegt NFA (2011) Hierarchical modelling of polystyrene surfaces (in preparation)
104. Reith D, Putz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic simulations. *J Comput Chem* 24:1624–1636
105. Rühle V, Junghans C, Lukyanov A, Kremer K, Andrienko D (2009) Versatile object-oriented toolkit for coarse-graining applications. *J Chem Theory Comput* 5:3211–3223
106. Fritz D (2009) Coarse-graining methods and polymer dynamics. PhD Thesis, University of Mainz
107. Carbone P, Karimi-Varzaneh HA, Chem X, Müller-Plathe F (2008) Transferability of coarse-grained force fields: the polymer case. *J Chem Phys* 128:064904
108. Villa A, van der Vegt NFA, Peter C (2009) Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Phys Chem Chem Phys* 11:2068–2076

109. Yesylevskyy SO, Schäfer LV, Sengupta D, Marrink SJ (2010) Polarizable water model for the coarse-Grained MARTINI force field. *PLoS Comput Biol* 6:e1000810
110. Wu Z, Cui Q, Yethiraj A (2010) Coarse-grained model for water: the importance of electrostatic interactions. *J Phys Chem B* 114:10524–10529
111. Woodward CE, Jönsson B (1991) *Chem Phys* 155:207–219
112. Stevens MJ, Kremer K (1993) *Macromolecules* 26:4717
113. Stevens MJ, Kremer K (1993) *Phys Rev Lett* 71:2228
114. Stevens MJ, Kremer K (1995) *J Chem Phys* 103(4):1669–1690
115. Kremer K, Grest GS (1990) *J Chem Phys* 92:5057–5086
116. Limbach HJ (2001) Ph.D. thesis. Johannes Gutenberg–Universität Mainz. <http://archimed.uni-mainz.de/pub/2002/0121/>
117. Dünweg B, Stevens MJ, Kremer K (1995) In: Binder K (ed) *Monte Carlo and molecular dynamics simulations in polymer science*. Oxford University Press, New York
118. Holm C, Rehahn M, Oppermann W, Ballauff M (2004) *Adv Polym Sci* 166:1–27
119. Dobrynin AV (2008) *Curr Opin Colloid Interface Sci* 13:376–388
120. Micka U, Holm C, Kremer K (1999) *Langmuir* 15:4033
121. Limbach HJ, Holm C, Kremer K (2002) *Europhys Lett* 60:566–572
122. Mann BAF, Everaers R, Holm C, Kremer K (2004) *Europhys Lett* 67:786–792
123. Mann BAF, Holm C, Kremer K (2005) *J Chem Phys* 122:154903
124. Dobrynin AV, Rubinstein M, Obukhov SP (1996) *Macromolecules* 29:2974
125. Limbach HJ, Holm C (2002) *Comput Phys Commun* 147:321–324
126. Limbach HJ, Holm C (2003) *J Phys Chem B* 107:8041–8055
127. Liao Q, Dobrynin AV, Rubinstein M (2003) *Macromolecules* 36:3399–3410
128. Liao Q, Dobrynin AV, Rubinstein M (2003) *Macromolecules* 36:3386–3398
129. Liao Q, Dobrynin AV, Rubinstein M (2006) *Macromolecules* 39:1920–1938
130. Deserno M, Holm C (2001) Cell-model and Poisson-Boltzmann-theory: a brief introduction. In: Holm C, Kékicheff P, Podgornik R (eds) *Electrostatic effects in soft matter and biophysics*, volume 46 of NATO science series II – mathematics, physics and chemistry. Kluwer, Dordrecht
131. Blaul J, Wittemann M, Ballauff M, Rehahn M (2000) *J Phys Chem B* 104:7077–7081
132. Deserno M, Holm C, Blaul J, Ballauff M, Rehahn M (2001) *Eur Phys J E* 5:97–103
133. Antypov D, Holm C (2006) *Phys Rev Lett* 96:088302
134. Antypov D, Holm C (2007) *Macromolecules* 40:731–738
135. Arnold A, Holm C (2005) *J Chem Phys* 123:144103
136. Deserno M, Holm C (2002) *Mol Phys* 100:2941–2956
137. Dünweg B, Ladd AJC (2009) *Adv Poly Sci* 221:89–166
138. Gompper G, Ihle T, Kroll DM, Winkler RG (2009) *Adv Poly Sci* 221:1–87
139. Slater GW, Holm C, Chubynsky MV, de Haan HW, Dubé A, Grass K, Hickey OA, King-sburry C, Sean D, Shendruk TN, Zhan L (2009) *Electrophoresis* 30:792–818
140. Grass K, Böhme U, Scheler U, Cottet H, Holm C (2008) *Phys Rev Lett* 100:096104
141. Frank S, Winkler RG (2008) *Europhys Lett* 83:38004
142. Grass K, Holm C (2010) *Faraday Discuss* 144:5770
143. Smiatek J, Schmid F (2010) *J Phys Chem B* 114:6266
144. Netz RR (2003) *Phys Rev Lett* 90:128104
145. Hsiao P-Y, Luijten E (2006) *Phys Rev Lett* 97:148301
146. Liu H, Zhu Y, Maginn E (2010) *Macromolecules* 43:48054813
147. Schmitt J, Decher G, Hong G (1992) *Thin Solid Films* 210/211:831
148. Decher G (1997) *Science* 277:1232–1237
149. Netz RR, Joanny JF (1999) *Macromolecules* 32:9013
150. Park SY, Rubner MF, Mayes AM (2002) *Langmuir* 18:9600–9604
151. Castelnovo M, Joanny JF (2000) *Langmuir* 16:7524–7532
152. Shafir A, Andelman D (2006) *Eur Phys J E* 19:155–162
153. Wang Q (2006) *J Phys Chem B* 110:5825–5828

154. Hoda N, Larson RG (2009) *J Phys Chem B* 113:4232–4241
155. Wang Q (2009) *Soft Matter* 5:413–424
156. Zhao W, Zheng B, Haynie DT (2006) *Langmuir* 22:6668–6675
157. Reddy G, Chang R, Yethiraj A (2006) *J Chem Theory Comput* 2:630–636
158. Chialvo AA, Simonson JM (2008) *J Phys Chem C* 112:19521–19529
159. Horinek D, Serr A, Geisler M, Pirzer T, Slotta U, Lud SQ, Garrido JA, Scheibel T, Hugel T, Netz RR (2008) *Proc Natl Acad Sci* 105:2842–2847
160. Qiao B, Cerdá JJ, Holm C (2010) *Macromolecules* 43:7828–7838
161. Qiao B, Cerdá JJ, Holm C (2011) *Macromolecules* 44:1707–1718
162. Messina R, Holm C, Kremer K (2003) *Langmuir* 19:4473–4482
163. Messina R (2003) *J Chem Phys* 119:8133–8139
164. Messina R (2004) *Macromolecules* 37:621–629
165. Messina R, Holm C, Kremer K (2004) *J Polym Sci B Polym Phys* 42:3557
166. Panchagnula V, Jeon J, Dobrynin AV (2004) *Phys Rev Lett* 93:037801
167. Messina R (2004) *Phys Rev E* 70:051802
168. Patel PA, Jeon J, Mather PT, Dobrynin AV (2005) *Langmuir* 21:6113–6122
169. Panchagnula V, Jeon J, Rusling JF, Dobrynin AV (2005) *Langmuir* 21:1118–1125
170. Jeon J, Panchagnula V, Pan J, Dobrynin AV (2006) *Langmuir* 22:4629–4637
171. Guyomard A, Muller G, Glinel K (2005) *Macromolecules* 38:5737–5742
172. Kovacevic D, van der Burgh S, de Keizer A, Stuart MAC (2002) *Langmuir* 18:5607–5612
173. Gopinadhan M, Ahrens H, Gunther JU, Steitz R, Helm CA (2005) *Macromolecules* 38:5228–5235
174. Liu G, Zhao J, Sun Q, Zhang G (2008) *J Phys Chem B* 112:3333–3338
175. Riegler H, Essler F (2002) *Langmuir* 18:6694–6698
176. Tanchak OM, Yager KG, Fritzsche H, Harroun T, Katsaras J, Barrett CJ (2008) *J Chem Phys* 129:084901
177. Kolasinska M, Warszynski P (2005) *Bioelectrochemistry* 66:65–70
178. Buron CC, Filiatre C, Membrey F, Bainier C, Charraut D, Foissy A (2007) *Colloids Surf A* 305:105–111
179. Steitz R, Jaeger W, von Klitzing R (2001) *Langmuir* 17:4471–4474
180. Schoeler B, Kumaraswamy G, Caruso F (2002) *Macromolecules* 35:889–897
181. Glinel K, Moussa A, Jonas AM, Laschewsky A (2002) *Langmuir* 18:1408–1412
182. Salomäki M, Kankare J (2008) *Macromolecules* 41:4423–4428
183. El Haitami AE, Martel D, Ball V, Nguyen HC, Gonthier E, Labbe P, Voegel JC, Schaaf P, Senger B, Boulmedais F (2009) *Langmuir* 25:2282–2289
184. Muller M, Meier-Haack J, Schwarz S, Buchhammer HM, Eichhorn EJ, Janke A, Kessler B, Nagel J, Oelmann M, Reihs T, Lunkwitz K (2004) *J Adhes* 80:521–547
185. Gopinadhan M, Ivanova O, Ahrens H, Günther JU, Steitz R, Helm CA (2007) *J Phys Chem B* 111:8426–8434
186. Cerdá JJ, Qiao B, Holm C (2009) *Soft Matter* 5:4412–4425
187. Cerdá JJ, Qiao B, Holm C (2009) *Eur Phys J Spec Top* 177:129–148
188. Carrillo JMY, Dobrynin AV (2007) *Langmuir* 23:2472–2482
189. Patel PA, Jeon J, Mather PT, Dobrynin AV (2006) *Langmuir* 22:9994–10002
190. Kulcsar A, Voegel JC, Schaaf P, Kekicheff P (2005) *Langmuir* 21:1166–1170
191. Gavryushov S, Linse P (2006) *J Phys Chem B* 110:10878–10887
192. Neff PA, Wunderlich BK, von Klitzing R, Bausch AR (2007) *Langmuir* 23:4048–4052
193. Schönhoff M, Ball V, Bausch AR, Dejgunat C, Delorme N, Glinel K, von Klitzing R, Steitz R (2007) *Colloids Surf A* 303:14–29
194. Wende C, Schönhoff M (2010) *Langmuir* 26:8352–8357
195. Kusalik PG, Mandy ME, Svishchev IM (1994) *J Chem Phys* 100:7654–7664
196. Mahoney MW, Jorgensen WL (2000) *J Chem Phys* 112:8910–8922
197. Tyagi S, Arnold A, Holm C (2007) *J Chem Phys* 127:154723
198. Tyagi S, Arnold A, Holm C (2008) *J Chem Phys* 129:204102

199. Tyagi S, Süzen M, Sega M, Barbosa M, Kantorovich SS, Holm C (2010) *J Chem Phys* 132:154112
200. Sagui C, Darden T (2001) *J Chem Phys* 114:6578
201. Banerjee S, Board JA Jr (2005) *J Comput Chem* 26:957–967
202. Maggs AC (2002) *J Chem Phys* 117:1975
203. Pasichnyk I, Dünweg B (2004) *J Phys Condens Matter* 16:3999–4020
204. Rottler J (2007) *J Chem Phys* 127:134104–134109
205. Meller LNA, Branton D (2001) *Phys Rev Lett* 86:3435
206. Keyser UF, Koeleman BN, van Dorp S, Krapf D, Smeets RMM, Lemay SG, Dekker NH, Dekker C (2006) *Nat Phys* 2:473–477
207. Howorka S, Siwy Z (2009) *Chem Soc Rev* 38:2360–2384
208. van Dorp S, Keyser UF, Dekker NH, Dekker C, Lemay SG (2009) *Nat Phys* 5:347–351
209. Muthukumar M (1999) *J Chem Phys* 111:10371
210. Milchev A, Binder K, Bhattacharya A (2004) *J Chem Phys* 121:6042–6051
211. Gracheva ME, Xiong A, Aksimentiev A, Schulten K, Timp G, Leburton JP (2006) *Nanotechnology* 17:622–633
212. Forrey C, Muthukumar M (2007) *J Chem Phys* 127:015102
213. Luan B, Aksimentiev A (2008) *Phys Rev E* 78:21912
214. Melchionna S, Bernaschi M, Fyta M, Kaxiras E, Succi S (2009) *Phys Rev E* 79:30901
215. Kesselheim S, Sega M, Holm C (2011) *Comput Phys Commun* 182:33–35
216. Fixman M, Skolnick J (1978) *Macromolecules* 11:863
217. Poma AB, Delle Site L (2010) *Phys Rev Lett* 104:250201

Coarse-Grained Modeling for Macromolecular Chemistry

Hossein Ali Karimi-Varzaneh and Florian Müller-Plathe

Abstract The physical phenomena and properties of macromolecules such as polymers or biological materials cover a wide range of length and time scales: from Ångströms and subpicoseconds to millimeters and minutes. Multiscale simulation methods link different computer simulation approaches, which cover these scales and the respective levels of resolution. Different simulation methods that bridge the atomistic description of the system to a coarser level have been developed in order to reach the mesoscopic time and length scales important for many material properties. Here, we give a short introduction to multiscale simulation approaches in macromolecular chemistry. Then, we review the coarse-grained simulation models developed to drive a simple model from a more detailed one. Some different methodological aspects such as time scale and dynamics in coarse-grained simulations and several typical problems are briefly addressed, finishing with a look at future challenges.

Keywords Atomistic simulation · Coarse-graining simulation · Computational chemistry · Multiscale modeling

Contents

1	Introduction	296
2	Mapping Scheme	297
3	Different Methods for Developing the Coarse-Grained Models	303
3.1	Structure-Based CG Model	304
3.2	Dynamic-Based CG Model	314
3.3	Coarse-Graining in Time	315
4	Back-Mapping	316
5	Outlook	316
	References	318

1 Introduction

Polymeric systems are characterized by a wide range of length scales that extend from Ångströms for the distance between the bonded atoms to at least micrometers for the contour length of the chain. The corresponding time scales associated with motions on such length scales are even broader: bond vibrations occur on the scale of picoseconds (10^{-13} s) and chain relaxation and morphology formation can occur over seconds, minutes, or hours, depending on molecular weight, temperature, and density. For this reason, an equally wide range of simulation methods at different levels of resolution and, consequently, including differing degrees of freedom is employed to study them [1].

Quantum mechanical (QM) methods present the most detailed picture of the system by using different levels of approximations to solve the Schrödinger equation and evaluate electron wave functions. A work by Martonak et al. [2] shows that even at room temperature, quantum effects are crucial to understanding the anisotropic thermal expansion of polyethylene crystals. It is clear that if the number of details that a simulation technique describes is increased, then the accessibility of long time scales and large length scales is decreased. Therefore, the QM methods can be used only for very short time and length scales, which are typically of the order of Ångströms and picoseconds, respectively. However, QM methods are extremely valuable in providing important information for preparation of an atomistic model about the basic structure of molecules, i.e., bond lengths, bond angles, torsion and associated force constant, partial charges, and torsional barriers.

In classical molecular dynamics (MD) simulations, the charge distributions are approximated either by putting fixed partial charges on interaction sites or by adding an approximate model for polarization effects. Thus, in MD simulations the time scale of the system is not dominated by the motion of the electrons, but mainly by the time of rotational motions and intramolecular vibrations, which are orders of magnitude slower than those of electron motions. Consequently, the time step of integration is larger, and trajectory lengths are in the order of nanoseconds and accessible lengths in the order of 10–100 Å. In MD methods, the Hamilton equations of motion [3] are integrated to move particles to new positions and to assign new velocities at these new positions, which is, in principle, probing the whole phase space. Consequently, MD simulation is a powerful technique for computing the equilibrium and dynamic properties of classical many-body systems [4]. Over the last 20 years, with rapid development of computers, polymeric systems have been the subject of intense study using MD simulations, but MD simulations using atomistic force fields are still unable to access the time scales necessary to achieve chain relaxation for polymeric systems of intermediate or high molecular weights [5]. To equilibrate the dense polymeric systems with long chains, advanced Monte Carlo (MC) methods have been developed that probe the configuration space by trial moves of particles. Within the so-called Metropolis algorithm, the energy change between two consecutive steps is used to accept or reject the new configuration to find the system in its energy-minimum state [4, 6–9].

Nevertheless, the size of the systems that can be efficiently simulated is still limited by the performance of present-day computers.

In order to study polymeric systems, particularly their ability to self-assemble over tens or hundreds of nanometers, it is necessary to reduce the number of degrees of freedom. Simple generic models (such as bead-and-spring or lattice models) are found to be suitable for study of scaling properties of macromolecular systems [10–12], both for static and dynamic properties, by reducing the computational complexity to the absolute minimum, namely connectivity and excluded volume plus some specific interactions, if needed [5, 6, 10]. Thus they allow for much longer effective time and length scales than more detailed models. The bead-and-spring models [10] are the most elementary MD models that are still able to resolve polymer chains. In these models, the individual polymer chain is modeled by mass points that repel each other via most often a purely repulsive Lennard–Jones (Weeks Chandler Anderson) potential to produce the excluded volume interaction, and they are connected along the chain by, for example, a finite extensible nonlinear elastic potential (FENE) to take care of the connectivity. These models have been used frequently to study the dynamics of short- and long-chain polymer melts as well as the relaxation properties of crosslinked polymer melts [13, 14]. However, for a complete understanding of polymer properties (such as viscosity or crystallization), both generic chain length and connectivity information (namely, mesoscopic model) and detailed material-specific information about the chemical composition (namely, atomistic model) are necessary, and one needs to combine the models. One approach to attain this goal is the development of “systematically coarse-grained” models, which can be used to predict the quantities for a specific material. The coarse-grained (CG) models allow a simplified picture and interpretation of complex molecular phenomena by averaging over unessential atomic details. However, the models still retain some of the material-specificity of the parent atomistic models. They are obtained by lumping groups of chemically connected atoms into “superatoms” and deriving the effective CG interaction potentials from the microscopic details of the atomistic models.

2 Mapping Scheme

The position of the beads in the CG model can be related to the atomistic coordinates of a structure by using a mapping scheme. It is clear that there is no unique way to map a given set of atoms onto a CG description. Depending on the specific system and on the properties that one wants to see reflected at the coarse level, one can define criteria to determine the mapping scheme. Examples for such criteria are the ability of the mapping scheme to account for stereoregularity of chain molecules or to capture certain geometry changes. Figure 1 shows different mapping schemes that have been used recently by our group to develop CG models for different polymers {polyamide-66 (PA-66) [15, 16], poly(methyl methacrylate) (PMMA) (Karimi-Varzaneh et al., unpublished results), and polystyrene (PS)

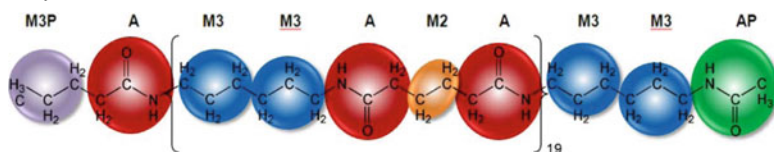
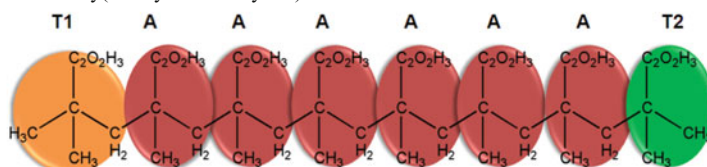
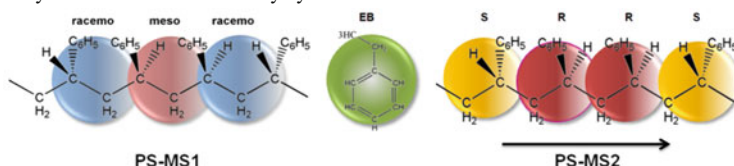
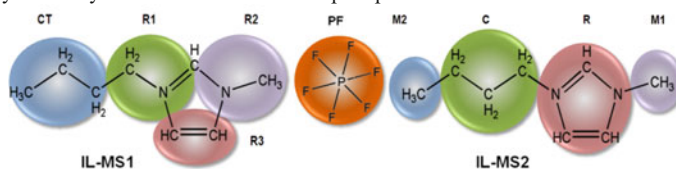
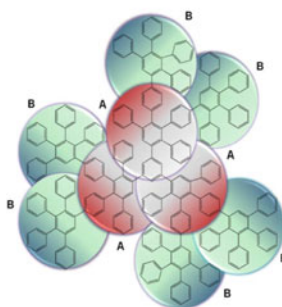
a Polyamide-66**b** Isotactic Poly(methyl methacrylate)**c** Ethylbenzene and atactic Polystyrene**d** 1-alkyl-3-methylimidazolium Hexafluorophosphate**e** Polyphenylene Dendrimer

Fig. 1 Chemical structure and mapping schemes for some of the discussed CG examples. (a) PA-66 [15, 16]: The centers of the beads *A*, *M2*, *M3* are located, respectively, in the carbonyl carbon, in the center of mass of the ethylene unit, and in the central carbon atom. The end groups are treated separately because their local dynamics differ from the internal monomers, and to conserve the total mass of the system. *AP* is the end monomer containing the amide group, and *M3P* is the end monomer containing the aliphatic moiety. (b) Isotactic PMMA (Karimi-Varzaneh et al., unpublished results): Every monomer is represented by one bead placed on the center of mass of the monomer (*A*), and the end groups (*T1* and *T2*) are treated differently to internal beads. (c) EB and atactic PS: *PS-MS1* is mapping scheme 1 for PS [17]; *meso* indicates that in this study two

[15, 17, 18]], for ionic liquids (1-alkyl-3-methylimidazolium hexafluorophosphate [C₄mim][PF₆] [19]), and for dendrimers (polyphenylene dendrimer [20]). For example, in the case of PS chains (Fig. 1c), two different mapping schemes have been proposed (PS-MS1 and PS-MS2). In PS-MS1, when two consecutive chiral carbon atoms have the same absolute configurations, the bead is called “meso” and if the configurations are different the bead is labeled “racemo”. The superatoms are located on the methylene carbon between the two stereocentres. The meso and racemo superatoms are treated as two different monomer species. In this way, the model is able to account for the tacticity of the polymer chain, and it accurately describes static properties of melts in a broad range of molecular weights [17]. In PS-MS2, bead types (*R* and *S*) are defined according to the absolute configuration of the monomer given by the asymmetric –CHR group defined against a given direction of the carbon backbone, and they are placed at the centers of mass of chemical repeat units. As will be discussed in Sect. 3, the CG potential in this model is temperature-independent [18]. Two different mapping schemes are also proposed by Hermendaris et al. [21] that use two beads per monomer in order to model the tacticity of PS. In one case, the CH₂ group of the backbone chain represents one CG effective bead, whereas the remaining CH group of the monomer in the backbone and the phenyl ring are mapped to another effective CG bead. In the other case, a CG bead includes information from three consecutive CH_x groups along the backbone (the CH₂ of a PS monomer plus the half mass of each one of the two neighboring CH groups along the chain backbone), while the other CG bead is just the phenyl ring. Their results showed that relatively small changes in how to divide the monomer into two beads influence the performance of the models and the reproduction of the local chain conformations and melt packing observed in atomistic simulations of the polymer.

To see how the different mapping schemes proposed in the literatures for PS were able to reproduce the atomistic structure, we plotted the intermolecular radial distribution functions (RDFs) between the centers of mass of PS monomers, ethylbenzene (EB) molecules, or both, obtained from the CG model proposed by

Fig. 1 (continued) consecutive chiral carbon atoms have the same absolute configurations and *racemo* indicates that the configurations are different. *PS-MS2* is mapping scheme 2 for PS [18]. The CG beads represent each EB molecule or each styrene unit, and the tacticity of the PS chains are captured in the CG model by defining different bond/angle types between *S* (phenyl ring located behind) and *R* (phenyl ring in front) types of beads. (d) [C₄mim][PF₆]: *IL-MS1* is mapping scheme 1 and *IL-MS2* is mapping scheme 2 [19]. In the alkyl tail of *IL-MS1*, one bead is placed every three carbon atoms (*CT*) while the imidazolium ring is split into three beads (*R1*, *R2*, and *R3*). Note that one carbon and one hydrogen atom are shared equally among the *R1* and *R2* beads. In *IL-MS2*, *M1* indicates the methyl group directly bonded to the imidazolium ring and *M2* the methyl group at the end of the aliphatic chain. Bead *R* indicates the imidazolium ring. In the aliphatic chain, three methylene units are collected into one single bead *C*. In both models, the anion (PF₆[−]) is described by one bead (*PF*). (e) First and second generation of polyphenylene dendrimer [20]: One type of bead (*A*), which contains the entire repeat unit, placed at the center of mass, is used to build the model for the first generation of dendrimers. For the second generation, bead types *A* (the inner beads) and *B* (the outer beads) are used

Hermendaris et al. [22] and Qian et al. [18] (for PS-MS2) at a temperature of 463 K (see Fig. 2). The RDFs are compared with the united-atom model of [22]. The figure shows that the mapping scheme of Qian et al. [18] reproduces the atomistic

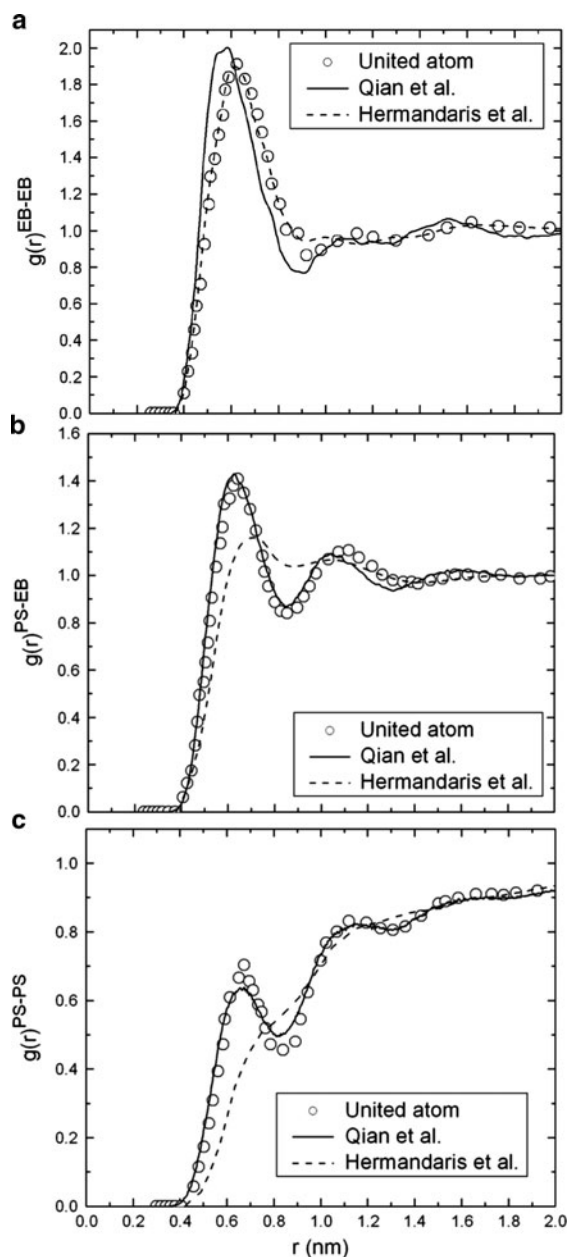


Fig. 2 Intermolecular RDFs between centers of mass of EB molecules (a), PS and EB molecules (b), or PS monomers (c), obtained from the CG models proposed by Hermendaris et al. [22] and Qian et al. [18] (for PS-MS2) compared with the united-atom model of [22] for a mixture of PS (96-mers) and EB (10 mol%) at 463 K

structure much closer than the mapping scheme of Hermendaris et al. [22] for different RDFs. The difference between RDFs of the two mapping schemes could also be due to the different ways of developing the CG force fields, which we will discuss in Sect. 3. However, it is clear that the different mappings for PS implicitly provide a bias of the resulting model towards the description of certain properties: if the superatom location is on the main chain, at the center of mass of the repeat unit, or at the center of the phenyl ring it will lead to a better reproduction of the main-chain structure or the phenyl ring arrangements, for example.

Two mapping schemes have also been used for $[C_4mim][PF_6]$ in Fig. 1d (IL-MS1 and IL-MS2). In IL-MS1, every three carbon atoms in the alkyl tail is represented by one bead (bead CT), whereas the imidazolium ring is split into three beads (R1, R2, and R3). This model has been proposed for study of the orientation of the imidazolium ring in the CG model. In IL-MS2, three different superatoms are used to describe the methyl moieties (identified by labels M1 and M2) and the imidazolium ring (bead R). In the aliphatic chain, three methylene units are collected into one single site called C. We have found that the two schemes, which retain different features of the original molecule, perform differently depending on the properties. Both mapping schemes are able to reproduce with sufficient accuracy the experimental X-ray and neutron scattering results (Fig. 3). However, the difference between the models becomes more apparent in the details of the position of the two major peaks ($\sim 1.5\text{\AA}^{-1}$ and $\sim 1\text{\AA}^{-1}$) and of the specific interactions that contribute to them. The inset in Fig. 3 compares the position of the two peaks as obtained from the CG simulations and as measured experimentally. The figure shows that IL-MS1 better fits the data corresponding to the lower q ,

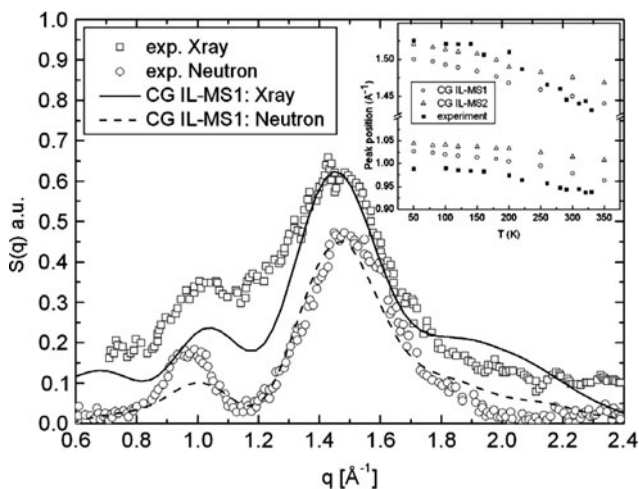


Fig. 3 Neutron and X-ray scattering structure factors of $[C_4mim][PF_6]$ obtained from CG simulation (IL-MS1) and experiment at 200 K. The *inset* shows the positions of the two major peaks of the X-ray scattering structure factors of $[C_4mim][PF_6]$ for the two mapping schemes (IL-MS1 and IL-MS2) at different temperatures. Neutron and X-ray scattering data are from [23]

whereas IL-MS2 agrees better with the position of the peak at higher q . This can be understood by looking at the different details retained in the two mapping schemes: IL-MS1, which describes the aromatic ring with more details using three different beads and preserving its flat shape, predicts a better position of the lower q peak, which in fact is due to the interactions between the rings [24–26]. In contrast, IL-MS2, which describes the alkyl tail more accurately with two beads instead of one, predicts better the position of the peak at $1.5\text{\AA}^{-1}$, which arises from the tail–tail interactions [24–26].

In addition to the positions of the superatoms, their number is also important. The number of real atoms collected into one superatom is often referred to as the degree of coarse-graining. It cannot be chosen completely arbitrarily, or else the essential physics of the polymer may be lost, as the following example shows. If polyethylene were to be coarse-grained, one could combine 1, 2, 3, or n CH_2 units into one superatom. For efficiency, one should opt for a high degree of coarse-graining n . The volume of the spherical superatom has to be approximately the sum of the volumes of the constituting methylene units, which fixes its diameter. For large enough n , this will lead to neighboring superatoms whose excluded volumes no longer overlap. As a consequence, polymer chains can cut through each other in a simulation and are not forced to reptate through the tubes formed by their neighbors. It is obvious that the dynamics of such a polymer melt will be qualitatively wrong, so that no diffusion coefficients, viscosities, elastic, or rheological parameters can be obtained. In the case of polyethylene, the largest possible n is 3. If the polymer has bulkier, more spherical subunits then n can be considerably larger. An example is the above-mentioned PS model where a styrene repeat unit (16 atoms) can be combined into one superatom without any harm to the dynamics because PS has a much fatter envelope than polyethylene.

The choice of the number and position of the superatoms is the responsibility of the human researcher. Guidelines can be given, but it cannot be automated. One possible principle for selecting the superatom center is that the bonds between superatoms can be represented by a single harmonic potential. One can calculate the bond distributions between the beads for different possible positions of the center of the superatom. Mapping schemes with a more localized bond length distribution are often preferred for technical reasons. For example, in the case of PA-66 (Fig. 1a), the main concern is on the M3 bead and we can consider two possibilities for the center of the bead: the center of mass (CM) of the bead or the central carbon atom. Figure 4 compares the distributions for the two different mapping schemes. As shown in this figure, the mapping with beads centered on the CM of M3 gives a three-peaked bond length distribution, whereas a more localized, double peak is found in the case of the beads centered on the second carbon atom of M3. In the light of this result, we locate the M3 center on the carbon atom of the backbone (rather than in the CM of the bead).

Another criterion that needs to be taken into account relates to the statistical correlations of internal degrees of freedom. The mapping scheme should be chosen such that these correlations are as weak as possible so that the intramolecular potentials can be separated, to a good approximation, into bond stretching, bond

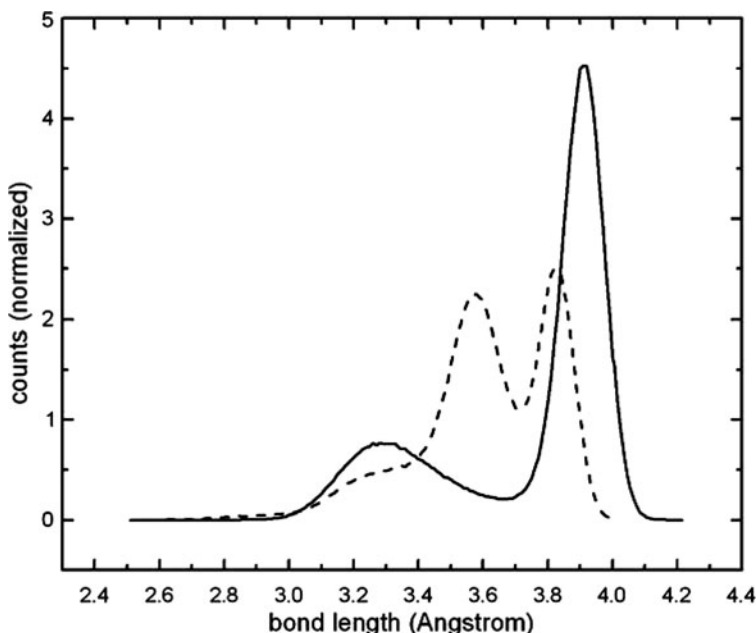


Fig. 4 Histogram of M3–M3 bond length (see Fig. 1a) extracted from atomistic simulation at 300 K for center of bead located at the center of mass (*dashed line*) and for the second carbon as the center of M3 (*solid line*)

angle bending, and torsions, as outlined in Sect. 3.1. A direct way to check the correlation is by plotting contour plots of combinations of energies (or probability distributions) of different bonded potentials in the CG description obtained from the atomistic simulation [21]. Once the position of the mapping centers is chosen, we have to define the type and characteristics of the effective interactions between such centers.

3 Different Methods for Developing the Coarse-Grained Models

The development of system-specific CG models for polymers is a very active research field. Coarse-graining means that about 90% of the degrees of freedom of an atomistic model are discarded. As a consequence, a CG model cannot be expected to reproduce all results of the parent atomistic model. Before developing a CG model one must, therefore, decide which polymer properties it shall reproduce. One can determine a set of target properties of microscopic character (such as forces or instantaneous energies) or structural averages (such as RDFs, average energies, and pressure), which the CG model is required to keep. Depending on the target properties, the way the effective CG potentials are derived, and also on

the degree of coarse-graining, various models and methods have been proposed in the literature [15, 17, 27–44].

3.1 *Structure-Based CG Model*

One way to develop a CG model is using a structure-based coarse-graining approach, where the direct link to the chemistry is achieved through structurally defined bonded and non-bonded effective CG potentials derived from the atomistic model. In this class of methods, the determination of interaction potentials for the CG model is based on the assumption that the total potential energy can be separated into bonded and non-bonded contributions. The bonded interactions are derived such that the conformational statistics of a single molecule is represented correctly in the CG model. A very important criterion for a mapping scheme, as mentioned in Sect. 2, is its ability to decouple internal degrees of freedom so that the intramolecular (bonded) potentials can be separated into bond, angle, and torsion terms. One option for deriving the CG bonded potentials is to use a simple Boltzmann inversion to convert the distributions of interparticle distances or angles into potentials. Another option is to determine analytical potentials that reproduce the probability distributions for the bonded part, for example by fitting the (multi peaked) bonded distributions by a series of Gaussian functions that can then be inverted analytically, resulting in smooth potentials and forces [45].

Similar to the bonded interaction functions, there are two options for deriving the non-bonded potentials: either using analytical potentials or using numerically derived tabulated potentials. In the first case, analytical potentials of various types can be used. The “normal” Lennard–Jones 12-6 potential is frequently used, which is sometimes too repulsive for the CG soft beads [44, 46], and for softer cases the Lennard–Jones-type (e.g., 9-6 or 7-6) [21, 47], Buckingham, or Morse potentials [48] have been employed. The potential parameters are chosen in such a way that the CG model reproduces satisfactorily the physical properties of the atomistic simulation or available experimental data. This task can be done automatically by a computer in a more efficient way than the usual manual trial and error method by using, e.g., the simplex algorithm [49, 50]. In this algorithm, a penalty function that compares the calculated values of selected properties with their target values from atomistic simulations or experiment is minimized by adjusting the force field parameters. In a parameter space of dimension N , $N+1$ preliminary molecular dynamics simulations with slightly different starting parameter sets are performed. Then, the physical properties of interest and the penalty function for each parameter combination are calculated. If for one of these sets the value of the penalty function is below a certain user-defined threshold, the corresponding force field is supposed to be satisfactory. Otherwise, a new molecular dynamics simulation is run with a new parameter set provided by a simplex move, and this process is repeated until the penalty function converges. However, slow convergence of the

analytical potentials and the manual process of selecting a good functional form of the potential are disadvantages of the first case of deriving the non-bonded CG potentials.

Concerning the second option to generate numerically a tabulated potential that closely reproduces a given melt structure, the iterative Boltzmann inversion (IBI) method [29, 41, 51, 52] has been developed.

3.1.1 Iterative Boltzmann Inversion Method

The main feature of the IBI method is the automatic and iterative way of determining the effective bead–bead interactions that match a set of structural quantities (such as intermolecular RDFs) calculated from a more detailed reference simulation model (i.e., atomistic). Henderson [53] proved that at a given density and temperature, there is a unique mapping between the RDF and the intermolecular potential. Thus, a potential that reproduces the target RDF is a fixed point of the iteration and, if the algorithm converges, a valid solution is obtained for the CG potential. For a complete polymer model, one assumes that the total potential energy U^{CG} can be separated into bonded (covalent) and non-bonded contributions:

$$U^{\text{CG}} = \sum U_{\text{b}}^{\text{CG}} + \sum U_{\text{nb}}^{\text{CG}}, \quad (1)$$

where U_{b}^{CG} and $U_{\text{nb}}^{\text{CG}}$ represent the bonded and non-bonded part of the potential, respectively. The bonded interactions are derived such that the conformational distribution P^{CG} , which is characterized by specific CG bond lengths r between adjacent pairs of CG beads, angles θ between neighboring triplets of beads, and torsions φ between neighboring quadruplet of beads, i.e., $P^{\text{CG}}(r, \theta, \varphi)$, in the CG simulation is reproduced. If one assumes that the different internal CG degrees of freedom are uncorrelated, then $P^{\text{CG}}(r, \theta, \varphi)$ factorizes into independent probability distributions of bond, angle, and torsional degrees of freedom:

$$P^{\text{CG}}(r, \theta, \varphi) = P^{\text{CG}}(r)P^{\text{CG}}(\theta)P^{\text{CG}}(\varphi). \quad (2)$$

To obtain the bonded potentials, the individual distributions $P^{\text{CG}}(r)$, $P^{\text{CG}}(\theta)$, and $P^{\text{CG}}(\varphi)$ are first fitted by a suitable sum of Gaussians functions and then Boltzmann inverted. It should be noted that the bond length and bond angle probability distributions are normalized by taking into account the corresponding metric, namely r^2 for bond lengths and $\sin(\theta)$ for bending angles. It should be noted that the Boltzmann inversion of a distribution leads to a potential of mean force (PMF), i.e., a free energy, which is only in certain limiting cases identical to a potential energy. This means that using a free energy in place of a potential energy is wrong in a strict statistical–mechanical sense. In the case of bonded interactions, however, which are rather stiff and energy-dominated and which separate well from the

remaining degrees of freedom, this approach is nevertheless often a good numerical approximation.

Non-bonded interactions are derived as effective non-bonded potentials $U_{\text{nb}}^{\text{CG}}(r)$ from a given target intermolecular RDF, $g^{\text{target}}(r)$, obtained from atomistic reference simulations or experimental data. First, a reasonable initial guess is needed. It can be obtained by directly Boltzmann-inverting the RDF (which is a probability distribution):

$$F(r) = -k_{\text{B}}T \ln(g^{\text{target}}(r)), \quad (3)$$

where k_{B} is the Boltzmann constant and T is the temperature. It is important to notice that $F(r)$ is a free energy and cannot be used directly as a two-body interaction potential in the CG model because it incorporates multibody contributions of all the other particles in the system in a statistically averaged way (see above). However, it is usually sufficient as an initial guess, $U_{\text{nb},0}^{\text{CG}}(r)$, for the iterative procedure whereby these multibody contributions are eliminated and an effective two-body interaction potential is determined that reproduces the target structure. Simulating the system with $U_{\text{nb},0}^{\text{CG}}(r)$ now yields a corresponding RDF, $g_0^{\text{CG}}(r)$, which is different from $g^{\text{target}}(r)$. Therefore, the CG potential needs to be improved, and this can be done by adding to $U_{\text{nb},0}^{\text{CG}}(r)$ a correction term $-k_{\text{B}}T \ln(g_0^{\text{CG}}(r)/g^{\text{target}}(r))$. This step is iterated:

$$U_{\text{nb},i+1}^{\text{CG}}(r) = U_{\text{nb},i}^{\text{CG}}(r) + k_{\text{B}}T \ln\left(\frac{g_i^{\text{CG}}(r)}{g^{\text{target}}(r)}\right) \quad (4)$$

until the reference $g^{\text{target}}(r)$ is reproduced and the potential is stationary, $U_{\text{nb},i+1}^{\text{CG}}(r) = U_{\text{nb},i}^{\text{CG}}(r)$. The convergence can be measured quantitatively by evaluating the following error function:

$$f_{\text{target}} = \int_0^{r_{\text{cutoff}}} \int w(r)(g_i^{\text{CG}}(r) - g^{\text{target}}(r))^2 dr, \quad (5)$$

where $w(r) = \exp(-r/\sigma)$ is a weighting function to penalize more strongly the deviations at small distances.

Since the IBI method has no obvious way in which the system energy or pressure can influence the value of the potential at a particular distance, the following approach can be used to add this information. Adding to the non-bonded potential a weak linear potential term ΔV , which goes to zero at the cutoff and whose slope is positive or negative (V_0), does not significantly change the RDFs produced by the model but changes the pressure down or up, respectively. The so-called ramp correction is of the form:

$$\Delta V = V_0 \left(1 - \frac{r}{r_{\text{cutoff}}}\right). \quad (6)$$

This correction can be inserted into the Boltzmann-inversion iterations to adjust the pressure to the target value.

On the basis of the CG simulations performed by our group, and in order to validate the workflow of developing CG models using the IBI method, a new program package (called IBIsCO) has been developed recently especially for CG simulations using Gaussian potential functions and/or tabulated interaction potentials derived by the IBI approach [54]. Various standard ensembles (NVT, NPT, and NVE) are available in IBIsCO. The techniques of dissipative particle dynamics (DPD) [55] and Lowe–Andersen (LA) [56] equations of motion are also embedded in IBIsCO. Besides their use as thermostats for the generation of the canonical ensemble, DPD and LA can also be used as techniques to compensate for the effects of lost degrees of freedom in CG models on the dynamics. These techniques slowdown the too-fast dynamics in CG models due to the softness and the lack of friction [57], which will be discussed in more detail later. IBIsCO also includes an implementation of the reverse nonequilibrium molecular dynamics method for the calculation of viscosities [58]. A detailed description of the IBIsCO code is presented in [54].

Because the RDF incorporates temperature, density, composition, and other dependencies into the effective pair interaction, the force field developed by IBI can have a severely limited range of applicability, and transferability of the CG force field is still a challenge [15, 18]. In the literature, several attempts at using IBI force field mapped at a specific temperature in a broader range have been reported. Vettorel and Meyer [59] faced the problem in studying the crystallization of a CG model of polyethylene, and the effect of the temperature changes in the model was checked by looking at the different effective potentials and monitoring their behavior as the temperature was modified. They found that the bond potential is temperature-independent, whereas the PMF obtained by direct Boltzmann inversion of the angle distributions depends on the temperature chosen for the mapping. However, after the iterative procedure that leads to the optimized potentials, the mismatch tends to disappear. Similar results were obtained for the non-bonded interactions. These observations allowed the authors to use the same potential for studying the crystallization of polyethylene. Ghosh and Faller [60] investigated a small organic glass former (*ortho*-terphenyl) using a mesoscale model composed of only a single interaction center. The authors used the same IBI potential at different temperatures and compared the resulting structural properties (in their case only the RDF) with the corresponding atomistic ones. In this way, they found that the CG potential depends not only on the structure but implicitly also on the temperature at which it has been optimized. In our recent studies [15, 18], we chose the strategy followed by Ghosh and Faller to investigate the transferability of the IBI force field by comparing the CG structural and dynamical properties with the atomistic reference calculations. Moreover, we investigated whether the degree of coarse-graining (how many real atoms per bead) and the size of the macromolecule affect the transferability. We analyzed the polymer case by investigating bulk melts of PS and PA-66 whose CG models differ in the chain length and in the number of atoms per bead (Fig. 1a,c). We found that the finer model used for PA-66 allowed us to use

one IBI potential over the entire temperature range of interest (300–600 K), and all properties investigated showed good agreement with experimental and atomistic results. In contrast, for PS, by using PS-MS2 the analysis of the intramolecular distribution of parameters such as distances and angles, as well as the RDFs, showed that the PS IBI force field can be confidently applied only in a small temperature range (~ 50 K) around the optimization temperature. Within this range, the density and the static properties of the PS bulk are in reasonable agreement with experimental and atomistic values; however, for temperatures further from the optimization point, the IBI potential cannot correctly reproduce the behavior of the polymer. By changing the mapping scheme from PS-MS2 to PS-MS1, the CG potential turned out to be very robust and transferable between different temperatures (over a range of 100 K) [18]. Figure 5 shows the density changes with the temperature for the atomistic and CG simulations of PS using PS-MS1 and PS-MS2. These results show that the transferability of the CG force field developed by the IBI method depends strangely on the location of the superatom within the real monomer, the number of degrees of freedom removed during the CG procedure, and the polymer under investigation.

Concerning the transferability of CG force field for PA-66 to different temperatures, we explored different thermodynamic and structural properties of the system at different temperatures [15, 16, 61]. The hydrogen bonding is one of the intermolecular interactions that most influences the dynamics of molecular systems, being responsible for the structure, function, and dynamics of many chemical systems from inorganic to biological compounds [62]. Due to the simplification of the CG

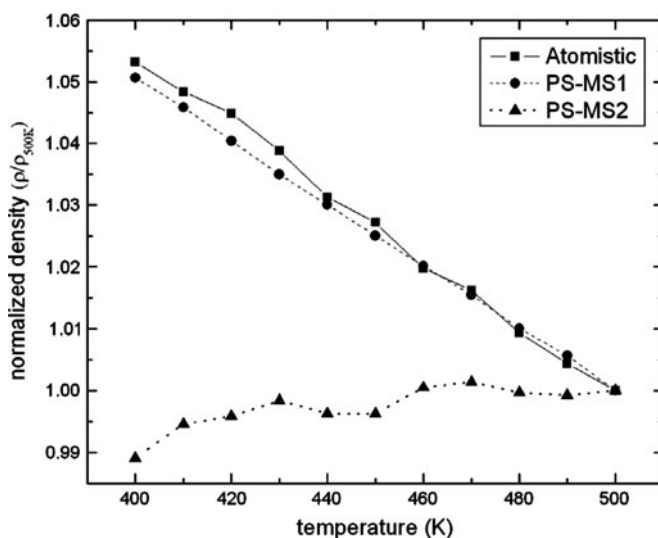


Fig. 5 Density change with the temperature for the atomistic and CG simulations of PS using PS-MS1 and PS-MS2 (see Fig. 1c). The density values have been normalized with respect to the reference value at 500 K

models, the atoms directly involved in the hydrogen bonding (oxygen, nitrogen, or fluorine as hydrogen bond donors and acceptors) as well as the hydrogen atom itself are usually ‘coarse-grained away’, i.e., lumped together with other atoms into beads. Several models have been developed to describe hydrogen bonding, especially in studying protein folding [63], and have met with different degrees of success. In the case of synthetic macromolecules, the presence of the hydrogen bond strongly affects their conformation, chemical–physical properties, crystallization, self-assembly behavior, and many other global properties. Since the hydrogen bonds are only present in an effective and averaged way, it is therefore particularly interesting to see whether and how the properties directly affected by the hydrogen bonds are preserved in the CG model. In addition, the possibility of correctly describing the hydrogen bond dynamics using a CG model would be of great importance for further improvements of CG force fields. In polyamides, nearly all the amide groups that are separated by a sequence of methylene groups are hydrogen-bonded [64]. The large number of hydrogen bonds forms an extended three-dimensional network whose dynamic rearrangement influences several properties of the material, such as the glass transition temperature and the melting point. For these reasons, understanding the thermal mechanical properties of polyamides by studying the thermal stability of hydrogen bonds has been a popular topic in previous research [64–68]. In our recent publications [16, 61], we first describe the detailed analysis of the effect of temperature on the local and global dynamics of unentangled PA-66 using atomistic molecular dynamics simulations. The local dynamics was mainly investigated by looking at the hydrogen bond dynamics and calculating the hydrogen-bond relaxation time and lifetime by means of specific correlation functions. The influence of the relaxation of the hydrogen-bond network on the global dynamics of the polymer was also analyzed. Our results show that the global dynamics of PA-66 is intimately related to the relaxation of the hydrogen-bond network formed among the amide groups. Then, we studied a CG model of the same PA-66 system (as shown in Fig. 1a), focusing on the dynamics (and thermodynamics) of the hydrogen bond [16]. The ability of the CG model to capture correctly the dynamics of the hydrogen-bond network at different temperatures was tested. To address this issue we then used the same correlation functions that were employed in the analysis of hydrogen bond dynamics in atomistic simulations. From a quantitative analysis of the hydrogen bond dynamics and thermodynamics, it turned out that the CG model is characterized by a weaker hydrogen-bond network than the corresponding atomic model. The weakness of the CG hydrogen bonding might be due to the lack of directionality as a consequence of the mapping scheme where the donor and acceptor atoms are lumped into spherical beads. Hence, as happens for biological systems, the necessity to introduce explicitly a new interaction accounting for the directionality of the hydrogen bonding interactions and their increasing strength with the decrease in the temperature is probably fundamental to the analysis of processes that are governed by their dynamics, such as self-assembly or crystallization in the polymers.

Transferability is also an issue for mixtures of different species. The canonical protocol to derive CG potentials for, say, mixtures of A and B would be to run an

atomistic reference simulation of a small A–B mixture and to generate CG potentials for A–A, B–B, and A–B interactions from it. This would have to be repeated at every composition of the mixture that is of interest. The approach is straightforward, but inefficient and cumbersome. Moreover, it precludes the coarse-graining of systems in which A and B phase-separate. For these reasons it is therefore desirable to come up with schemes in which coarse-graining is done for the individual components A and B separately, and the mixed interaction potentials are then obtained via some combinations rules, similar to the widely used combinations rules for atomistic force fields [69]. Some progress has been achieved recently in a study of PS solutions in EB [18]. Here, it was determined that IBI-derived CG potentials for EB and PS could be combined by taking their geometric average. The resulting CG potential for mixed EB–PS interactions successfully described structural and thermodynamic properties of the solutions at all compositions studied. Further research is needed in order to establish whether a geometric combination rule is a general option or whether its success is coincidental and due to the chemical similarity of the two components.

Investigations of dynamic properties such as mean-squared displacement, diffusion constant, and Rouse-mode analysis necessitate the transition from unentangled to entangled motion for IBI force fields, and it turns out that such structure-based CG potentials can be used for a qualitative study of the dynamics of polymer systems [10, 36]. The CG force fields in general do reproduce, e.g., the scaling behavior of the dynamics. However, since many of the original degrees of freedom are removed in the CG description, the effective CG potentials are softer compared to the atomistic ones, and this results in a reduced effective friction between the beads. Thus, CG simulations cannot be used directly for quantitative predictions of the dynamics. Of course, the three basic units (particle mass, size, and energy scale) define a time scale in the MD simulation of the CG systems, but the time in the CG description does not correspond to the real physical time of the underlying mobility. One of the main problems of such CG models is the artificial dynamics, which are too fast compared to either atomistic or experimental reference data [15, 70]. To re-establish the correct dynamics in CG simulations, different approaches have been proposed. Izvekov and Voth [71] proposed an approach within the coarse-graining framework of force matching (see below) that reproduced correct dynamics in the CG simulation. However, in order to map the time accurately between the atomistic and the structure-based CG models one can use one of the following two methods. The first method is to gauge the CG dynamics by equating a scalar dynamical quantity like the diffusion coefficient or the viscosity [15, 70]. The results of the CG model could thus be matched to the value from long atomistic MD runs or experiments. By doing this, only the asymptotic long time scale regime is compared, and one hopes that one single time-scaling factor covers all dynamic processes. The second way to map the time is to match the mean-square displacement (MSD) of the monomers [37, 72, 73], if there is data available from atomistic MD simulations. The time-scaling factor determines the real unit to which the CG time corresponds.

According to the workflow presented here to develop the CG force fields using the IBI method, one obtains potentials for bonded and non-bonded interactions at the same time on the basis of the same atomistic simulation; thus there is no clear separation between the optimization procedures for bonded and non-bonded interaction potentials. One can achieve this separation by deriving CG bond length, bond angle, and torsional distributions from the atomically detailed conformations sampled by a single (chain) molecule in vacuum, if the conformational sampling of the molecule in vacuum and in the bulk (or solution) phase does not differ substantially [48]. The IBI method has the advantage that detailed structural information is included into the CG model, and it has been used successfully for molecular liquids [18], polymer melts [15, 37, 73], dendrimers [20], polymer solutions [18], polymer blends [74], and ionic liquids [19].

However, there can be limits to this approach because it is not always clear whether the chosen CG mapping scheme can converge to an optimal fit. For liquid mixtures or solutions, the situation is more complex because several RDFs that mutually affect each other need to be simultaneously reproduced. In addition, for dilute solutions, where we have a low concentration of solute, the solute–solute RDFs converge very slowly in the CG simulations. In this case, the PMF between the solute molecules can be obtained using free-energy calculation methods such as umbrella sampling or constraint dynamics. Recently, these methods have been used in an iterative optimization approach to study self-assembling dipeptides at the CG scale [75, 76]. The PMF between solute molecules in a solvent box, $V_{\text{PMF}}^{\text{AA}}(r)$, is calculated by all-atom simulation from n distance constraint simulations:

$$V_{\text{PMF}}(r) = \int_{r_m}^r \left[\langle f_c \rangle_s + \frac{2k_B T}{s} \right] ds + C, \quad (7)$$

where f_c is the constraint force, and r_m is the maximum distance between the centers of mass of the two molecules. This potential was successfully employed to simulate the aggregation process of a hydrophobic dipeptide in solution with an implicit solvent representation in a CG model [76]. Since the so-obtained PMF incorporates the thermally averaged contributions from solute and solvent degrees of freedom, it cannot be directly used as CG potential if the CG model has an explicit solvent representation. To determine the solvent contribution that needs to be removed from $V_{\text{PMF}}^{\text{AA}}(r)$, the PMF calculations with the CG potential are run, while the direct solute–solute interactions are excluded. The effective solute–solute potential can then be obtained by subtracting $V_{\text{PMF,excl}}^{\text{CG}}(r)$ from the all-atom PMF, $V_{\text{PMF}}^{\text{AA}}(r)$ [76]. This subtraction procedure removes the solvent contribution from the PMF, and is similar to iteration steps in the IBI method. Recently, the method has also been used to develop a CG model for PS [77]. To derive the non-bonded interactions, PS oligomer pairs were simulated in vacuum with a detailed atomistic model. The effective non-bonded potentials obtained in this procedure include the effects of multibody correlations related to the chain connectivity (Brini et al., 2010, unpublished results).

3.1.2 Multiscale Coarse-Graining Method (Force Matching Method)

With the goal of providing a systematic multiscale approach to coarse graining, Izvekov and Voth introduced the multiscale coarse-graining (MS-CG) method (force matching method) [78, 79]. In this method, the forces in the CG system are determined such that they are mapped to corresponding sums of forces in the corresponding atomistic system [80–82]. The MS-CG method has been applied to the development of accurate CG models for peptides [83, 84], pure bilayers [78], mixed bilayers [85], carbohydrates [83], simple fluids [71, 79], ionic liquids [86, 87], soot nanoparticles [88], and mixed-resolution models of transmembrane proteins [89]. The MS-CG theory can also serve as a basis for achieving more correct dynamic behavior (e.g., self-diffusion) in the CG model [71]. If no approximations are introduced into the method, the MS-CG variational principle provides a computational algorithm for determining the many-body CG free-energy surface for a given atomically detailed model. The bonded parameters of the potentials developed by MS-CG method are found to be transferable to different temperatures, whereas the non-bonded potentials are less transferable. However, the MS-CG models are well transferable to different system sizes [86]. Recently, a three-body potential has been used to develop a one-site CG model for water to improve the results over the two-body approximation [90]. However, the effects of electrostatic interactions and direct comparison with the other methods need further investigation.

3.1.3 MARTINI Force Field

The MARTINI force field, in close connection with atomistic models, has been developed as another method for obtaining the interaction potentials between the CG beads. The method's philosophy of the coarse-graining approach is substantially different from the other methods [32, 91]. Instead of focusing on an accurate reproduction of structural details at a particular state point for a specific system, the aim is for a broader range of applications without the need to reparameterize the model each time by extensive calibration of the chemical building blocks of the CG force field against thermodynamic data. Currently, the MARTINI force field provides parameters for a variety of biomolecules, including many different lipids, cholesterol, and all amino acids. Properties accurately reproduced include structural [32, 92, 93], elastic [32, 91], dynamic [32], and thermodynamic data [91, 93, 94]. In order to parameterize the non-bonded interactions of the CG model, a systematic comparison with experimental thermodynamic data has been performed. Specifically, the free energy of hydration, the free energy of vaporization, and the partitioning free energies between water and a number of organic phases were calculated for each of the different CG particle types. To parameterize the bonded interactions, the method uses structural data that are either directly derived from the underlying atomistic structure (such as bond lengths of rigid structures) or obtained from comparison with fine-grained simulations. In the latter procedure, the fine-grained

simulations are first converted into a “mapped” CG simulation by identifying the center of mass of the corresponding atoms as the mapped CG bead. Second, the distribution functions are calculated for the mapped simulation and compared to those obtained from a true CG simulation. Subsequently, the CG parameters are systematically changed until satisfactory overlap of the distribution functions is obtained. The MARTINI force field has also been applied recently to model polymers such as polyethylene glycol [95] and PS [96]. To reproduce the specific structural properties of polymer systems, the radius of gyration of the polymer chains has also been used as a target in the parameterization of the non-bonded interactions for the two different mapping schemes proposed for PS [96]. Different aspects of the CG force field compared to the previous models developed for PS have been discussed in [96].

The range of applications of the MARTINI force field is very broad. There are, however, certain important limitations that should be kept in mind. For example, the model has been parameterized for the fluid phase. Thus, properties of solids, such as crystal packing, are not expected to be accurate. On the other hand, both the gas and the solid phase appear somewhat too stable with respect to the fluid phase [91], and therefore the thermodynamic behavior of solid–fluid and gas–fluid interfaces should be interpreted with care, at least at the quantitative level.

3.1.4 Newton Inversion Method

Another systematic way to construct CG models from detailed atomistic simulations is the Newton inversion method [97]. In this method, the structural information extracted from atomistic simulations is used to determine effective potentials for a CG model of the system. Suppose the effective potentials in the CG model are determined by a set of parameters $\{\lambda_i\}$ where i runs from 1 to the number of parameters in the potential. The set of target properties that are known from atomistic simulations is represented by $\{A_j\}$, where j changes from 1 to the number of target properties. By means of the Newton inversion method, a set of nonlinear multidimensional equation between $\{\lambda_i\}$ and computed average properties $\{\langle A_j \rangle\}$ is solved iteratively. At each iteration of the Newton inversion, the effect of different potential parameters on different averages can be calculated by the following formula [97]:

$$\frac{\partial \langle A_j \rangle}{\partial \lambda_i} = -\beta \left(\left\langle \frac{\partial H}{\partial \lambda_i} A_j \right\rangle - \left\langle \frac{\partial H}{\partial \lambda_i} \right\rangle \langle A_j \rangle \right), \quad (8)$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant and H is the Hamiltonian of the CG system. By using (8) and solving the system of linear equations (2), the parameters $\{\lambda_i\}$ corresponding to the target values of $\{\langle A_j \rangle\}$ can be found:

$$\Delta \langle A_j \rangle = \sum_i \frac{\partial \langle A_j \rangle}{\partial \lambda_i} \Delta \lambda_i + O(\Delta \lambda^2), \quad (9)$$

where the second-order corrections are neglected. Simulation starts from some initial potential determined by a trial set of parameters and, after running the simulation, the deviation of computed average properties $\{\langle A_j \rangle\}$ from the target values ($\Delta\langle A_j \rangle$) as well as (8) is determined. Then, from (2) the corrections to the potential parameters $\Delta\lambda_i$ can be found. The procedure is repeated with the new parameter set until convergence is reached. In the case where the parameters $\{\lambda_i\}$ are the values of the pair potential at a number of points covering the whole range of distances, and the target properties are the values of RDF at the same set of points, the method becomes equivalent to the inverse Monte Carlo approach [98, 99]. The method has been used successfully to develop a united atom model for water, a CG model for an equimolar mixture of L- and D-proline dissolved in dimethyl sulfoxide, and a CG model of dimyristoyl phosphatidylcholine lipid molecules. However, the transferability of the CG potentials needs to be checked in every case [97].

3.2 *Dynamic-Based CG Model*

An alternative way to develop a CG force field is a starting from the dynamic properties of the system. In this case, the Langevin-equation formalism [10, 100] is used to describe the dynamic evolution of the system, and the friction coefficients that partially slow down the dynamics are determined from atomistic reference simulations using force–velocity and velocity–velocity correlation functions [34, 71]. This method is usually used to study complex liquids [101] or biomolecular systems [85]. In the same class of methods also fall those that tune the friction coefficients until the dynamic properties match the atomistic ones [33]. In any case, it is of interest to understand the physical origins of the acceleration of the CG dynamics for specific cases, to assess the methods mentioned above, and to gain a better understanding of the effect of coarse graining on the dynamics of a system. However, this class of method could fail to reproduce the structure of the system, since the developments of the CG force field only take care of the dynamic properties of the system. There is currently much research being carried out to investigate, whether it is possible to derive coarse-grained potentials that are both dynamically and structurally consistent with the underlying higher-resolution description. In a recent work of Qian et al. [57], the DPD [55] and LA [56] equations of motion have been applied in CG simulations to slow down the dynamics of the CG model obtained through the IBI method. The simulation results showed that both DPD and LA could re-introduce friction into the system and compensate for the dynamic effects of coarse-graining. Thus, the too-fast dynamics of CG models in molecular dynamics can be corrected and can be slowed down to match reality. Empirical rules have been found for the control parameters (noise strength in DPD and bath collision frequency in LA) in CG simulation of liquid EB [57]. Further work needs to be done to establish how transferable these rules are among different systems.

The different simulation hierarchies (QM, atomistic MD, and CG simulations) can be used to address phenomena or properties of a given system at several levels

of resolution and, consequently, on several time and length scales. The easiest way to combine different simulation models on different scales is to treat them separately and sequentially by simply passing information (structures, parameters, energies etc.) from one level of resolution to the next. A step beyond these sequential schemes is involved in those approaches where the scales are coupled in a concurrent fashion within a unified computational scheme. In these approaches, two or more levels of resolution are used at the same time in the simulation. A dual-scale approach has already been used to study the interaction between bisphenol-A polycarbonate and a nickel surface [102, 103]. In this method, the regions with different resolutions are fixed and the exchange of particles among the different regions is not allowed. While this may not be a crucial point for hard matter, it is certainly a strong limitation for soft matter (i.e., complex fluids) since relevant density fluctuations are arbitrary. An even more sophisticated multiscale approach allows adaptive switching between resolution levels for individual molecules on the fly, e.g., depending on their spatial coordinates. Recently, such an adaptive resolution scheme (AdResS) has been developed in which molecules can freely exchange between a high-resolution (CG) and a low-resolution (atomistic) region by changing the molecular degree of freedoms [104–110]. In this case, the atomistic and the CG scales can be coupled via a position-dependent interpolation formula on the atomistic and CG force in such a way that allows a smooth transition from atomistic to CG trajectories without altering the equilibrium of the system [111]. The method has been already used for liquid water [105] and for a polymer–solvent system in which the water molecules within few solvation shells around the polymer chains are considered atomistically while outside the water is treated on a rather coarse level [106]. It has even been augmented by a continuum region, and a methane-like liquid has been simulated using this triple-resolution scheme [112].

3.3 *Coarse-Graining in Time*

Although CG models have been successfully used to simulate large systems for very long time and length scales, the lack of detailed atomistic information in CG simulations still limits the systems and the properties that can be studied using these models. As an alternative to the spatial coarse-graining techniques, Violi [113] proposed a novel method to describe the evolution of reactive systems (diffusion processes and chemical reactions) over long time scales while preserving an all-atom description of the system by coarse-graining in time. The method combines the MD methodology with kinetic Monte Carlo (KMC) to allow the extension of the accessible time scales compared to the direct MD simulation [114]. In the KMC step, the structure of the growing species is modified during the reaction and then the newly formed structure is relaxed towards thermal equilibration using an MD run. The MD describes the local phase space changes and rearrangement reactions and allows for relaxation as well as processes very far from equilibrium. The KMC method is responsible for the conformational changes that jump to a completely

different area of phase space and allows much larger time-scale changes to the system than the MD simulation. The method has been used to study the formation of carbonaceous nanoparticles with an average diameter of 50 nm by using the AMPI (atomistic model for particle inception) code [113].

4 Back-Mapping

The combination of CG simulation with an efficient back-mapping methodology (i.e., reintroduction of atomistic detail) is also a powerful tool for efficiently obtaining well-equilibrated atomistic structures. In general, the back-mapping procedure has no unique solution because every CG structure corresponds to an ensemble of atomistic microstates. Therefore, one needs to find one representative all-atom structure with the correct statistical weight of those degrees of freedom that are not resolved in the CG description. Several slightly different strategies for reintroducing atomistic detail into a CG structure have been presented [13, 16, 115, 116]. When the mesoscale model is tailored on the atomic contour using atomic distributions to build up the CG force field, zooming back to the atomic description is usually a simple geometrical problem. The general strategy is to use reasonably rigid all-atom chain fragments (corresponding to a single or a small set of CG beads) that were taken from a correctly sampled distribution of all-atom chain structures. An alternative for the case of more flexible low-molecular-weight molecules is to restrain some atomistic coordinates to the CG structure to avoid deviation of the atomistic structure too strongly from the CG reference [117]. In some cases, as for Santangelo and coworkers [115], if the model is particularly coarse or the beads contain asymmetric atoms and the polymer chain has a specific tacticity, a more sophisticated method must be followed: For instance, the atomic fragment inserted into the CG model must be chosen from several that correspond to the same type of bead. The structures resulting from the back-mapping procedure can be directly compared to experimental data (e.g., X-ray or neutron scattering) or they can be used in further computations, for example to determine dynamic data (e.g., the permeabilities of small molecules in large polymeric systems) [118–120]. Additionally, the combination of CG simulations, where the CG model is based on an underlying atomistic description, with a back-mapping procedure can be further employed to validate the atomistic force field on time and length scales not accessible to atomistic simulations.

5 Outlook

The key motivation for CG molecular modeling and simulation derives from the need to bridge the atomistic and mesoscopic scales. Typically, there are two to three orders of magnitude in length and time separating these regimes. Only at the mesoscopic scale can one see the emergence of important phenomena

(e.g., self-assembly in biomolecular or soft matter systems). CG simulations, especially as the aim is to make increasing contact with experimental results for complex systems, therefore play a significant role in the exploration of mesoscopic phenomena and, in turn, of the behavior of real biomolecular and materials systems.

Although CG models provide a highly efficient computational tool for rapidly investigating different properties of the system with a desired resolution, they face a number of significant challenges before they can become widely utilized by the research community, especially by experimental researchers as a tool to help interpret their experiments. As discussed before, the structure-based CG models are state-point dependent, which means that the potentials obtained at a given thermodynamic condition do not generally provide a good description of the structure and other properties at other conditions. Thus, one needs to test the transferability for each CG model individually. Our results show that for a defined mapping scheme, IBI potentials develop independently and with different shape, and give comparable self-diffusion coefficients for high-enough temperatures. At high temperatures, the specifics of a force field become unimportant and only global properties such as excluded volume and bead connectivity prevail. We have also shown that the scaling factor measuring the artificial speed-up of the CG model over the parent atomistic model depends on the simulation temperature. A key goal then is both to define and to understand what is and what is not transferable in a given CG model and why. Recent work by Harmandaris et al. [73] shows that the dependence of polymer dynamics on density is not described accurately with the CG model, whereas the dependence on chain length is the same as in atomistic simulations. Thus, at high molecular weights where the change in the polymer dynamics is entirely due to the increase of the molecular weight we will have a constant scaling factor between the atomistic and the CG model. The asymptotic plateau value of the scaling factor allows us to quantitatively predict the diffusion coefficient (and of other dynamic properties) of higher molecular weight polymer melts directly from the CG simulations. Since we have IBIsCO as a powerful tool for CG simulations, combining these results with the recent work of Qian et al. [57] to control the fast dynamics in the CG models could lead to a robust method for calculating the viscosity of long polymer chains.

Another challenge involves the establishment of a proper formal connection between the behavior of the CG representation of the system and the underlying all-atom (full atomic resolution) model. In many systems, the formation (e.g., self-assembly) and dynamics of large-scale structures and conformations cannot be decoupled from local, chemical processes and specific intermolecular interactions. A hydrogen bond is an attractive interaction acting between an electronegative atom (the acceptor) and a hydrogen atom bonded to a donor nitrogen, oxygen, or fluorine. Due to the simplification of the CG models, the atoms directly involved in the hydrogen bonding (donor and acceptor) are usually ‘coarse-grained away’ i.e., lumped with other atoms into beads. The poorly described hydrogen bonding interactions can lead to an unphysical CG dynamics that prevents the correct description of the collective properties of the polymers. A method that explicitly introduces an orientation-dependent CG hydrogen bonding potential would allow

the study of those collective phenomena in materials that, driven by the presence of hydrogen bonds, cannot be investigated with an atomistic approach. Polymer crystallization and self-assembly of block copolymers could be the first objects of investigation. Because hydrogen bonding is the driving force in many biological processes, the new force field approach could be particularly suitable for the study of biopolymers such as polysaccharides and of biomaterials in which a polymer interacts with a biological system.

The proposed CG models can also be used for the study of systems more complicated than bulk polymer melts. Possible examples are the study of the diffusion of a penetrant in a polymer matrix, or of block copolymers, blends, etc. [121, 122]. In addition, the method can be directly incorporated into multiscale methodologies, which include multiple levels of simulation, and where both atomistic and mesoscopic descriptions are needed at the same time, but in different regions. An example is the study of the long time scale dynamics of polymers near solid attractive surfaces, where an atomistic description is needed very close to the surface but a mesoscopic description can be used for length scales far from the surface.

Acknowledgments The authors thank Dr. V. Harmandaris and Dr. Qian for providing the data for RDFs of PS and EB in united atom and CG simulations.

References

1. Kremer K (2006) Polymer dynamics: long time simulations and topological constraints. In: Ferrario M, Ciccotti G, Binder K (eds) Computer simulations in condensed matter systems: from materials to chemical biology, vol 2. Lecture notes in physics, vol 704. Springer, Berlin, Heidelberg, p 341
2. Martonak R, Paul W, Binder K (1997) *J Chem Phys* 106:8918
3. Goldstein H, Poole C, Safko J (2002) Classical mechanics. Addison Wesley, San Francisco
4. Harmandaris VA, Mavrantzas V (2004) Molecular dynamic simulations of polymers. In: Theodorou DN, Kotelyanski M (eds) Simulation methods for polymers. Marcel Dekker, New York
5. Baschnagel J, Binder K, Doruker P, Gusev AA, Hahn O, Kremer K, Mattice WL, Müller-Plathe F, Murat M, Paul W, Santos S, Suter UW, Tries V (2000) *Adv Polym Sci* 152:41
6. Binder K (1995) Monte Carlo and molecular dynamics simulations in polymer science. Oxford University Press, New York
7. Theodorou DN (2010) *Ind Eng Chem Res* 49:3047
8. Mavrantzas VG, Boone TD, Zervopoulou E, Theodorou DN (1999) *Macromolecules* 32:5072
9. Karayiannis NC, Giannousaki AE, Mavrantzas VG, Theodorou DN (2002) *J Chem Phys* 117:5465
10. Kremer K, Grest GS (1990) *J Chem Phys* 92:5057
11. Muller M, Katsov K, Schick M (2006) *Phys Rep* 434:113
12. Reynwar BJ, Illya G, Harmandaris VA, Müller MM, Kremer K, Deserno M (2007) *Nature* 447:461
13. Auhl R, Everaers R, Grest GS, Kremer K, Plimpton SJ (2003) *J Chem Phys* 119:12718
14. Grest GS, Kremer K (1986) *Phys Rev A* 33:3628

15. Carbone P, Karimi-Varzaneh HA, Chen X, Müller-Plathe F (2008) *J Chem Phys* 128:064904
16. Karimi-Varzaneh HA, Carbone P, Müller-Plathe F (2008) *J Chem Phys* 129:154904
17. Milano G, Müller-Plathe F (2005) *J Phys Chem B* 109:18609
18. Qian H-J, Carbone P, Chen X, Karimi-Varzaneh HA, Liew CC, Müller-Plathe F (2008) *Macromolecules* 41:9919
19. Karimi-Varzaneh HA, Müller-Plathe F, Balasubramanian S, Carbone P (2010) *Phys Chem Chem Phys* 12:4714
20. Carbone P, Negri F, Müller-Plathe F (2007) *Macromolecules* 40:7074
21. Harmandaris VA, Reith D, van der Vegt NFA, Kremer K (2007) *Macromol Chem Phys* 208:2109
22. Harmandaris VA, Adhikari NP, van der Vegt NFA, Kremer K (2007) *Macromolecules* 40:7026
23. Triolo A, Mandanici A, Russina O, Rodriguez-Mora V, Cutroni M, Hardacre C, Nieuwenhuyzen M, Bleif HJ, Keller L, Ramos MA (2006) *J Phys Chem B* 110:21357
24. Lopes J, Padua AAH (2006) *J Phys Chem B* 110:3330
25. Bhargava BL, Devane R, Klein ML, Balasubramanian S (2007) *Soft Matter* 3:1395
26. Urahata SM, Ribeiro MCC (2004) *J Chem Phys* 120:1855
27. Praprotnik M, Delle Site L, Kremer K (2008) *Annu Rev Phys Chem* 59:545
28. Izvekov S, Parrinello M, Burnham CJ, Voth GA (2004) *J Chem Phys* 120:10896
29. Luybartsev AP, Laaksonen A (1995) *Phys Rev E* 52:3730
30. Spyriouni T, Tzoumanekas C, Theodorou D, Müller-Plathe F, Milano G (2007) *Macromolecules* 40:3876
31. Nielsen SO, Srinivas G, Klein M (2005) *J Chem Phys* 123:124907
32. Marrink SJ, Vries AHd, Mark AE (2004) *J Phys Chem B* 108:750
33. Padding JT, Briels WJ (2002) *J Chem Phys* 117:925
34. Li XJ, Kou DZ, Rao SL, Liang H (2006) *J Chem Phys* 124:204909
35. Haliloglou T, Mattice WL (1998) *J Chem Phys* 108:6989
36. Sun Q, Faller R (2006) *Macromolecules* 39:812
37. Harmandaris VA, Adhikari NP, van der Vegt NFA, Kremer K (2006) *Macromolecules* 39:6708
38. Larson RG (2004) *Mol Phys* 102:341
39. Fukunaga H, Takimoto J, Doi M (2002) *J Chem Phys* 116:8183
40. Reith D, Meyer H, Müller-Plathe F (2001) *Macromolecules* 34:2335
41. Reith D, Pütz M, Müller-Plathe F (2003) *J Comput Chem* 24:1624
42. Tschöp KKW, Batoulis J, Bürger T, Hahn O (1998) *Acta Polym* 49:61
43. Paul W, Binder K, Kremer K, Heermann D (1991) *Macromolecules* 24:6332
44. Abrams C, Kremer K (2003) *Macromolecules* 36:260
45. Milano G, Goudeau S, Müller-Plathe F (2005) *J Polym Sci Pol Phys* 43:871
46. Queyroy S, Neyertz S, Brown D, Müller-Plathe F (2004) *Macromolecules* 37:7338
47. Meyer H, Müller-Plathe F (2002) *Macromolecules* 35:1241
48. Peter C, Delle Site L, Kremer K (2008) *Soft Matter* 4:859
49. Faller R, Schmitz H, Biermann O, Müller-Plathe F (1999) *J Comput Chem* 20:1009
50. Meyer H, Biermann O, Faller R, Reith D, Müller-Plathe F (2000) *J Chem Phys* 113:6264
51. Soper AK (1996) *Chem Phys* 202:295
52. Reith D, Meyer H, Müller-Plathe F (2002) *Comput Phys Commun* 148:299
53. Henderson RL (1974) *Phys Lett A* 49:197
54. Karimi-Varzaneh HA, Qian H-J, Chen X, Carbone P, Müller-Plathe F (2011) *J Comput Chem*. DOI: 10.1002/jcc.21717
55. Groot RD, Warren PB (1997) *J Chem Phys* 107:4423
56. Lowe CP (1999) *Europhys Lett* 47:145
57. Qian H-J, Liew CC, Müller-Plathe F (2009) *Phys Chem Chem Phys* 11:1962
58. Bordat P, Müller-Plathe F (2002) *J Chem Phys* 116:3362
59. Vettorel T, Meyer H (2006) *J Chem Theory Comput* 2:616

60. Ghosh J, Faller R (2007) *Mol Simul* 33:759
61. Karimi-Varzaneh HA, Carbone P, Müller-Plathe F (2008) *Macromolecules* 41:7211
62. Steiner T (2002) *Angew Chem Int Ed* 41:48
63. Sancho DD, Rey A (2007) *J Comput Chem* 28:1187
64. Murthy NS (2006) *J Polym Sci B Polym Phys* 44:1763
65. Garcia D, Starkweather HW (1985) *J Polym Sci Polym Phys Ed* 23:537
66. Bessler E, Bier G (1969) *Makromol Chem* 122:30
67. Trifan DS, Terenzi JF (1958) *J Polym Sci* 28:443
68. Schroeder LR, Cooper SL (1976) *J Appl Phys* 47:4310
69. Waldman M, Hagler AT (1993) *J Comput Chem* 14:1077
70. Chen X, Carbone P, Cavalcanti WL, Milano G, Müller-Plathe F (2007) *Macromolecules* 40:8087
71. Izvekov S, Voth GA (2006) *J Chem Phys* 125:151101
72. Leon S, van der Vegt NFA, Delle Site L, Kremer K (2005) *Macromolecules* 38:8078
73. Harmandaris VA, Kremer K (2009) *Macromolecules* 42:791
74. Sun Q, Faller R (2006) *J Chem Theory Comput* 2:607
75. Villa A, van der Vegt NFA, Peter C (2009) *Phys Chem Chem Phys* 11:2068
76. Villa A, Peter C, van der Vegt NFA (2009) *Phys Chem Chem Phys* 11:2077
77. Fritz D, Harmandaris VA, Kremer K, van der Vegt NFA (2009) *Macromolecules* 42:7579
78. Izvekov S, Voth GA (2005) *J Chem Phys B* 109:2469
79. Izvekov S, Voth GA (2005) *J Chem Phys* 123:134105
80. Noid WG, Chu J-W, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC (2008) *J Chem Phys* 128:244114
81. Noid WG, Chu JW, Ayton GS, Voth GA (2007) *J Phys Chem B* 111:4116
82. Noid WG, Liu P, Wang Y, Chu J-W, Ayton GS, Izvekov S, Andersen HC, Voth GA (2008) *J Chem Phys* 128:244115
83. Liu P, Voth GA (2007) *J Chem Phys* 126:045106
84. Zhou J, Thorpe IF, Izvekov S, Voth GA (2007) *Biophys J* 92:4289
85. Izvekov S, Voth GA (2006) *J Chem Theory Comput* 2:637
86. Wang Y, Izvekov S, Yan T, Voth GA (2006) *J Phys Chem B* 110:3564
87. Wang Y, Voth GA (2005) *J Am Chem Soc* 127:12192
88. Izvekov S, Viola A, Voth GA (2005) *J Phys Chem B* 109:17019
89. Shi Q, Izvekov S, Voth GA (2006) *J Phys Chem B* 110:15045
90. Larini L, Lu L, Voth GA (2010) *J Chem Phys* 132:164107
91. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) *J Phys Chem B* 111:7812
92. Baron R, Trzesniak D, de Vries AH, Elsener A, Marrink SJ, van Gunsteren WF (2007) *Chem Phys Chem* 8:452
93. Monticelli L, Kandasamy S, Periole X, Larson RG, Tieleman DP, Marrink SJ (2008) *J Chem Theory Comp* 4:819
94. Marrink SJ, Mark AE (2004) *Biophys J* 87:3894
95. Lee H, Vries AH, Marrink S-J, Pastor RW (2009) *J Phys Chem B* 113:13186
96. Rossi G, Monticelli L, Puisto SR, Vattulainen I, Ala-Nissila T (2011) *Soft Matter* 7:698
97. Lyubartsev A, Mirzoev A, Chen L, Laaksonen A (2010) *Faraday Discuss* 144:43
98. Lyubartsev AP, Laaksonen A (1995) *Phys Rev E* 52:3730
99. Lyubartsev AP, Laaksonen A (2004) On the reduction of molecular degrees of freedom in computer simulations. In: Karttunen M, Vattulainen I, Lukkarinen A (eds) *Novel methods in soft matter simulations*. Springer, Berlin, Heidelberg, pp 219–244
100. Schneider J, Hess W, Klein R (1985) *J Phys A* 18:1221
101. Jiang W, Wang YT, Yan TY, Voth GA (2008) *J Phys Chem C* 112:1132
102. Abrams CF, Delle Site L, Kremer K (2003) *Phys Rev E* 67:021807
103. Delle Site L, Leon S, Kremer K (2004) *J Am Chem Soc* 126:2944
104. Praprotnik M, Delle Site L, Kremer K (2005) *J Chem Phys* 123:224106

105. Praprotnik M, Matysiak S, Delle Site L, Kremer K, Clementi C (2007) *J Phys Condens Mater* 19:292201
106. Praprotnik M, Delle Site L, Kremer K (2007) *J Chem Phys* 126:134902
107. Ensing B, Nielsen SO, Moore PB, Klein ML, Parrinello M (2007) *J Chem Theory Comp* 3:1100
108. Poblete S, Praprotnik M, Kremer K, Site LD (2010) *J Chem Phys* 132:114101
109. Matysiak S, Clementi C, Praprotnik M, Kremer K, Site LD (2008) *J Chem Phys* 128:024503
110. Poma AB, Site LD (2010) *Phys Rev Lett* 104:250201
111. Junghans C, Praprotnik M, Site LD (2009) In: Grotendorst J, Attig N, Blügel S, Marx D (eds) *Multiscale simulation methods in molecular sciences*, NIC Series vol 42. Institute for Advanced Simulation, Forschungszentrum Jülich, Germany, p 359
112. Delgado-Buscalioni R, Kremer K, Praprotnik M (2008) *J Chem Phys* 128:114110
113. Violi A (2009) In: Voth GA (ed) *Coarse-graining of condensed phase and biomolecular systems*. CRC, Boca Raton, pp 433–448
114. Violi A (2004) *Combust Flame* 139:279
115. Santangelo G, Di Matteo A, Müller-Plathe F, Milano G (2007) *J Phys Chem B* 111:2765
116. Hess B, Leon S, van der Vegt NFA, Kremer K (2006) *Soft Matter* 2:409
117. Chen X, Carbone P, Santangelo G, Di Matteo A, Milano G, Müller-Plathe F (2009) *Phys Chem Chem Phys* 11:1977
118. Hess B, Peter C, Ozal T, van der Vegt NFA (2008) *Macromolecules* 41:2283
119. Ozal T, Peter C, Hess B, van der Vegt NFA (2008) *Macromolecules* 41:5055
120. Hess B, van der Vegt NFA (2008) *Macromolecules* 41:7281
121. Merkel TC, Bondar VI, Nagai K, Freeman BD, Pinnau I (2000) *J Polym Sci B Polym Phys* 38:415
122. Kormsmeier RW, Meerwall EV, Peppas NA (1986) *J Polym Sci B Polym Phys* 24:409

Index

A

Ab initio MD simulations 109
Adaptive partitioning (AP) 50
Adiabatic approximation 6
Adrenergic receptor 16
Aluminum chloride 140
Always stable predictor–corrector (ASPC)
109, 120
Ammonia 139, 231
Angle bending 211
Assisted model building with energy
refinement (AMBER) 222
Associated liquids 109
Atomic orbital basis methods 123
Atomistic simulation 251, 295
Atoms in molecules (AIM) 215

B

Auger decay 23
B3LYP 7, 79
Basis sets 109, 122
truncation errors (BSSE) 79
Benzene, alkylation 93
Bis(diphenylphosphino)-ethane 142
Bond order dependent valence terms 17
Bond stretching 210
Born–Oppenheimer approximation 5, 73, 111
MD (BOMD) simulations 109, 114
Bose gas 11

C

Carbon 140
Carbon dioxide, binary mixtures 235
Carbon monoxide 145
Carbon nanotubes (CNTs) 18
nucleation/growth 1
Car–Parrinello MD simulations 109, 115

Catalysis 109
Cell multipole method (CMM) 12
Chemical reactor 93
Chemistry at Harvard molecular mechanics
(CHARMM) 222
Cholesky decomposition (CD) 76
Chrysene 196
Coarse-grain (CG) models 28, 251, 295
Collective variable 109
Complete basis set (CBS) models 78
Complete-active-space second-order
perturbation (CASPT2) 77
Computational chemistry 295
Computational heterogeneous catalysis 69
Condensed-phase optimized molecular
potentials for atomistic simulation
studies (COMPASS) 222
Conductor-like screening model
(COSMO) 216
Configurational-bias Monte Carlo
(CBMC) 183
Confinement 157
Continuum thermodynamics 157
Coupled cluster (CC) 5

D

Degrees of freedom (DOF) 27
Density fitting (DF) 76
Density functional theory (DFT) 5
DFT plus dispersion (DFT-D) 80
Diamond 140
nanoparticle, Auger decay 25
Diffusion coefficients 229
1,3-Dimethylimidazolium chloride 136, 139
Dimethylsulfoxide (DMSO) 140
Discrete variable representation 109, 122
Dispersion 109
Dissipative particle dynamics (DPD) 279, 307

Distributed multipole analysis (DMA) 216
 DNA 1, 28, 171, 277, 286
 meso-scale dynamics 28
 Dynamic partitioning 49

E

Effective core potential 109, 127
 Electrical conductivity 229
 Electrocapillarity 158
 Electrochemistry 143
 Electron force field (eFF) 17
 Electron localization function (ELF) 137
 Electrostatic multipole moments 216
 Electrowetting, nanoscale 155
 Electrowetting on dielectric (EWOD) 156
 Enhanced sampling 109, 130
 Equations of motion (EOM) 4
 1-Ethyl-3-methyl-imidazolium nitrate [emim]
 [NO₃] 58
 Ethylbenzene 298
 Euler–Lagrange relation 112
 EXAFS 133, 135
 Excess coordination numbers 266
 Extended system dynamics 89
 Extracellular matrix (ECM) 33

F

Fermi gas 11
 Fictitious mass 109
 Finite extensible nonlinear elastic potential
 (FENE) 297
 First-principles-based multiscale-
 multiparadigm simulation 3
 First-principles molecular dynamics
 simulations 109
 Force fields (FF) 10, 201, 203
 parameterization 213
 transferable 218
 Force matching method 312
 Formamide 139
 Free energy surface (FES) 109, 131

G

G protein-coupled receptors (GPCRs) 13
 Gas chromatography 181
 Gas–liquid chromatography (GLC) 185
 Gas–liquid interactions 51
 Gas–liquid scattering 43
 Gas–solid chromatography (GSC) 185
 Gaussian and plane waves (GPW) 123
 Gaussian-type Geminals (GTG) 78

Generalized gradient approximation (GGA)
 7, 79
 Generalized valence bond (GVB) wave
 functions 5
 Gibbs ensemble Monte Carlo (GEMC) 183
 GIBBS99 force field 222
 Graphite 140
 Groningen molecular simulation
 (GROMOS) 222

H

Hard sphere (HS) 204
 Hartree–Fock (HF) 6, 74, 109
 Hybrid QM/MM molecular dynamics 43
 Hydrogen, electro-oxidation 145
 Hydrogen bonds 209
 interfacial 155
 network, polarized water 164
 Hydrogen fluoride 139
 Hydrophobic confinement 159
 Hyperthermal desorption/scattering (HDS) 53

I

Ideal adsorbed solution theory (IAST) 97
 Interfacial chemistry 43
 Ionic liquids 109, 138
 IR 133
 Iterative Boltzmann inversion 305

K

Kinetic energy pressure 22
 Kinetic Monte Carlo (kMC) 85
 Kirkwood–Buff (KB) theory 265
 Kohn–Sham orbitals (KS) 79, 116
 Kohn–Sham SCF scheme 74
 Kovats retention index 185

L

Langevin thermostat 88
 Lennard-Jones (LJ) potential 205
 Linear scaling 109
 Liouville equation 87
 Liquid, confinement 157
 Liquid metals/alloys 140
 Lithium cluster impacts 24
 Local density approximation (LDA) 7, 79
 Local spin density approximation (LSDA) 79
 Lowe–Andersen interface-fluid collision
 thermostat (LA-IFC) 89

M

Marcus theory 144
MARTINI force field 312
Maximally localized Wannier functions (MLWF) 136
Message passing interface (MPI) 89
Metadynamics 109, 131
Metal-free organic reactions 141
Metal–liquid interfaces 143
Methanol 138
N-Methylacetamide (NMA) 134, 138, 140
Micro-canonical ensemble (NVE) 88
Minimum energy path (MEP) 131
Minimum free energy paths (MFEP) 131
Model liquid surface 51
Molecular dynamics (MD) 2, 44, 69, 81, 112, 201
Molecular mechanics (MM) 11, 43
Molecular orbitals (MO) 6
Molecular simulation 181, 251
Molecular thermodynamics 159
Møller–Plesset perturbation theory (MP) 75
Monte Carlo 18, 69, 81, 181, 201, 226, 280, 296
 inverse 314
 kinetic 315
Multiple time scale (MTS) 88
Multireference configuration interaction (MRCI) 6, 77
Multireference self-consistent field (MC-SCF) 6
Multiscale modeling 1, 69, 295
 QM 4

N

Nanofluidics 155
Nanoparticle ordering 155
Nanotube growth 1
Nath, Escobedo, and de Pablo (NERD) force field 218, 221
Newton inversion method 313
NMR 133
Non-adiabatic molecular dynamics 1
Nucleation 20
Nudged elastic band (NEB) 91, 109, 131
Numerically tabulated atom-centered orbitals (NAOs) 109, 124

O

Optimized potential model for phase equilibria (OPPE) 218, 220

Optimized potentials for liquid simulations (OPLS) 218, 219
Organometallic structures 1
Osmotic coefficients 263
Oxygen, electroreduction 145

P

Partitioned rational function optimisation (P-RFO) 71
Phase equilibria 227
Phosphorus 140
Planck's quantum 4
Plane waves 109
Poisson–Boltzmann (PB) model 8
Polarizable continuum model (PCM) 216
Polarization 208
Polyamide–66 (PA–66) 297
Polycyclic aromatic hydrocarbons (PAHs) 196
Polyelectrolyte multilayers (PEMs) 279
Polyelectrolytes 251
Polymer-based hydrogel networks 33
Polyphenylene dendrimer 298
Polystyrene (PS) 270, 298
Polystyrene sulfonate (PSS) 275, 283
Poly(diallyldimethylammonium) (PDADMA) 283
Poly(methyl methacrylate) (PMMA) 297
Potential energy surface (PES) 5, 71
Potential of mean force (PMF) 264
Protein structure prediction 1
Pseudopotential 109

Q

QM/MM 89
Quadratic configuration interaction method (QCISD) 75
Quantum 251
Quantum chemistry 69, 73
Quantum mechanics (QM) 2, 43, 214
Quantum Monte Carlo (QMC) 5

R

Radial distribution functions (RDFs) 299
Random phase approximation (RPA) 126
Rare event 109
Rayleigh–Schrödinger perturbation theory 75
Reactive dynamics (RD) 19
Reactive molecular dynamics 1
Reactivity, solvent/pH effects 8
ReaxFF 17, 85

Redox reactions, solution 144
Reduced density matrix (RDM) 77
Reference system propagator algorithm
(r-RESPA) 88
Relativistic electronic structure 109
Resolution of the identity (RI) 76
Restrained electrostatic potential (RESP) 215
Retention mechanism 181
Retention potential 109
Reversed-phase liquid chromatography
(RPLC) 181, 187
Room temperature ionic liquids (RTIL) 43, 58
Roothaan equations 74, 80
Rubredoxin 144

S

Schrödinger's equation 6, 73
Self consistent field (SCF) 109, 111
Self consistent reaction field (SCRF) 216
SHAKE algorithm 109
Shear viscosity 229
Shock Hugoniot 23
Soft-sphere (SS) potential 204
Solar cells, dye-sensitized 146
Solvent effects 109, 140
Solvent-separated ion pair (2SIP) 264
Squalane 43, 53
Stochastic rotational dynamics (SDR) 279
String method 109
SuperComBiHelix 15
Sutherland potential 205

T

Taylor expansion 113
Thermal conductivity 229
Thermodynamics 201
Tight binding (TB) methods 18
Tissue engineering 33

Torsional rotation terms 211
Transferable potential 17
Transferable potential for phase equilibria
(TraPPE) 218, 220
Transition state theory (TST) 70, 89
Transmembrane protein structures 1
Transport properties 229
Trapped desorption (TD) 53
Truncated and shifted Lennard-Jones potential
(TSLJ) 205
Two-electron reduced density matrix
(2-RDM) 77

U

Uracil 140

V

Valence coordinate cross terms 212
van der Waals interaction 109, 127, 204, 214
van der Waals term 11
Variational transition-state theory (VTST) 71

W

Wannier orbitals 109
Water 109, 138
 confined 155
Water-mediated ordering 170
Wave equation 4
Wave packet molecular dynamics (WPMD) 22
Wavefunction analysis 136
Wavelets 109, 123
Wetting, nanoscale 155
Wetting free energy 168

X

X3LYP 7