Molecular Modeling and Simulation
Applications and Perspectives

Randall Q. Snurr
Claire S. Adjiman
David A. Kofke *Editors*

Foundations of Molecular Modeling and Simulation

Select Papers from FOMMS 2015

Springer

Molecular Modeling and Simulation

Applications and Perspectives

Series editor

Edward Maginn, University of Notre Dame, Notre Dame, IN, USA

This series aims at providing a comprehensive collection of works on developments in molecular modeling and simulation, particularly as applied to the various research fields of engineering. The Series covers a broad range of topics related to modeling matter at the atomistic level. The series provides timely and detailed treatment of advanced methods and their application in a broad range of interrelated fields such as biomedical and biochemical engineering, chemical engineering, chemistry, molecular biology, mechanical engineering and materials science. The Series accepts both edited and authored works, including textbooks, monographs, reference works, and professional books. The series welcomes manuscripts concerned with developments in and applications of molecular modeling and simulation to contemporary research in myriad technological fields, including, but not limited to:

- New Materials Development
- Process Engineering
- Fuel Engineering
- Combustion
- Polymer Engineering
- Biomechanics
- Biomaterials
- Fluid Flow and Modeling
- Nano and Micro Fluidics
- Nano and Micro Technology
- Thin Films
- Phase Equilibria
- Transport Properties
- Computational Biology

More information about this series at http://www.springer.com/series/13829

Randall Q. Snurr · Claire S. Adjiman
David A. Kofke
Editors

Foundations of Molecular Modeling and Simulation

Select Papers from FOMMS 2015

Springer

Editors
Randall Q. Snurr
Department of Chemical and Biological Engineering
Northwestern University
Evanston, IL
USA

David A. Kofke
The State University of New York
Buffalo, NY
USA

Claire S. Adjiman
Department of Chemical Engineering
Imperial College London
London
UK

ISSN 2364-5083 ISSN 2364-5091 (electronic)
Molecular Modeling and Simulation
ISBN 978-981-10-1126-9 ISBN 978-981-10-1128-3 (eBook)
DOI 10.1007/978-981-10-1128-3

Library of Congress Control Number: 2016939101

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Series Editor's Preface

This is the first volume in the new series *Molecular Modeling and Simulation—Application and Perspectives*. The series aims at providing a comprehensive collection of works on developments in molecular modeling and simulation, particularly as applied to the various research fields of engineering. The goal is to cover a broad range of topics related to modeling matter at the atomistic level and to provide timely and detailed treatment of advanced methods and their application in a broad range of interrelated fields such as biomedical and biochemical engineering, chemical engineering, chemistry, molecular biology, mechanical engineering, and materials science. It is therefore fitting that the first volume contains papers arising from work presented at the 2015 Foundations of Molecular Modeling and Simulation (FOMMS) conference, held July 12–16, 2015 near Mount Hood, Oregon.

I wish to acknowledge the tireless efforts of the FOMMS 2015 conference cochairs Claire S. Adjiman (Imperial College London) and David A. Kofke (University at Buffalo) and conference chair Randall Q. Snurr (Northwestern University), who organized FOMMS 2015 and carried out the editorial duties associated with assembling this volume.

Edward Maginn

Preface

This volume contains ten papers from the 2015 conference on Foundations of Molecular Modeling and Simulation (FOMMS). The theme of this 6th FOMMS conference was Molecular Modeling and the Materials Genome. As in past conferences, the format consisted of invited lectures, contributed posters, and several workshops. A total of 172 people participated in FOMMS 2015, and 116 contributed posters were presented.

The conference began with a keynote address from Frank Stillinger of Princeton University, entitled “Chiral Symmetry Breaking via Computer Simulation.” The theme of the first session was Future Trends in Modeling, Simulation and Data Mining, and the session featured talks by Andrea Browning of Boeing, Alán Aspuru-Guzik of Harvard University, and Jinghai Li of the Chinese Academy of Sciences. The session on Biomaterials and Biological Systems consisted of talks from Sabrina Pricl of the University of Trieste and Yiannis Kaznessis of the University of Minnesota. Chris Wolverton of Northwestern University, Kristen Fichthorn of Penn State University, and Jonathan Moore of Dow Chemical spoke in the session on Energy and Environmental Applications, and the session on Complex Fluids and Materials featured talks by Edward Maginn of the University of Notre Dame, Coray Colina of Penn State University, and Marjolein Dijkstra of Utrecht University. Talks by Joachim Sauer of Humboldt University, Daniela Kohen of Carleton College, and Jeffrey Errington of the University at Buffalo were the focus of the session on Catalysis and Interfaces. The session on Reactive Force Fields featured presentations by Susan Sinnott of the University of Florida and Adri van Duin of Penn State University. The conference ended with the awarding of the FOMMS Medal to Carol Hall of North Carolina State University, who gave a memorable talk entitled “Protein Aggregation Simulations: Lessons Learned Over a Decade.”

The conference also featured three workshops. The first workshop on Data Mining, Machine Learning, and Materials Informatics was given by Jonathan Moore of Dow Chemical and Johannes Hachmann of the University at Buffalo. Joshua Anderson of the University of Michigan put on a workshop entitled “Using

GPUs for Bigger and Faster Simulations," and the final workshop, entitled "Solving Common Software Problems in Computational Labs," was led by Patrick Fuller of NuMat Technologies and Christopher Wilmer of the University of Pittsburgh.

The principal sponsor of FOMMS 2015 was the CACHE Corporation, with financial support coming from the Computational Molecular Science and Engineering Forum of the American Institute of Chemical Engineers, ExxonMobil, Imperial College London, the Journal of Physical Chemistry, Materials Design, the National Institute of Standards and Technology, the National Science Foundation, Northwestern University, Procter and Gamble, the Royal Society of Chemistry, Scienomics, Springer, the University of Minnesota Nanoporous Materials Genome Center, and UOP.

The ten papers in this volume represent the wide range of molecular modeling tools and applications discussed at the conference. The first paper, by Shao and Hall, presents a coarse-grained model that accounts for protein–protein interactions in a multiprotein system using discontinuous molecular dynamics simulations. The model should set the stage for simulating protein systems on longer timescales and deepening our understanding of processes such as protein crystallization, protein recognition, and protein purification. In the second paper, Sprenger et al. describe their use of molecular dynamics simulations with enhanced sampling methods to study how two types of defects in self-assembled monolayers affect the structure of adsorbed peptides. Moore et al. present the development of a coarse-grained force field for water via multistate iterative Boltzmann inversion. The model is derived to match the bulk and interfacial properties of liquid water. Hülsmann et al. discuss strategies and software for the semi- or fully-automated parameterization of force fields, including options for intramolecular and intermolecular interactions and a work flow combining global and local optimization procedures. In another paper focused on software and methods, Klein et al. describe open-source software called mBuild, which is a general tool designed to simplify the construction of complex, regular, and irregular structures for molecular simulation. Basic molecular components are connected using an equivalence operator which reduces and often removes the need for users to explicitly rotate and translate components as they assemble complex systems. In a methods-oriented contribution bridging quantum and classical mechanics, Subramanian et al. examine the Path Integral Monte Carlo performed with "semi-classical beads." They compare the rate of convergence with respect to the number and type of beads for computing fully quantum virial coefficients of helium-4.

Turning more toward applications, the paper by He et al. describes molecular simulations of the homogeneous nucleation of the ionic liquid [dmim$^+$][Cl$^-$] from its bulk supercooled liquid. Their work combines the string method in collective variables, Markovian milestoning with Voronoi tessellations, and order parameters for molecular crystals. Results include the free-energy barrier, the critical nucleus size, and the nucleation rate. Schweizer et al. study the influence of alloy composition on the structure of Raney nickel catalysts using molecular dynamics simulations and the competitive adsorption of benzene and cyclohexane on Raney nickel as a first step toward modeling the catalytic hydrogenation of benzene. Norman et al. present atomistic modeling related to hydrocarbon mixtures and gas hydrates in porous media, including molecular dynamics simulations to study the phase diagrams of hydrocarbon mixtures in the bulk and in confined geometries. Finally, Bamberger and Kohen report a combination of grand canonical Monte Carlo and MD simulations that provide new insight into an intriguing "cation gating" that allows carbon dioxide but not other adsorbates to permeate Na—Rho zeolites.

We thank all of the participants for their contributions to FOMMS 2015 and especially the authors and reviewers of the papers in this volume. Special thanks goes to the conference facilitator, Robin Craven; the Senior Advisors of FOMMS 2015, Peter Cummings, Joe Golab, Clare McCabe, Jonathan Moore, and J. Ilja Siepmann; and the conference Programming Committee.

Randall Q. Snurr
Claire S. Adjiman
David A. Kofke

Contents

About the Editors

Randall Q. Snurr is the John G. Searle Professor of Chemical and Biological Engineering at Northwestern University. He holds BSE and Ph.D. degrees in chemical engineering from the University of Pennsylvania and the University of California, Berkeley, respectively, and performed postdoctoral research at the University of Leipzig supported by a fellowship from the Alexander von Humboldt Foundation. Other honors include the Institute Award for Excellence in Industrial Gases Technology from the American Institute of Chemical Engineers, the Leibniz professorship at the University of Leipzig, and a CAREER award from the National Science Foundation. He was named a Highly Cited Researcher for the period 2002–2012 by Thomson Reuters. He was a senior editor of the Journal of Physical Chemistry and currently serves on the editorial boards of several journals. His research interests include development of new nanoporous materials for energy and environmental applications, molecular simulation, adsorption separations, diffusion in nanoporous materials, and catalysis.

Claire S. Adjiman is professor of chemical engineering at Imperial College London. She holds an MEng from Imperial College and a Ph.D. from Princeton University, both in chemical engineering. Her research interests lie in the area of integrated process and molecular/materials design, including the development of design methods, property prediction techniques, and optimization algorithms. She is the recipient of several prizes including a RAEng-ICI Fellowship (1998–2003), the Philip Leverhulme Prize for Engineering (2009), and the SCI Armstrong Lecture (2011). She holds an EPSRC Leadership Fellowship (2012–2017) and was elected Fellow of the IChemE in 2013. In 2011, she co-edited a book on Molecular Systems Engineering published by Wiley-VCH.

David A. Kofke received his B.S. in chemical engineering from Carnegie Mellon University and Ph.D. from the University of Pennsylvania, advised by Eduardo Glandt. Since 1989, he has been on the chemical engineering faculty of the University at Buffalo (SUNY), where he served as department chair for 6 years, and now holds the rank of SUNY Distinguished Professor. Author of over 130 refereed publications, Kofke's research currently focuses on rigorous molecular-based free-energy calculations for crystal structure prediction, and calculation of virial coefficients and other cluster integrals from molecular models. Among other awards, Kofke is the recipient of the triennial John M. Prausnitz Award for applied chemical thermodynamics, the Jacob F. Schoellkopf Medal, and the Himmelblau Award from the CAST division of AIChE. Prof. Kofke is a member since 1999 of the Board of Trustees of CACHE, where he served as President in 2010–2012. He is a Fellow of AIChE and AAAS.

A Discontinuous Potential Model for Protein–Protein Interactions

Qing Shao and Carol K. Hall

Abstract Protein–protein interactions play an important role in many biologic and industrial processes. In this work, we develop a two-bead-per-residue model that enables us to account for protein–protein interactions in a multi-protein system using discontinuous molecular dynamics simulations. This model deploys discontinuous potentials to describe the non-bonded interactions and virtual bonds to keep proteins in their native state. The geometric and energetic parameters are derived from the potentials of mean force between sidechain–sidechain, sidechain–backbone, and backbone–backbone pairs. The energetic parameters are scaled with the aim of matching the second virial coefficient of lysozyme reported in experiment. We also investigate the performance of several bond-building strategies.

Keywords Coarse-grained model · Protein–protein interactions · Discontinuous molecular dynamics · Square-well potential · Osmotic second virial coefficient

1 Introduction

Here, we report the development of a two-bead-per-residue protein model that can be used with discontinuous molecular dynamics (DMD) simulations to investigate protein–protein interactions in a multi-protein system. We expect that this new model will allow us to simulate multi-protein systems on longer timescales than what has heretofore been achievable, helping us to deepen our understanding of processes such as protein crystallization [1], protein recognition [2], and protein purification [3].

Protein models can be classified broadly into two types: all-atom if they describe every atom in the protein explicitly and coarse-grained if they group several atoms into one interactive site. All-atom force fields such as CHARMM [4], AMBER [5],

---

Q. Shao · C.K. Hall (✉)
Department of Chemical and Biomolecular Engineering,
North Carolina State University, Raleigh 27695, USA
e-mail: hall@ncsu.edu

R.Q. Snurr et al. (eds.), *Foundations of Molecular Modeling and Simulation*,
Molecular Modeling and Simulation, DOI 10.1007/978-981-10-1128-3_1

GROMOS [6, 7], and OPLS/AA [8] are very good at describing the behavior of a single protein and how it interacts with other molecules in explicit solvent. However, atomistic simulations are usually limited to one or several small proteins and timescales up to several hundred nanoseconds, effectively precluding the investigation of many interesting multi-protein problems. Coarse-grained models enable us to simulate larger systems for longer timescales using less computational resources. There are two major choices to be made in the development of coarse-grained models: (1) how to coarse-grain the protein geometry and (2) how to obtain the geometric and energetic parameters (see recent review papers [9–15] that summarize the various coarse-graining methods, coarse-grained protein models, and their applications). Coarse-grained protein models can be categorized based on how the atoms are grouped together to form the coarse-grained bead (four-bead-per-residue [16], two-bead-per-residue [17], one-bead-per-residue [18, 19], and ultra-coarse-grained [20]) and how the force field parameters are determined (e.g., Go-type [21], knowledge-based [22, 23], and physics-based [24]).

Coarse-grained models are usually more problem-specific than all-atom models because of transferability issues. Coarse-grained protein models are often developed with the goal of examining particular properties. Most of the current coarse-grained protein models focus on the folding/unfolding problem. It thus remains to be seen how well protein models developed based on such properties do in describing behavior that is a consequence of protein–protein interactions. For example, Stark et al. [25] found that the popular MARTINI force field predicts a second virial coefficient of lysozyme that differs considerably from the experimental value. This inconsistency between simulation and experiment points out the importance of developing protein models that are designed to apply to problems where protein–protein interactions play a major role.

It is also important that the method used to simulate multi-protein systems be fast. Most of the models used in simulating multi-protein systems are based on continuous intermolecular potentials like the Lennard–Jones potential. Simulations based on continuous potentials proceed by solving Newton equations at a uniformly spaced time intervals. They have an algorithm complexity of $O(N\log N)$, where N is the number of particles in the system. The big-O notation describes how the performance or complexity (referring to the number of operations) required to run an algorithm depends on the number of particles in the system. Therefore, the required computational time for continuous MD simulations increases dramatically with the number of beads in the system, limiting their application to relatively small systems.

Discontinuous molecular dynamics (DMD) simulations can be used to investigate large systems efficiently with moderate computational resources. DMD simulations were designed to be applicable to systems that interact via discontinuous potentials (square-well/square-shoulder and hard-sphere). They proceed by analytically calculating the next collision time. Several papers [26–28] describe the details of DMD simulations. The algorithm complexity of DMD simulations is $O(N\log N)$. (One paper by Paul [29] even claims a realization of the DMD method with an algorithm complexity of $O(1)$.) The enhanced algorithm complexity of DMD simulations compared to continuous MD simulations make them suitable for the investigation of long-time processes like spontaneous formation of amyloid structure, which are still challenging for continuous MD simulations.

This work reports our effort to develop a coarse-grained protein model that can be used to study protein–protein interactions in multi-protein systems via DMD simulations. We deploy a two-bead-per-residue protein model: one bead for the backbone and the other for the sidechain. The parameters of our protein model are obtained by coarse-graining atomistic simulation results for backbone–backbone, backbone–sidechain, and sidechain–sidechain interactions in explicit water. The rest of the paper is organized as follows. Section 2 describes the protein model in detail; Sect. 3 describes the atomistic and DMD simulations; Sect. 4 discusses the analysis leading to the final choice of model parameters; and Sect. 5 summarizes the current status of the model.

2 Model Description

We deploy a two-bead-per-residue protein model to represent the 20 natural amino acid residues. Since computational efficiency was a major consideration here, we tried to minimize the number of beads in the system and at the same time represent the chemical heterogeneity of the individual amino acid residues. Although a one-bead-per-residue model minimizes the number of beads in the system, we found that it made it harder to represent the difference among the various types of amino acid residue in DMD simulations. Protein models with more than two beads per residue do a good job of representing the chemical heterogeneity of the 20 residues (see, e.g., our protein model, PRIME20 [16]), but this increases the required computational resources. The two-bead-per-residue model is a good compromise for large proteins.

The 18 amino acid residues except glycine and proline are represented by two beads: one bead at the position of the C-α atom to represent the backbone entity and the other at the sidechain center of mass to represent the sidechain entity. Glycine and proline residues are represented solely by a single bead at the positions of their C-α atoms because either they do not have a sidechain or the sidechain is closely linked with the backbone. Figure 1 shows a schematic of the two-bead model for alanine.

The potential energy of the system is the sum of the intermolecular potential energy, intramolecular potential energy, and virtual bond energy for all the beads in the system (Eq. 1).

$$U_{\text{total}} = \sum U_{\text{inter}}(r) + \sum U_{\text{intra}}(r) + \sum U_{\text{bond}}(r) \tag{1}$$

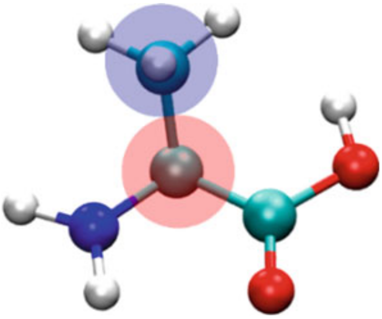

Fig. 1 Schematic of the two-bead-per-residue model. One bead is at Cα, and the other is at the center of mass of the sidechain

The intermolecular bead–bead interactions are represented by a single square well or single square shoulder potential as given in Eq. (2):

$$\begin{cases} U_{\text{inter}}(r) = \infty, & r \leq \sigma_1 \\ U_{\text{inter}}(r) = \epsilon, & \sigma_1 < r < \sigma_2 \\ U_{\text{inter}}(r) = 0, & r \geq \sigma_2 \end{cases} \tag{2}$$

where r is the bead–bead distance, σ_1 and σ_2 are geometric parameters, and ε is the energetic parameter. The geometric and energetic parameters (σ_1, σ_2, and ε) are derived from the potentials of mean force (PMFs) of sidechain–sidechain, sidechain–backbone, and backbone–backbone pairs from atomistic simulations in explicit water solvent as discussed in Sect. 4. A single square-well potential ($\varepsilon < 0$) indicates that the two entities attract each other in explicit water; a single square-shoulder potential ($\varepsilon > 0$) indicates that these two entities repel each other in explicit water; and a hard-sphere potential ($\varepsilon = 0$) indicates that the two entities just have an excluded volume interaction in water. The effect of water is taken into account in the parameters because the PMFs were obtained from the pair's interactions in explicit water solvent.

The intramolecular bead–bead non-bonded interactions consider excluded volume effects only. The hard-sphere potential is used to describe the intramolecular bead–bead non-bonded interactions (Eq. 3).

$$\begin{cases} U_{\text{intra}}(r) = \infty, & r \leq 0.8\sigma_1 \\ U_{\text{intra}}(r) = 0, & r > 0.8\sigma_1 \end{cases} \tag{3}$$

where r is the bead–bead distance and σ_1 is the geometric parameter in Eq. (2). The geometric parameters could, in principle, be obtained from the volumes of the individual beads, but to simplify the process, we choose to use $0.8\sigma_1$ as the geometric parameter. We found that this selection avoids overlap between beads in a protein and works well with the virtual bond setting, which is described in the next paragraph.

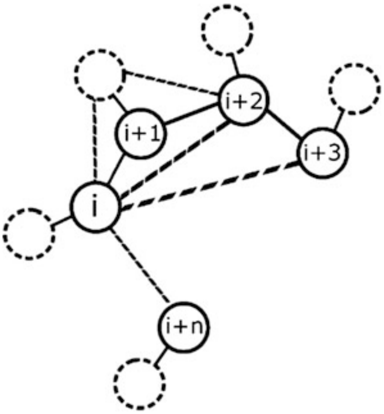

Fig. 2 Schematic describing virtual bonds. The *circles with solid borders* are backbone beads, and the *circles with dash-line borders* are sidechain beads. The virtual bonds connect these beads to keep the protein in its native state

We deploy virtual bead–bead bonds to maintain the protein in its native state. The virtual bond potential is a double hard wall (Eq. 4).

$$\begin{cases} U_{\text{bond}}(r) = \infty, & r \leq (1-\delta)\sigma \\ U_{\text{bond}}(r) = 0, & (1-\delta)\sigma < r < (1+\delta)\sigma \\ U_{\text{bond}}(r) = \infty, & r \geq (1+\delta)\sigma \end{cases} \tag{4}$$

where r is the bead–bead distance, σ is an equilibrium bead–bead distance obtained from the native state of the protein, and δ is the flexibility factor. Figure 2 shows a schematic describing the virtual bonds. The native state of a protein is its naturally folded structure. Here, we use the structure of a protein in the Protein Data Bank (PDB) as its native state. The virtual bonds can be divided into two categories depending on the indices of the connected beads along the amino acid sequence. The “local” category includes virtual bonds between beads whose index difference is less than four. They are used to maintain the protein local secondary structure. The other category (non-local) includes virtual bonds between beads far away from each other along the amino acid sequence. These bonds are used to maintain the tertiary and quaternary structures of a protein. Section 4.2 discusses the choice of the virtual bonds in detail.

3 Simulation Details

3.1 Atomistic Simulation

We conducted atomistic simulations to obtain the geometric and energetic parameters for the coarse-grained beads in the two-bead-per-residue model; these parameters are then used in the DMD simulations. The sidechain and backbone entities were generated from amino acid residues. Glycine and proline entities were generated by capping their N and C terminals with an acetyl group and an N-methyl