THE UNIVERSITY OF CHICAGO

BASIS SETS AND OPTIMIZATION FOR COARSE-GRAINED MODELS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTY

BY

THOMAS PHILIP DANNENHOFFER-LAFAGE

CHICAGO, ILLINOIS

JUNE 2018

Copyright © 2018 by Thomas Dannenhoffer-Lafage

All rights reserved

For my wife Camille and my mother Ann,

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank Professor Gregory A. Voth for serving as my research advisor. His support made it possible for me to study interesting problems in coarse-graining and his direction and insight were instrumental to the final form that my research took. He also created a stellar environment that brought together many of the most intelligent and creative people I have ever met, which helped me immensely. Also, I would like to thank Professor Suri Vaikuntanathan and Professor Timothy Berkelbach for serving as committee member for the defense of this thesis.

Next, I would like to thank those who served as mentors to me during the time I worked on this thesis. In particular, I would like to thank Professor Andrew White and Dr. Jacob Wagner, who helped me greatly to gain the skills necessary to complete the work presented. Also, I would like to thank Professor Marina Guenza, who first introduced me to computational research while I was at the University of Oregon.

Additionally, I would like to thank those who were collaborators on the research presented in this thesis, and the other work I completed while at the University of Chicago: Professor Andrew White, Dr. Jacob Wagner, Dr. Glen Hocky, Dr. Alexander Pak, Jaehyeok Jin, Aleksander Durumeric, Paul Calio, and Professor Gregory A. Voth.

Also, I must thank all the members of the Voth group who helped me during my time at the University of Chicago. Specifically, I would like to thank those who I had productive conversations with about the work presented and the field of science in general: Morris Cohen, Zack Jarin, Aleksander Durumeric, Paul Calio, Dr. Jacob Wagner, Dr. Glen Hocky, Dr. Alexander Pak, Dr. Tamara Bidone, Dr. Jesper Madsen, and Dr. Rui Sun.

Lastly, but certainly not least, I would like to thank everyone who gave the very necessary emotional support to reach this point. In particular, I would like to thank my cohorts at the University of Chicago, whose camaraderie helped me realize that my problems were not unique. Importantly, I would like to thank my wife Camille and my mother Ann, who encouraged me to be tenacious and creative, and always believed in my success.

Abstract

Coarse-Grained (CG) models provide a promising direction to study variety of chemical systems at a reduced computational cost. CG model are generated by reducing the representation of a molecular system from atoms to beads. However, how these models are parameterized can greatly affect the reliability and the insight that could be provided by CG models. In my thesis, work is presented on different parameterization schemes and basis sets that can be utilized to produce CG models. First, the affect of parameterizing models with the Experiment Directed Simulation (EDS) methodology is explored theoretically and practically. This provides a foundation for top-down information to be incorporated systematically into CG models via EDS. Second, an implementation of the EDS methodology that uses CG variables as targets is presented, called Coarse Grain Directed Simulation. This allows for small part of a much larger system to be modeled in the effective environment of the larger system while only minimally biasing the simulated part of the simulation. Thirdly, a reactive methodology call reactive Multiscale Coarse-Graining is discussed. This takes advantage of a matrix style Hamiltonian that allows for multiple states of a system to be represented, allowing for features such as bond breaking and forming within a coarse-grained simulation based on the free energy of the system. Also, a comparison of Multiscale Coarse-graind (MS-CG) and Relative Entropy Minimization (REM) parameterization methodologies is explored for the case of a CG lipid bilayer within an implicit solvent. This comparison explores the ability for MS-CG and REM to model solvent-solute interaction when the solvent particles have been integrated away, removing the vector of interaction between the solvent and solute particles. Taken together, this work provides the foundation for understanding how different types of information can be taken into account in CG models via these different parameterization schemes and basis sets.

Chapter 1

Introduction

The development of models has been essential to progress of science. The heliocentric model of the solar system with forces determined by gravity was able to reconcile Kepler's laws.[1] The quantum model of the hydrogen atom was used as the foundation for describing heavier atoms and covalent bond.[2-3]

While the theoretical models above were all important to the advancement of science, they were all limited to describing very simple system. While equations exist for describing the motions of very complex systems, the mathematics to solve these equations is still unknown. This has lead to the development of approximate models that aim to solve these equations under certain assumptions. These models have become very popular in the last five decades due to the advent modern computer. While it would take a very long time for a human to track the motion of ten particles, a modern computer may track the motion of millions of particles.[4]

These computer-generated models have lead to advancements in biophysics, medicine, and testing of material properties without the need to physically source or synthesize the material in question. This has made computer modeling an indispensible to development of new drugs and materials.[5-8] However, current computer simulations are still limited by the size of systems and the length of time that can be simulated given current computational resources. Current computational resources allow for simulations of drug molecules interacting with medium sized protein, but it would be impossible to simulate drug interactions in true physiological conditions.[4] One solution to this limitation is coarse-grained (CG) models. CG methods aim to create a CG model that can

be simulated at reduced computational cost by removing "non-essential" features of a system while still remaining predictive.[9-11]

There are two major components that go into the construction of a CG model, the mapping and force-field (FF). The mapping is a rule that determines the relationship between the higher resolution fine-grained (FG) and the CG model. There are two important components that go into deciding a mapping, deciding which atoms should correspond to which CG site and how the each atom within a site should be weighted.[12] While the rational for deciding which atoms should correspond to which sites has largely been based on intuition, there exist methodologies that will attempt to group together atoms that have similar motions.[13-14] Furthermore, some theoretical work has been done showing that the quality of CG model within a given basis set can be severely limited by the choice of map.[15] The choice of weighting for maps has been dominated by two main choices in the literature, center of mass and carbon-alpha. However research has shown that different choices may prove more effective depending on the essential physics of the system.[16] While mappings are surely a very important aspect of CG model building, this thesis will focus the FF aspect of the problem.

The FF determines how CG sites interact with one another. FFs also have two important components that go into their design, the basis set and the optimization. The basis set is the choice of functions that describe the interactions between sites and determine expressiveness between interaction sites. Optimization is how any parameters relating to the basis set should be fit. There are two major obstacles that go into building a FF for CG systems, the time consuming nature of FF optimization and the lack intuitive principles choosing the basis sets used. While one could follow a similar design scheme

of current atomistic FF, the apparent lack of transferability of CG models[17] and the fact that many of the atomistic FFs are almost 20 years old and still receive constant updates leads much to be desired.[18-21] These FFs were also guided by intuitive principles that helped guide what basis set should be picked for the models. The Lennard-Jones interaction was chosen because it has sixth order term that is related to dispersion interactions.[22] Partial point charges have their rationalization in quantum mechanics and can be determined by calculations and experiments.[23] No such intuitive principle exists to suggest that Lennard-Jones or partial point charges are an appropriate basis set for CG interactions. These problems lead for the need of FFs that have more flexible basis sets and can be designed systematically (and hopefully atomically).

Currently, there exist two philosophies for determining the force field optimization of CG models, top-down and bottom–up modeling. Top-down models aim to create a model that reproduces certain macroscopic quantities of a system in question and assumes that the resulting microscopic behavior of the system to be consistent with reality. The quantities that are reproduced typically come from experiment in top-down models. One popular top-down CG model is the MARTINI model.[24] Bottom-up models aim to create a model that reproduces some microscopic quantity of a system and assume that the resulting macroscopic behavior to be consistent with reality. The quantities that are reproduced typically come from a more detailed simulation in bottom-up models. Two methods popular methods for creating bottom-up models are the Relative Entropy Minimizaiton and the Multiscale Coarse-graining methods.[11, 25] It should be noted that these two approaches are not mutually exclusive. While the choice of modeling philosophy doesn't need to dictate the choice of basis set used, top-down models usually

resort to much simpler functional forms with few parameters to facilitate the guess-and-check methodologies that is usually required. Bottom-up methods usually use more complicated and more flexible basis sets since systematic methodologies exist for the optimization of any floating parameters.

While there exist many CG models that have been determined by both the top-down and bottom-up philosophies, many questions still exist about the nature of the force fields that they render. One issue is the choice of basis set when parameterizing a top-down model. When new experimental information becomes that a one would like to incorporate into a model, they run the risk of possibly harming other aspects of the model in the process. Furthermore, there usually isn't a systematic method to incorporate this information. Furthermore, how coarse-grained models should take into account reactivity is still a largely undeveloped. Also, while there has been some theoretical work, how models that are from different bottom-up methods differ in complex cases is still unknown. These questions will be explored in this thesis.

The rest of this thesis is organized as follows: chapter 2 will provide a background on statistical mechanics, coarse-graining, and other methods as they relate to the research presented in the other chapters. Chapter 3 discusses the incorporating top-down information into CG models by Experiment Directed simulations and how the changes can be understood through the Relative Entropy Minimization framework. Chapter 4 discusses how top-down restrictions based on coarse-grained expectation values can be used to simulate a subsection of a large system while only effectively simulating only a subsection of it. Chapter 5 discusses how the use of matrix based Hamiltonian will allow for a CG models to have reactivity within them and a method for parameterizing such

models. Chapter 6 compares a two lipid models that were parameterized by Relative Entropy Minimization and Multiscale Coarse-graining. Finally, chapter 7 will provide conclusion and discuss future directions.

Chapter 2

Statistical Mechanics and Coarse-Graining

The research presented in this thesis builds upon existing knowledge of statistical mechanics,[26] molecular dynamics (MD),[27-28] Coarse-graining (CG),[9-10, 29] and Empirical Valance bond (EVB) methods.[30] While these topics are much too large to cover in a high level detail, a brief overview of these topics as they relate to the research presented will be covered. The research presented also builds upon the Experiment Directed Simulation (EDS) Method,[31] which will also be discussed in this chapter.

2.1 Statistical Mechanics

Let us suppose imagine a many-body system in a particular microscopic state. The time evolution of a system would be described by Schrodinger's equations

$$i\hbar\frac{\partial}{\partial t}\psi = \mathbf{H}\psi \,, \tag{2.1}$$

Where $\hbar$ is the reduced Planck's constant, $\mathbf{H}$ is the Hamiltonian that describes the system, and ψ is the wavefunction for the system. A common assumption in statistical mechanics is that if one observes a system long enough, the trajectory of the system will visit all states accessible to the system. This assumption is typically referred to as the ergodic hypothesis. Thus, many measurements of a certain quantity taken over a trajectory and then averaged

$$A_{obs} = \frac{1}{N}\sum_{i=0}^{N} A_n \,, \tag{2.2}$$

where A_n is the value of a the observable in question at a given time. It is assumed to be equal to the ensemble average of the system

$$A_{obs} = \sum_{\nu} P_{\nu} A_{\nu} = \langle A \rangle . \tag{2.3}$$

While the above analysis most easily lends itself to an isolated system at a constant energy, it is typically more useful to consider a system at constant temperature. The statistics of a system at constant temperature can be justified by considering a subsystem of a much larger system at constant energy. The larger system acts as a bath to keep the subsystem at a constant temperature. The probability of finding this subsystem at a given energy is given by

$$P_{\nu} = Z^{-1} \exp(-\beta E_{\nu}), \tag{2.4}$$

Where Z is the partition function of the system, $\beta = \left(k_B T\right)^{-1}$ where k_B is the Boltzmann constant, and E_{ν} is the energy of a particular state of the system. . The partition function is described by the equation

$$Z = \sum_{\nu} \exp(-\beta E_{\nu}) . \tag{2.5}$$

While it would be necessary to employ quantum mechanics to get a complete description of a system under investigation, especially the behavior of the electrons, a classical description is typically sufficient when describing the behavior of the nuclei. In a classical description, the fluctuations of the electrons have been "integrated out," leaving an effective interaction between the nuclei. A classical Hamiltonian that describes this classical system is given by

$$H\left(\mathbf{r}^n, \mathbf{p}^n\right) = K\left(\mathbf{p}^n\right) + V\left(\mathbf{r}^n\right), \tag{2.6}$$

where $K\left(\mathbf{p}^n\right)$ is the kinetic energy part of the system and $V\left(\mathbf{r}^n\right)$ is the potential energy part of the system. In principle, the potential energy part could be described by a many-body expansion

$$V\left(\mathbf{r}^n\right)=\sum_i V_1\left(\mathbf{r}_i\right)+\sum_i\sum_{j>i} V_2\left(\mathbf{r}_i,\mathbf{r}_j\right)+\sum\sum\sum V_3\left(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k\right)+\ldots \qquad (2.7)$$

However, the potential energy part is typically truncated at the pair term that results in a potential energy term that is only dependent on the distance between two particles $V_2\left(\mathbf{r}_i,\mathbf{r}_j\right)=V_2\left(r_{ij}\right)$. Since the kinetic energy part of the Hamiltonian is typically assumed to obey ideal gas statistics, the resulting partition function for a classical system only depends n the positions of the particles

$$Z=\int d\mathbf{r}^n \exp\left(-\beta V\left(\mathbf{r}^n\right)\right). \qquad (2.8)$$

One important quantity that shows up commonly in statistical mechanics of classical liquids is the radial distribution function (RDF)

$$g\left(r\right)=V\frac{N-1}{N}\left\langle\delta\left(r-r_{ij}\right)\right\rangle \qquad (2.9)$$

The RDF can be used to describe many properties of a liquid if it is assumed to only interact among pairs. The RDF is related to the potential of mean force (PMF) by the reversible work theorem.

2.2 Molecular Dynamics

In order to get calculate the trajectory of a complex system, it is often necessary to simulate it. One common method to simulating the trajectory of a liquid is Molecular Dynamics (MD). This is done by calculating the force on each particle of a system in

question by taking the gradient of the potential energy function, $\nabla V\left(\mathbf{r}^n\right)$, used to describe the interactions between particles. The force is then related to the acceleration of the particle by Newton's second law, $\mathbf{F} = m\mathbf{a}$, and a system of coupled differential equations is then used to describe the motion of the particles. By discretizing time, the position of the particles at a future time can be approximately solved.

While the constant accelerated motion equations could be used to describe motion of the particles within a timestep, it is more precise to use the Verlet algorithm

$$\mathbf{r}\left(t+\Delta t\right) = 2\mathbf{r}\left(t\right) - \mathbf{r}\left(t-\Delta t\right) + \frac{1}{2}\mathbf{a}\left(t\right)\Delta t^2 \tag{2.10}$$

Which uses the position of the particle at the precious timestep and the position and acceleration of the particles at the current timestep. The velocity of the particles can be determined, if desired, from the equations

$$\mathbf{v}\left(t\right) = \frac{\mathbf{r}\left(t+\Delta t\right) - \mathbf{r}\left(t-\Delta t\right)}{2\Delta t}. \tag{2.11}$$

In principle, any pairwise formula could be used to describe the potential energy between atoms. However, for computational reason, the Lennard-Jones interaction is typically used

$$V_{LJ}\left(r\right) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{2.12}$$

Where ε is the depth of the well and σ is the point where the potential energy function becomes positive. This is typically complemented by pointwise electrostatic interaction

$$V_{elec}\left(r_{ij}\right) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \tag{2.13}$$

where q are the partial charges of the atom and ε_0 is the permittivity of free space. The electrostatic interactions are typically calculated by k-space based methodology such as Ewald summation or Particle-Particle Paricle-Mesh. To better simulate the bulk properties of a system, periodic boundary conditions (PBC) are typically employed.

Additionally, terms are typically employed to describe atoms that are involved in intramolecular interactions. One example is known as bonded interactions, which occur between two atoms that are involved in a covalent bond. Another example is the angled interaction, which occurs between three atoms, which share a central atom to which the other two are bonded. The form of these interactions are typically harmonic, which is the first nonconstant term of a Taylor expansion about a mimimum. Finally, dihedral interactions occur between a series of three bonds that have exactly one other bond in common. These interactions are typically some sort of periodic interactions.

2.3 Coarse-Grained Methods

Coarse-grained (CG) methods aim to create CG models, which are reduced representation of some fine-grained (FG) system. In this thesis, bottom-up methods for creating CG models are used and will be the focus of the section. CG and FG models are related to one another via a mapping operator. The mapping operator relates FG configurations to CG configurations. The most common type of mapping operators are linear mappings:

$$M\left(\mathbf{r}^n\right) = \sum_i c_{iI}\mathbf{r}_i \,, \tag{2.14}$$

Where c_{iI} is the mapping coefficient for atom i to CG site I. The mapping coefficients for a single CG site must sum to one. The types of mappings employed in this thesis are the

center of mass (COM) and the carbon alpha mapping. The COM mapping sets the mapping coefficient equal to the ratio of the FG atom's mass and the sum of the mass of all the particles mapped the CG bead. The carbon mapping sets the mapping coefficient of the alpha carbon to one and the rest of the atoms in a residue to zero. Other types of mapping have been to be effective choices in certain cases.

A bottom-up CG method aims to create interactions between CG particles that results in a CG trajectory that is consistent with a FG trajectory. A CG trajectory is consistent when the probability of finding a certain CG configuration is equal to the probability of finding all the atomistic configurations that map to it. Each method discussed below achieves this in different ways.

2.3.1 Relative Entropy Minimization

Relative entropy minimization attempts to create a model by minimizing the distance-like metric between two probabilities distributions

$$S_{rel} = k_B \int d\mathbf{r} P_T\left(\mathbf{r}^n\right) \ln\left(\frac{P_T\left(\mathbf{r}^n\right)}{P_M\left(\mathbf{r}^n\right)}\right), \tag{2.15}$$

where P_T is the probability distribution of the target ensemble, typically a mapped atomistic trajectory, and P_M is the probability distribution of the CG model to be fit. The relative entropy is well known from information theory, but can be related to molecular modeling via a maximum likelihood analysis.[32] If a canonical distribution is assumed, the relative entropy can be expressed as

$$S_{rel} = \beta\left\langle U_M - U_T \right\rangle_T - \beta\left(A_M - A_T\right) + \left\langle S_{map} \right\rangle_T, \tag{2.16}$$

where A is Helmholtz free energy of the target and model system and S_{map} is the mapping entropy.[11]

While any form can be chosen for the potential energy function in the CG model, a linear expansion of basis functions

$$U_{CG}\left(\mathbf{R}^N\right)=\sum \lambda_i \phi_i\left(\mathbf{R}^N\right), \tag{2.17}$$

Where, is typically chosen because of its convex nature with respect to the coefficients.[33] If this linear expansion of basis sets is used, derivatives with respect to the basis set coefficients can be taken

$$\frac{\partial S_{rel}}{\partial \lambda_i}=\beta\left\langle \frac{\partial U_{CG}}{\partial \lambda_i}\right\rangle_T-\beta\left\langle \frac{\partial U_{CG}}{\partial \lambda_i}\right\rangle_M, \tag{2.18}$$

And second derivative

$$\frac{\partial^2 S_{rel}}{\partial \lambda_i \partial \lambda_j}=\beta\left\langle \frac{\partial^2 U_{CG}}{\partial \lambda_i \partial \lambda_j}\right\rangle_T-\beta\left\langle \frac{\partial^2 U_{CG}}{\partial \lambda_i \partial \lambda_j}\right\rangle_M+\beta^2\left\langle \frac{\partial U_{CG}}{\partial \lambda_i}\frac{\partial U_{CG}}{\partial \lambda_j}\right\rangle-\beta^2\left\langle \frac{\partial U_{CG}}{\partial \lambda_i}\right\rangle\left\langle \frac{\partial U_{CG}}{\partial \lambda_j}\right\rangle, \tag{2.19}$$

which allows for the minimum solution to be determined using a Newton-Raphson gradient decent method.[34]

2.3.2 Multiscale Coarse-Graining

It can be shown that a CG model that is consistent with a its underlying FG model will have the same forces as the mapped FG trajectory

$$\mathbf{F}_I\left(\mathbf{R}^N\right)=\left\langle \sum_{i\in I}\mathbf{f}_i\left(\mathbf{r}^n\right)\right\rangle. \tag{2.20}$$

Thus, MS-CG models attempt to determine interactions between CG beads by the force-matching method

$$\chi^2 = \left\langle \left\| \mathbf{F}_{CG} - \mathbf{F}_{AA} \right\|^2 \right\rangle_T . \tag{2.21}$$

If a linear expansion of basis functions is used, the force-matching equation can be solved using a linear least squares fit, which does not require multiple iterations.[12, 16, 35-43]

2.4 Empirical Valance Bond

One deficiency of standard MD simulations is that they typically do not allow for bond breaking and forming. One method for overcoming this is the Empirical Valance Bond (EVB) method, which was first introduced by Warshel and Weiss. The EVB method introduces a matrix Hamiltonian

$$\mathbf{H} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \tag{2.22}$$

where V_{11} and V_{22} are standard non-reactive potential energies for the reactive and product states and V_{12} are the coupling between these two states. The matrix Hamiltonian is then diagonalized and the lowest energy solution is then used as the solution used for determining the energy of the system. Forces are then calculated using the Hellman-Feynman theorem. Off-diagonal elements are typically constants that are chosen such the resulting potential energy surface agrees with some empirical data.

2.5 Experiment Directed Simulation

One typical problem with molecular models is that improving them when new experimental information becomes available can be a time consuming process and changing parameters can sometimes provide cause undesired effects to other properties of the system. The Experiment Directed Simulation (EDS) method provides a framework to change a force field such it agrees with new experimental information while minimally

biasing the existing force field. EDS can be derived by introducing a new constraint to the maximum entropy derivation of the Boltzmann distribution, which results in the distribution function

$$P(\mathbf{r}) = \frac{e^{-\beta\left(U(\mathbf{r})+\alpha f(\mathbf{r})\right)}}{\int d\mathbf{r}\, e^{-\beta\left(U(\mathbf{r})+\alpha f(\mathbf{r})\right)}} \tag{2.23}$$

Where. It has been shown that the resulting EDS model is the one that changes the distribution functions the least from the original model, but still creates agreement with new experimental observation. The EDS parameter is solved by percoordinate adaptive online stochastic gradient descent.

Chapter 3

A Direct Method for Incorporating Experimental Data into Multiscale Coarse-grained Models

This chapter is reprinted with permission from J. Chem. Theory Comput. 2016, 12 (5), 2144-2153 41 Copyright 2015 American Chemical Society.

3.1 Introduction

An important aspect of molecular dynamics (MD) simulations is the creation of force-fields that agree with experimental data. Two common pathways to generating force-fields are bottom-up and top-down approaches. Top-down approaches rely on adjusting force-field parameters to match experimental data.[22, 44-46] Historically, this was achieved by matching thermodynamic data, such as pressure, using assumed force-field functional forms such as Lennard-Jones and Coulomb electrostatic interactions.[22] Later, more sophisticated experimental techniques such as NMR and vibrational spectra were used to parameterize force-fields.[44-45] Additionally, more complex interactions were incorporated into force-fields such as multi-body interactions.[22] Alternatively, bottom-up methods use quantum mechanical (QM) simulations to build force-field parameters.[47-50] Some of these methods rely on building force-fields that match the calculated observables from the quantum simulations, whereas others use the forces or energies directly calculated from the QM simulation.

While both top-down and bottom-up methods are used for building force-fields, both have deficiencies. Top-down methods need to be constantly updated to match new experimental data as it becomes available and may miss essential physics of the underlying mechanics. In addition, there may not be a set of parameters that creates

agreement with experiment for the set of force-field functionals being used. There is also the issue of over-fitting the simulation to available experimental data. Bottom-up methods require highly expensive QM simulations to make them be accurate. To work around the high computational costs, molecules are frequently simulated far from the condensed phase or use a limited amount of sampling. An effective middle-of-the-road approach would be to combine top-down and bottom-up approaches.

One technique for combining top-down and bottom-up techniques is take a force-field generated by bottom-up techniques and update the force-field by including a bias. A bias is introduced by including new functional forms to an existing force-field to match simulation data with experimental observables. One example is a harmonic constraint, but it can alter dynamical properties and distribution of the biased portion of the system.[51-52] Also, a coupling constant that creates agreement between the simulation results and experimental data may not exist. Islam and co-workers,[53] as well as Roux and Islam,[54] have developed the restrained ensemble MD method to reduce the undesired effects of the harmonic bias.[55-57] While the restrained ensemble MD simulation converges to a minimal bias in the limit of a large number of replicas, the multiple replicas greatly increase the computational cost of this method and does not provide a trajectory.[58]

Recently, White and Voth[31] developed the experiment directed simulation (EDS) method for biasing simulations to match experimental data. This method uses the maximum entropy argument of Pitera and Chodera,[59] based on the maximum entropy derivation of statistical mechanics of Jaynes,[32] which shows that the optimal bias is linear with the biased observable. White and Voth[31] used a per-coordinate adaptive online stochastic gradient decent method[60] to find the parameter that creates agreement with the

experimental value within a single simulation. EDS provides a rapid and computationally inexpensive way to incorporate experimental data into atomistic simulations parameterized from bottom-up methods. However, while it has been shown empirically that biasing methods that use the maximum entropy argument to bias observables causes improved agreement with an ideal model that agrees with experiment, there has been no proof that the use of this biasing form creates improvement in general. Among other things, the present paper provides such a proof.

While accurate force-fields are important to obtain reliable information from MD simulation, sometimes systems are too large or need to be run for such long time scales that the desired information cannot be easily extracted because of computer hardware limitations. These limitations have inspired the use of coarse-grained (CG) models, in which degrees of freedom of the system are eliminated. Similar to molecular force-fields, CG force fields can also parameterized in top-down or bottom up approaches.[9, 29, 61-62] However, since CG observables cannot typically be recovered from experiments, most methods must rely on all-atom simulations as their "experimental data" or to match bulk properties of the system.[29] Several systematic bottom-up approaches have been developed to parameterize CG methods from atomistic simulation, such as force-matching,[12, 25, 35, 63] relative entropy minimization,[11] inverse Monte Carlo,[64] iterative Boltzmann inversion,[65] and the generalized Yvon-Born-Green equations.[66] Relative entropy minimization, which is based on a minimizing the loss of Shannon information, is used a measure of the overlap of a model with a target ensemble.[11] Inverse Monte Carlo and iterative Boltzmann inversion can be posed as methods to minimize the relative entropy using different minimization schemes.[33, 67] Similarly, force-matching and

generalized Yvon-Born-Green equations find force-fields using the same equations,[68] while their motivations differ. While coarse-graining methods may converge to same CG potential in the appropriate limits (especially the exact limit), the different methods give different CG potentials where there is limited sampling and approximate basis sets to represent the CG effective potential.[69] As such, the local nature of force-matching, implemented as Multiscale Coarse-graining (MS-CG),[12, 25, 35, 63] provides an attractive route to develop bottom-up CG models because of incomplete sampling of the many-body all-atom probability density function. MS-CG also does not require iterative CG simulations to obtain the CG potential. However, this MS-CG approach (and other rigorous coarse-graining approaches based on an underlying all-atom force-field) may only be as good as the accuracy of the underlying atomistic force-field will allow. The subject of this paper is to show how the introduction of EDS into a potentially inaccurate atomistic force field will lead to an improved MS-CG effective potential without requiring the need to reparameterized the atomistic force-field.

The present work will first demonstrate how using a method that introduces a minimal bias, such as EDS, reduces the relative entropy of the biased atomistic system with respect to a hypothetical target atomistic model that is defined to completely agree with experiment. The utility of the biased trajectories calculated by EDS will then be demonstrated by generating MS-CG models that in turn show better agreement with experimental data. These CG models built from biased trajectories are shown to have properties closer to the accurate reference model, even when the measured CG observables are not directly biased. This combined EDS-MS-CG method therefore

provides a rapid way to build improved MS-CG models, even when only partial knowledge of the most accurate all-atom force-field is available to the simulator.

3.2 Theory

3.2.1 Minimizing Relative Entropy

Assume there exists an ensemble of particles described by the probability distribution $P(\mathbf{r}) \propto e^{-\beta(U(\mathbf{r})+\alpha f(\mathbf{r}))}$, where $\beta = 1/k_B T$ and $U(\mathbf{r})$ is the atomistic (all-atom) potential energy as a function of the n atomistic particle positions, the latter being denoted by the vector $\mathbf{r}$, i.e, $\mathbf{r} \equiv \mathbf{r}^n = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n\}$, with $\int d\mathbf{r} \equiv \prod_{i=1}^{n} \int d\mathbf{r}_i$. While a potential energy function may be built from sophisticated top-down or bottom-up methods, an imperfect model would create disagreement between some simulation observables and an experimental observation. Pitera and Chodera[59] presented a form of linear bias to create agreement between experimental observations and simulated observables while adding minimal additional information into the simulation. By adding an additional constraint to the maximum entropy approach to thermodynamics,[32] that the ensemble average of the model matches the experimental observation, a new probability distribution that is minimally biased from an established model is given by

$$P(\mathbf{r}) = \frac{e^{-\beta(U(\mathbf{r})+\alpha f(\mathbf{r}))}}{\int d\mathbf{r}\, e^{-\beta(U(\mathbf{r})+\alpha f(\mathbf{r}))}} \tag{3.1}$$

where α is a flexible parameter that is determined during a simulation using a method like EDS.[31]

While eq 1 presents a new ensemble that both agrees with experimental observations and has the least additional information included into the model to create such agreement, it has not been shown whether or not the model has improved agreement with other experimental observables in the more general sense. One such measure that could quantify improvement to the biased system is relative entropy measure proposed Roux and Weare,[58] given by

$$S_{rel} = k_B \int d\mathbf{r}\; P_T(\mathbf{r}) \ln\left(\frac{P_T(\mathbf{r})}{P_M(\mathbf{r})}\right), \tag{3.2}$$

where $P_T(\mathbf{r})$ is a target distribution that one is hoping to match and $P_M(\mathbf{r})$ is a model distribution that is being modified to match the target. Relative entropy stems from information theory and is typically used to quantify the distance between two distributions, while technically not being a proper metric.[70] While relative entropy is commonly used as an objective function for optimizing CG models with the mapped all-atom configurations as the reference, as proposed by Shell,[11] it will be used here as an objective function to analyze the difference between different molecular models at the same resolution.

For the following argument, consider three models: a target model indicated with a subscript *T*, an unbiased model indicated with subscript *U*, and a biased model indicated with subscript *B*. The target is a hypothetical model that perfectly agrees with experiment for all observables, known and unknown. The unbiased model will be some existing (and imperfect) model of the molecular system in question. The biased model is a model that has been biased by EDS, resulting in a potential energy function as given in eq 1. We assume that the model is in the same thermodynamic state as the target model, that the

form of the observable operator does not change with the addition of bias, and that the observable operator is known without error. Furthermore, we assume that the biased and target model ensemble averages for the observable being biased are in agreement, such that

$$\left\langle f(\mathbf{r})\right\rangle_T = \left\langle f(\mathbf{r})\right\rangle_B \tag{3.3}$$

In other words, we assume that there exists an α in eq 1 that creates agreement between the biased simulation and experimental observation and that the method used for this was able to successfully determine the value of α that obtains this agreement.

We will start our analysis by investigating the relative entropy of the biased system with respect to our hypothetical target by rewriting eq 2 as,

$$S_{rel} = k_B \int d\mathbf{r} P_T(\mathbf{r}) \ln\left(\frac{P_T(\mathbf{r})}{P_B(\mathbf{r})}\right) \tag{3.4}$$

A commonly used expression for the relative entropy in the canonical ensemble is given by[11]

$$TS_{rel} = \left\langle U_B(\mathbf{r}) - U_T(\mathbf{r})\right\rangle_T + k_B T \ln\left(Z_B\right) - k_B T \ln\left(Z_T\right) \tag{3.5}$$

Now, the biased potential is written in term of the unbiased potential to get

$$TS_{rel} = \alpha\left\langle f(\mathbf{r})\right\rangle_T + \left\langle U_U(\mathbf{r}) - U_T(\mathbf{r})\right\rangle_T + k_B T \ln\left(Z_B\right) - k_B T \ln\left(Z_T\right) \tag{3.6}$$

Next, the derivative of the relative entropy is taken with respect to the flexible EDS parameter α

$$T\frac{\partial S_{rel}}{\partial \alpha} = \left\langle f(\mathbf{r})\right\rangle_T + \frac{\partial}{\partial \alpha} k_B T \ln\left(Z_B\right) = \left(\left\langle f(\mathbf{r})\right\rangle_T - \left\langle f(\mathbf{r})\right\rangle_B\right) \tag{3.7}$$

where the potential energy terms and the partition function of the target and unbiased model have vanished because they are not a function of α. When the consistency

condition of eq 3 is satisfied, eq 7 will be equal to zero and the relative entropy will be at an extremum.

To analyze the extremum as a minimum or maximum, the second derivative of the relative entropy can be calculated with respect to α and is given by

$$T\frac{\partial^2 S_{rel}}{\partial \alpha^2} = -\frac{\partial \langle f(\mathbf{r}) \rangle_B}{\partial \alpha} \tag{3.8}$$

where the expectation of the target ensemble vanishes again since it is not a function of α. Now, from the analysis of White and Voth we have the result[31]

$$\frac{\partial \langle f(\mathbf{r}) \rangle_B}{\partial \alpha} = -\beta Var\left(f(\mathbf{r}) \right)_B \tag{3.9}$$

Since β and the variance are always positive quantities, this result when placed into eq. 3-8 proves convexity of the relative entropy with respect to α. Since the relative entropy is convex and at a minimum when the biased model matches the target, the EDS minimal bias that matches an observable with experiment also always decreases the relative entropy of the biased model with respect to the target model.

This above argument can be extended to the EDS biasing of multiple observables by noting that the mixed second derivatives are given by

$$T\frac{\partial^2 S_{rel}}{\partial \alpha_i \partial \alpha_j} = \beta Cov\left(f_i(\mathbf{r}), f_j(\mathbf{r}) \right)_B \tag{3.10}$$

which results in Hessian that is equivalent to a covariance matrix. Since the determinant of a covariance matrix is positive as long as observables are not completely correlated, this surface will be convex as well.

While other techniques for relative entropy minimization exist such as those used in building multiscale models, these other techniques of relative entropy minimization

differ from EDS in that the target distribution must be known completely for any optimization to occur. This is usually realized by starting from an all-atom distribution that is known completely as the target distribution and defining a CG distribution as the model. When a method like EDS is used, however, the above analysis shows that only expectation values of the target distribution (certain experimental data) need to be known to improve the agreement of the model with the target distribution that yields that experimental data. As such, this alleviates the need for determining the probability distribution of a target system in detail by some additional technique, experimental or otherwise, to help decrease the relative entropy of the model relative to the target, which is often impossible in practice.

3.2.2 EDS with Multiscale Coarse-graining: A New Paradigm

In one way or another, all coarse-graining methods attempt to make the following replacement of variables:

$$\int d\mathbf{r} \exp[-\beta U(\mathbf{r})] = C \int d\mathbf{R} \exp[-\beta U_{CG}(\mathbf{R})] \tag{3.11}$$

where a new set of CG coordinates are defined here, i.e., $\mathbf{R} \equiv \mathbf{R}^N = \{\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_N\}$ and $\int d\mathbf{R} \equiv \prod_{I=1}^{N} \int d\mathbf{R}_I$, with $N < n$ where n is the full set of atomistic coordinates $\mathbf{r}$, and $U(\mathbf{r})$ is, as defined earlier, the atomistic force-field (potential energy function). (In the above equation the ideal gas pre-factor C has been renormalized to reflect the change in number of degrees of freedom in the CG system.) Importantly, in Eq. 3-11 the expression for the free energy is re-written in terms of an integral over the set of CG "particle" or "site" positions, $\mathbf{R}$, that are N in number. Moreover, the Boltzmann factor is now written in terms of the all-important *effective potential* $U_{CG}(\mathbf{R})$ for the CG variables. The CG

variables, by virtue of the definition of coarse-graining, are fewer in number than the atomistic variables (often substantially so). The promise of CG modeling is thus to substantially reduce the computational challenge of evaluating Boltzmann averages through a combination of the fewer number of CG degrees of freedom and also the likelihood that the CG effective potential $U_{CG}(\mathbf{R})$ will be smoother and/or simpler than the full all-atom resolution one, as defined earlier by $U(\mathbf{r})$.

One can formally derive the expression for the CG effective potential in eq 3-11 by casting coarse-graining in the context of rigorous statistical mechanics as follows:[12, 25, 35, 63] Each CG site coordinate $\mathbf{R}_I$ is defined to be a function of the positions of some set of atoms in a given molecule: $\mathbf{R}_I = M_{\mathbf{R}_I}(\mathbf{r})$, where $M_{\mathbf{R}_I}(\mathbf{r})$ are mathematical mapping operators. One example of such a mapping function is the center of mass of a group of atoms. With this mapping in hand, one arrives at the formal definition of the CG effective potential $U_{CG}(\mathbf{R})$, given by

$$U_{CG}(\mathbf{R}) \equiv -k_B T \ln\left[\int d\mathbf{r} \prod_{I=1}^{N} \delta(M_{\mathbf{R}_I}(\mathbf{r}) - \mathbf{R}_I) \exp[-\beta U(\mathbf{r})]\right] + const. \qquad (3.12)$$

Equation 10b is both physically informative and also reveals just how challenging the task of finding $U_{CG}(\mathbf{R})$ can be, because this equation reveals that the CG effective "potential" is in fact a many-dimensional free energy surface (i.e., the so-called many-body potential of mean force, or PMF) for the CG variables $\mathbf{R}$. This is because, in a formal sense, certain degrees of freedom have been integrated out in evaluating eq 10b. As such, the effective CG potential will contain "missing entropy" effects arising from the atomistic degrees of freedom that have been integrated over when transforming the equation to the CG variables. These entropic effects can be illusive and their behavior

hard to predict, so therein lies the origin of one of the key challenges in coarse-graining at the fundamental level. In its most rigorous form, coarse-graining can be considered a process of "renormalization" of interactions into a new representation having a lower overall dimensionality.

By virtue of the above analysis, the MS-CG theory shows that those regions of atomistic configuration space that are consistent with a specific set of values of **R** contribute to the integral that gives the CG effective potential for that set of values of **R**. The exact CG effective potential of Eq. 3-12 is formally defined in terms of the exact equilibrium distribution function of the CG coordinates. If one treats Eq. 3-6 as a an effective potential energy function, adds a corresponding kinetic energy term, and constructs a Hamiltonian for the CG degrees of freedom, and if this Hamiltonian is used to generate MD equations of motion in the usual way, then it is straightforward to show[12, 68] that the calculated static equilibrium distribution function of the coarse grained degrees of freedom will be exact. In other words, the CG effective potential $U_{CG}(\mathbf{R})$ gives the exact partition function and hence exact free energy if sampled from a Boltzmann distribution for the CG variables depending on that effective CG potential.

In practice, the CG potential must be some approximation to the exact expression embodied in Eq. 3-12. One approach such as the MARTINI model[24, 71-72] is to develop the CG effective potential through *ad hoc* procedures by fitting the CG model directly to experimental data and then assuming that the observables calculated by such a model can be easily defined in the CG mapping. This approach has proven quite popular due to its ease of use, especially by those who seek to get results from computer software but do not wish to become immersed in the complexities of coarse-graining, but the link to

rigorous statistical mechanics (and certainly to the underlying atomistic potential energy function) are mostly lost. By contrast, the MS-CG approach[12, 16, 25, 35-43, 63] provides a direct variational methodology to define the effective CG potential based on a statistical mechanical formulation of force-matching. The MS-CG variational functional is defined as

$$\chi^2\left[\vec{F}^{CG}\right] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \left| \vec{F}_I^{CG}(\mathbf{R}) - \vec{f}_I^{AA}(\mathbf{r}) \right|^2 \right\rangle \tag{3.13}$$

where $\vec{F}_I^{CG}(\mathbf{R}) = -\vec{\nabla}_I V_{CG}(\mathbf{R})$ is the force on the coarse-grained site *I* as a function of the configuration of the coarse-grained system at **R**, while $\vec{f}_I^{AA}(\mathbf{r})$ is the all-atom (AA) projection of the forces from the atoms making up that CG site. (Note that the all-atom force projection, $\vec{f}_I^{AA}(\mathbf{r})$, defined here should not be confused with the EDS bias function $f(\mathbf{r})$ defined earlier in eq 1.) As an alternative to the MS-CG approach, relative entropy minimization between AA and CG models has also been shown to be a useful approach,[11, 33, 67] which gives equivalent results in the exact coarse-graining limit.[69]

At this juncture in our analysis, the formal combination of MS-CG with EDS is clear and relatively straightforward. In order to accomplish such a synthesis of methods, one simply replaces the CG site-projected AA forces in eq 10c, $\vec{f}_I^{AA}(\mathbf{r})$, by the equivalent CG-site projected AA *biased* forces from the EDS distribution function given in eq 1. The resulting EDS-MS-CG method might be viewed in essence as a new paradigm in CG modeling and simulation, in that it combines the "bottom-up" features of the MS-CG approach with the "top-down" features of EDS (and hence, at least in spirit, with the philosophy of fitting CG models to experimental data as in the MARTINI

approach). A key feature here is that, in contrast to an *ad hoc* approach, this combination of methods is done within a rigorous statistical mechanical relative entropy framework, as described in the previous subsection.

In the next section, the results from the combined EDS-MS-CG approach are numerically illustrated for two liquid state systems, liquid methanol and ethylene carbonate.

3.3 Simulations

All MD simulations of methanol and ethylene carbonate were performed with the software LAMMPS.[73] A control simulation of methanol was performed using the OPLS all-atom force field.[47] Methanol was simulated with a timestep of 1 fs and used non-bonded Lennard-Jones interactions with a radial cutoff of 1 nm. Particle-particle particle-mesh (PPPM) was used to treat long-range electrostatics.[74] The system was equilibrated for 5 ns at constant *NPT* at 1 atm and 300 K, setting its volume to the average of the last 2 ns of *NPT* simuation, and then simulated longer for 1 ns at constant *NVT* at 300 K. A box of 1000 methanol molecules were simulated under a Nose-Hoover thermostat[75] at 300 K with a thermostat time constant of 100 fs. The was then simulated for 4 ns, collecting configurations every 250 fs.

A control simulation of ethylene carbonate was performed using the force-field described by Masia et al.[49] Ethylene carbonate was simulated with a timestep of 1 fs and used non-bonded Lennard-Jones with a cutoff of 1.25 nm. PPPM was again used to treat long-range electrostatics.[74] A box of 250 ethylene carbonate molecules was simulated with a Nose-Hoover thermostat at 325 K and thermostat time constant of 50 fs.[49, 75] The

system was equilibrated for 1 ns, then simulated in production model for 9 ns, collecting configurations every 150 fs.

A "deficient" (i.e., inaccurate) model for methanol was generated by multiplying the atomic partial charges by a factor of 0.5, while the deficient model for ethylene carbonate was generated by multiplying the atomic partial charges by a factor 1.5. For purposes of comparison, the deficient models of ethylene carbonate and methanol were simulated under the same conditions as the control simulations.

The EDS biased models were generated by biasing the deficient models with the EDS method implemented in the PLUMED package.[76] Interested readers are directed elsewhere for the algorithmic details,[31] but briefly and as described earlier, EDS biases a force field by introducing a new term to the potential energy of the system that is linear with the biased observable, as in eq 1,

$$U'(\mathbf{r}) = U(\mathbf{r}) + \alpha f(\mathbf{r}), \tag{3.14}$$

where the flexible parameter α is determined by stochastic gradient decent. Coordination number and its moments were the observables chosen to be matched with the reference simulation in the present applications. Since the typical definition of the coordination relies of using a step function with an ill-defined derivative, the following function was used to define the coordination number observable operator:[77]

$$f(r) \approx \begin{cases} \dfrac{1-\left(\dfrac{r-r_o}{w}\right)^6}{1-\left(\dfrac{r-r_o}{w}\right)^{12}}, r > r_o \\ 1, otherwise \end{cases}, \tag{3.15}$$

where r_o is the start and w is the width of the sigmoidal function. Moments of the coordination number were calculated by

$$\left\langle r^m f(r) \right\rangle = \int r^m f(r) g(r)\, dr\,, \tag{3.16}$$

where $g(r)$ is the radial distribution function (RDF) of the atomic sites in question. The coupling constant range was set to be 1000 kcal/ mol and iteration duration of 1.0 ps. The inaccurate models were simulated for 100 ps before EDS was initiated. The bias was then equilibrated for 1 ns. Statistics were next gathered for 9 ns.

All CG forces were calculated using the MS-CG methodology.[25] In all CG models, pairwise interaction potentials were used for intermolecular interactions. The one-site methanol forces were calculated with a nonbonded cutoff of 1.0 nm and sixth order spline basis functions with a resolution 0.04 nm. CG simulations were initiated from a mapped configuration from the sampling run. Ethylene carbonate CG forces were calculated with a nonbonded cutoff of 1.25 nm and sixth order spline basis functions with a resolution of 0.05 nm. For the three-site model, fourth order spline basis functions with a resolution of 0.01 nm were used. Bonded terms were calculated by using a third order spline basis functions. The bonded terms were then fitted to a harmonic potential. The CG simulations were equilibrated for 1,000,000 timesteps, then configurations were sampled every 150 timesteps for 1,000,000 timesteps. Figure 3-1 depicts the relationship between the reference, deficient, and biased models, and how the EDS is used to create a biased model for MS-CG .

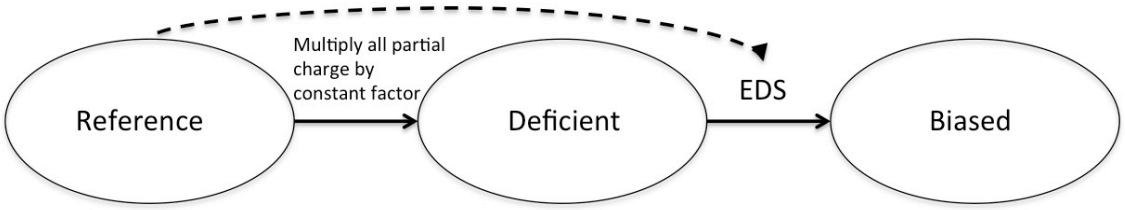

Figure 3-1: Representation of the relationship between the different models created in this work. The reference system is the target model to be reproduced. It is considered that the accurate CG model and comes from the application of MS-CG to the accurate atomistic force field. The deficient model was created by multiplying all the partial charges by a constant factor of 0.5 for methanol and 1.5 for ethylene carbonate. The deficient model was then biased with EDS so that the coordination number and its moments matched the reference model, creating the biased model. The resulting EDS model was then coarse-grained with MS-CG and compared to the MS-CG model for the reference system.

3.4 Results

In this study, an established force-field's parameters were changed systematically and then biased with EDS so that the observables of the systematically changed force-field agreed with the original (considered to be accurate) reference force-field. Throughout this section, the accurate force-field will therefore be called the reference model and the systematically changed force-field will be called the deficient model. We use a reference simulation as our target so that we can compare distributions and CG models. Normally of course, one would not have knowledge of the ideal model to which to compare the EDS results , but instead only the experimental observables. Likewise, the trajectories of the ideal model needed to build a CG model would be unavailable to the simulator.

As noted earlier, the deficient models were created by reparameterizing a currently existing force-field so this latter reference model could represent the ideal target. Partial charge was chosen as the parameter to alter since it is usually parameterized in bottom-up approaches from gas-phase data, sometimes away from the state point of the condensed phase. Furthermore, the long-range nature of partial charges can cause large effects on the model when the partial charges are varied.[78] This makes varying the partial charge parameter directly to improve one property, such as the coordination number, an ineffective route since other properties that already agree with experiment will likely be changed in the process. On the other hand, using EDS to improve a small set of system properties is ideal as only properties directly correlated will be affected, and only by a minimal amount. Additionally, when EDS is used to match properties, the relative entropy with an ideal target is guaranteed to decrease as proven earlier, while modifying the force-field parameters directly by hand could actually increase the relative entropy.

For all the EDS biased models, a simulation observable is matched to agree with the reference model. Coordination number was chosen as the observable to be matched for all the models described. While the RDF is the typical observable used for comparing two models, biasing to match the entire would require determining a coupling constant at each pairwise distance or the use of a free energy method to design a free-energy surface along the collective variable.[31, 79] Furthermore, obtaining a RDF directly from experiment can be an uncertain task. To demonstrate the flexibility of the method, a much more gentle bias of the coordination number, which requires much less information about the system, was implemented instead.

3.4.1 Methanol

Methanol was chosen as the first test case for the EDS-MS-CG methodology because of the existing studies in the literature on MS-CG models for this system.[35] Also, methanol is an amphiphile with hydrogen bonding and dipole-dipole interactions, which have a strong dependence on the partial charges. As noted earlier, the deficient model was generated by multiplying the partial charge of each of the atoms by a factor of 0.5.

In the EDS biased simulation, the coordination number of the oxygen atoms with respect to other oxygen atoms was biased to match the reference model. This was chosen since the oxygen atom is both the most highly charged atom and most massive atom in the molecule. Additionally, information about the coordination number of oxygen can often be interpreted from experimentally scattering data. The parameters for the coordination number were chosen such that the sigmoidal part of the coordination number operator were centered at location of the first desired peak, which corresponded to a r_o = 0.25 nm and a width w of 0.1 nm. The cutoff of the observable operator was set to be 1.0 nm. In order to better reproduce the shape of the RDF, the zeroth through third moments of the coordination number were also matched with EDS. The EDS bias converged within 500 ps to the value in the reference simulation. While the second solvation shell in the all-atom MD simulation is overstructured in the deficient system, as seen in figure 3-2a, the bias was able to correct it to make the first shell more dense much like the reference. Further more, the center of the first peak is closer to zero in the biased system, as in the reference. Also, the all-atom RDF of the biased system after the second solvation shell also agrees much better with the reference system.

A center of mass (COM) CG mapping was used to create MS-CG models of all three atomistic models. The solvation shell structure of the MS-CG model parameterized from the biased trajectory (MS-CG:B) more closely matches the solvation structure of the reference CG trajectory, even though only one of the constituent atoms of the CG site is biased to match the reference distribution. In figure 3-2b, the MS-CG model parameterized from the deficient system (MS-CG:D) only has one solvation shell within 0.6 nm while the reference system has two shells. The MS-CG:B is able to reproduce the two-shell structure of the reference. (Note here that the reference for the CG models is the CG site projection taken from the all-atom MD data for the reference models.) Also, the solvation structure after the first two solvation shells is much better reproduced in MS-CG:B than in the MS-CG:D. The biasing method of EDS is therefore able to improve other properties not directly related to the observables being matched, such as the center of mass RDF, and these improvements are captured by the MS-CG.

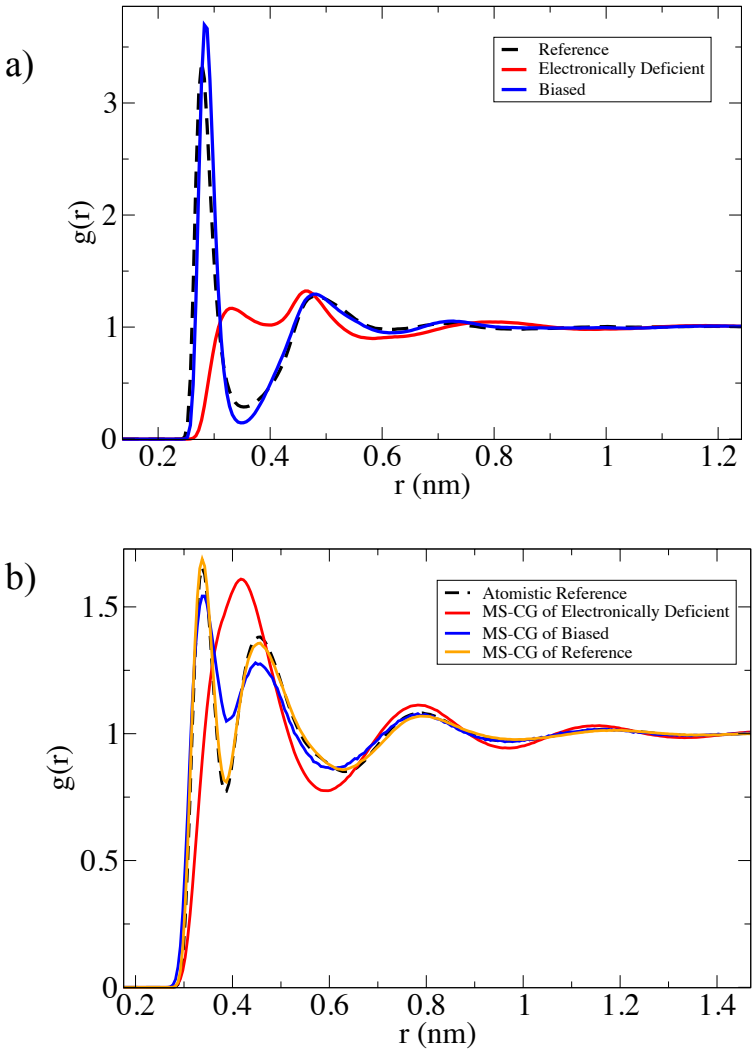

Figure 3-2: RDFs of the methanol simulations. Biased simulations were created by performing an EDS simulation on the electronically deficient model (scaled partial charges) by matching the first four moments of the coordination number of the oxygen with itself to the reference simulation. Panel (a) shows the all-atom RDF of the oxygen with itself in the atomistic MD simulations. Panel (b) shows the RDFs of one-site MS-CG models parameterized using the different atomistic trajectories along with the mapped reference trajectory. Both panels show clear improvements of peaks heights and positions between deficient and biased models when compared to the reference.

3.4.2 Ethylene Carbonate

Ethylene carbonate was chosen as the second test case. Ethylene carbonate is a commonly used solvent in lithium ion batteries, which are widely commercially available due to their large power density.[80] While widely produced, understanding of the charge cycles and the formation of films on electrolyte surfaces is still not complete.[81-82] As new experimental data is becoming available, EDS allows for direct implementation of new experimental data into molecular simulations. However, due to the large length and time scales at which processes relating electrode film formation and other degradation take place, all-atom simulations are still challenging to investigate these phenomena.[83-84] Thus, CG simulations are an attractive technique to study these processes computationally. The marriage of MS-CG and EDS can provide a rapid way to generate improved CG models.

As noted earlier, the "electronically deficient" model of ethylene carbonate was generated by multiplying the atomic partial charges of each atom by a factor of 1.5. In the biased simulation, the coordination number of the carbonyl carbon was biased to the match reference simulation since it is the atomic site with the greatest partial charge. The coordination number was matched at two different distances so that the first two-solvation shells of ethylene carbonate could be matched separately from one another. The center of the coordination number biases were set to the peak of the first two solvation shells of ethylene carbonate at r_o = 0.435 nm and r_o = 0.60 nm. While the coordination number captures the area underneath solvation shells, the third moment of these two peaks was also biased as well to match the reference simulation to capture the shape of the solvation shell. The width parameter was set to 0.1 nm in all biases. All biases

converged within the first 500 ps to the value in the reference simulation. As seen in figure 3-3a, the deficient model over-structures the first two solvation shells at the all-atom level, while the peak heights of the biased system match the all-atom reference RDF much better than the deficient model. The RDF after the first two peaks also matches the reference distribution much better than the deficient system, even though there is no bias directly acting on any particles relating to pairs at that distance.

Two MS-CG models were then made of ethylene carbonate. A one-site model of ethylene carbonate was created by using a center of mass mapping. As seen in figure 3-3b, peak heights of the MS-CG:B RDF are much closer match the reference distribution than the RDF of MS-CG:D. In fact, the MS-CG:B matches the reference CG distribution more closely than the MS-CG model that was parameterized directly from the reference trajectory (MS-CG:R).

A three-site MS-CG model of ethylene carbonate was also created, where one site was the COM of the carbonyl carbon and carbonyl oxygen, and the other sites were one side oxygen, a carbon, and two hydrogens of the ethylene. A comparison of the COM RDFs of the three-site CG models is presented in figure 3-3c. As with the methanol MS-CG models, the peak height of the first solvation shell of MS-CG:B the atomistic reference is a closer match to the reference than MS-CG:D, as well the RDF past the first solvation shell. Another interesting result of CG models built from all-atom trajectories biased with EDS is the improvement of observables not directly biased by EDS. This is illustrated in figure 3-3d, which shows the RDFs of the CG sites containing a side oxygen and an ethylene carbon and two hydrogens. Even though the sites being measured do not contain a site being biased by EDS, the peak heights of the first two solvation shells more

closely match the reference model, as well as the structure past the first two solvation shells.

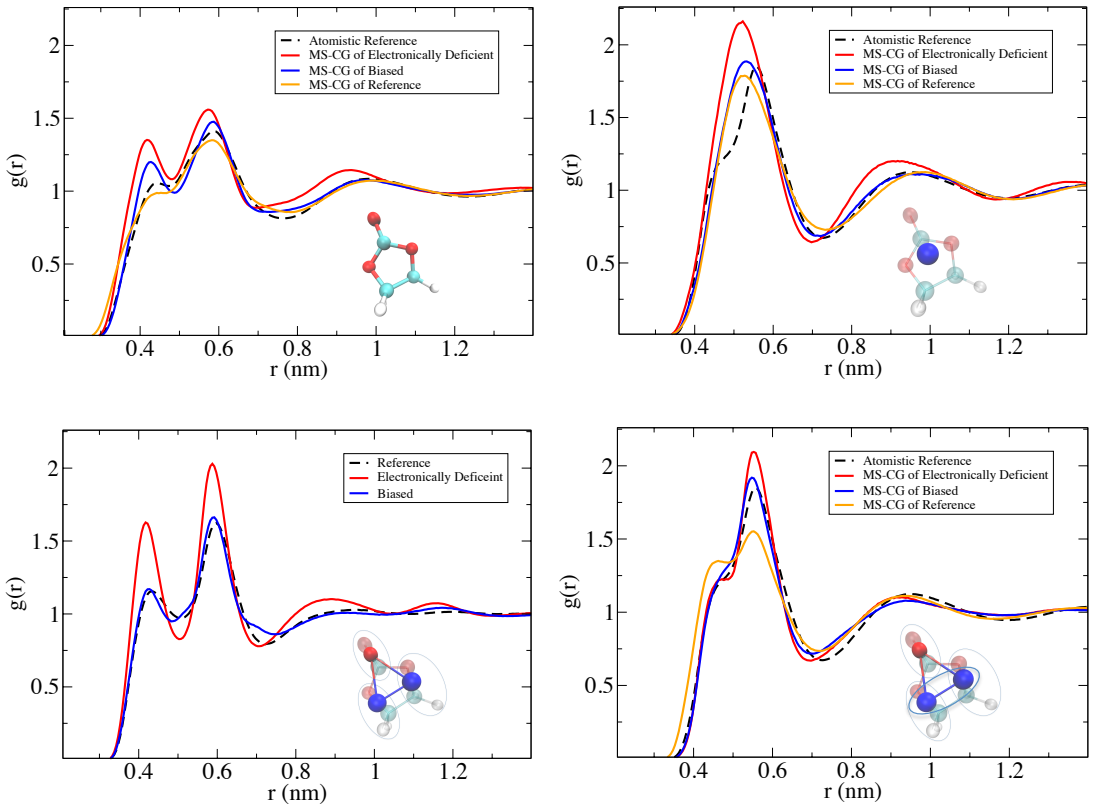

Figure 3-3: Radial distribution functions of the ethylene carbonate models. Biased systems were generated from the electronically deficient system while matching the zeroth and third moments of the coordination number for the first two solvation shells of the carbonyl carbon with itself. Panel (a) shows the all-atom RDF of the carbonyl carbon with itself. Panel (b) shows the RDF of the one-site MS-CG models parameterized using the different atomistic simulations along with the atomistic references trajectory mapped on to the CG site. Panel (c) shows the RDF of the center of mass of the three-site MS-CG models compared with a atomistic trajectory mapped on to the CG sites. Panel (d) shows the RDF of site type 2 with itself from the three-site model, whose constituent atoms were not directly biased by EDS.

3.5 Discussion

In both the methanol and the ethylene carbonate cases, the MS-CG:B result (the EDS-MS-CG approach) clearly matches the atomistic reference CG distribution better than does the MS-CG:D result. Interestingly, the one-site MS-CG:B model of of ethylene carbonate, in which the EDS-MS-CG is applied to the deficient model, matches the "exact" data from the atomistic reference better than does the MS-CG:R, i.e., the system in which the reference system is coarse-grained with MS-CG directly. This should not suggest that MS-CG models of an EDS biased deficient system will always give better data than direct MS-CG models of the ideal reference system. In the case of ethylene carbonate, the direct MS-CG model of the reference is likely less accurate because the essential physics of the system cannot be completely captured by a pairwise CG basis set for this mapping. A similar case has been seen with linear hexanes where angle distributions of hexane are by only bond and angle terms.[43] In the instance of one-site CG ethylene carbonate, the pairwise basis set appears to captures the biased data at the CG level more accurately than it does the reference data directly, but this may not always be the case.

While it has been shown that the relative entropy of a system decreases by performing a minimal bias, it should be noted that the form of the bias is important for the result to hold. There exist other types of biasing techniques, such as harmonic constraints, that can also create agreement between simulation observable and experimental observations, but these do not necessarily decrease the relative entropy of the model with respect to an ideal target. While this has been discussed with graphical comparison by Pitera and Chodra,[59] there has been no analysis of the improvement by

using other forms of biasing techniques. For instance, if a similar analysis to the one performed earlier is carried out for a harmonic constraint bias, given by

$$U_{HR} = U_U + \frac{k_r}{2}(f(\mathbf{r}) - \langle f \rangle_T)^2 \tag{3.17}$$

where k_r is the flexible parameter, one cannot show that the relative entropy is guaranteed to decrease if agreement between experimental observation and simulation observable is achieved. If the first derivative of the relative entropy with respect to harmonic constraint parameter is taken, one gets

$$\frac{\partial S_{rel}}{\partial k_r} = \frac{1}{2T}(\langle f^2 \rangle_T - \langle f^2 \rangle_B + \langle f \rangle_T^2 - \langle f \rangle_B \langle f \rangle_T) \tag{3.18}$$

Now if one assumes that the consistency condition of eq 3 is achieved with the harmonic constraint, eq 18 becomes

$$\frac{\partial S_{rel}}{\partial k_r} = \frac{1}{2T}(\langle f^2 \rangle_T - \langle f^2 \rangle_B) \tag{3.19}$$

which need not be zero if only the mean value of f is constrained to agree with an experimental observation. Furthermore, if the second derivative is taken

$$\frac{\partial^2 S_{rel}}{\partial k_r{}^2} = \frac{\beta}{2T}(\langle f^4 \rangle_B - \langle f^2 \rangle_B^2 + 3\langle f^3 \rangle_B \langle f \rangle_T + 2\langle f^2 \rangle_B \langle f \rangle_T^2 + \langle f^2 \rangle_B \langle f \rangle_B \langle f \rangle_T + \langle f^2 \rangle_B^2 \langle f \rangle_T^2) \tag{3.20}$$

which is not necessarily convex. The sign of eq 20 depends on the sign and magnitude of the observable being matched. This analysis suggests that minimal bias, such as the one utilized by EDS, will decrease the relative entropy, but a harmonic bias may increase or decrease the relative entropy depending on the unbiased model. Thus, the functional form of the bias that creates agreement with experiment is as important as the agreement

between experiment and simulation if one wishes to decrease the relative entropy between the biased system and the target system.

3.6 Conclusion

In this paper an analysis has been presented of how the relative entropy of a model with respect to a hypothetical ideal target decreases when a minimal bias is applied to the system, such as the one generated by EDS. Then, an approach was developed for using an EDS bias with known experimental information, applied to an imperfect atomistic model (potential energy function), to improve the properties of an MS-CG model. As cases to test this new approach, EDS was used to improve "deficient" models of methanol and ethylene carbonate by matching the coordination number of the oxygen of methanol and the carbonyl carbon of ethylene carbonate to an accurate reference system. The resulting MS-CG models showed substantial improvements in the solvation structure and peak heights of the solvation shells compared to "exact" data generated from atomistic MD for the accurate reference system projected onto the CG variables. Biasing RDFs associated with the constituent atoms of one site of an MS-CG model also improved the RDFs of sites whose constituent atoms were not directly biased, illustrating that the properties of the model are improved in general, not only for just those properties subject to the EDS bias. This new EDS-MS-CG methodology can therefore be used to generate improved CG models from atomistic forcefields as experimental data becomes available without the time-consuming re-parameterizations of force fields, or to provide reasonable CG models when one has imperfect knowledge of all of the interactions in a given system of interest.

Chapter 4

Coarse-Grained Directed Simulation

This chapter is reprinted with permission from J. Chem. Theory Comput. 2017, 13, 4593-4603 Copyright 2017 American Chemical Society.

4.1 Introduction

It is a fundamental challenge of molecular dynamics (MD) simulation that the accessible time and length scales are limited by the level of detail at which a system is described.[28] As a consequence, computational studies generally compromise on the size of the system under consideration in order to achieve a desired level of detail. A prime example arises in the study of enzymes where reactivity must be incorporated into the simulation, but it is thus far impractical to treat the entire system quantum mechanically. Hence, in order to do a quantum calculation on the active site while including the important role of the protein environment, hybrid methods have been developed to couple the protein's fluctuations into the reactive subsystem. These ideas led to the commonly used practice of QM/MM simulation, and contributed to the 2013 Nobel Prize in Chemistry.[85]

The continued development of enhanced sampling techniques, as well as advances in computational power, have made it possible to compute complex free energy surfaces for quantum mechanical reactions or conformational transitions taking place in biomolecular systems.[86-87] However, many problems of current focus do not occur isolated in solution, but rather in a much larger macromolecular context. For example, we might be interested in a process occurring for a particular protein embedded in a membrane or one within a much larger many-protein complex. In these cases, free energy sampling methods, which aim to enhance sampling along "slow" degrees of freedom in a low dimensional set of

collective variables (CVs), are constrained by our ability to sample the many other degrees of freedom in the system. In the past, such simulations have been made to be more computationally tractable by extracting the key subcomponent of the system and adding restraints or harmonic biases to keep this subsystem in approximately the same configuration as it would sample in its native (much larger) context. Unfortunately, doing so neglects both the effect of long range forces coming from the other molecules in the complex, and alters the fluctuations sampled by the subsystem, which in many cases are strongly coupled with dynamics in the larger system. In this work, we focus on the latter problem, namely, we wish to have a protein of interest isolated in solution sample configurations as if it were embedded within a larger multi-protein complex, thus greatly reducing the computational cost and hence enabling studies where extensive free energy sampling is possible. In our case, the larger complex can also first be simulated for some limited amount of MD sampling time in order to obtain information about the conformational ensemble of the subsystem of interest (alternatively, one could imagine approximating this information from normal mode or elastic network model analysis performed on experimental structures[88]).

In an ideal world with perfect sampling, one could simulate the subsystem by first finding an exact potential of mean force, integrating over all the other configurations of the larger system. In the canonical ensemble, for a subsystem with *N* particles having the coordinates $q = (q_1, q_2, \dots, q_{3N})$ and a supersystem with *N*+*M* particles with coordinates $r = (q_1, q_2, \dots, q_{3N}, q_{3N+1}, \dots, q_{3N+3M})$ and overall potential energy function $U(r)$, this corresponds to first calculating:

$$F\left(X\right)=-k_B T\ln\left(\frac{\int dr\delta\left(q-X\right)e^{-\frac{U(r)}{k_B T}}}{\int dre^{-\frac{U(r)}{k_B T}}}\right). \tag{4.1}$$

This new many-body potential of mean force (PMF), $F(X)$, captures the effects of the exact coupling to the larger supersystem, and could be used to compute any desired observable function of the subsystem variables, $\langle f\rangle$, as if we had performed the computation on the supersystem:

$$\left\langle f\right\rangle=\frac{\int drf\left(q\right)e^{-\frac{U(r)}{k_B T}}}{\int dre^{-\frac{U(r)}{k_B T}}}=\frac{\int drf\left(q\right)e^{-\frac{F(X)}{k_B T}}}{\int dre^{-\frac{F(X)}{k_B T}}}. \tag{4.2}$$

However, we know in practice this cannot be done for anything but the simplest cases, both because the amount of sampling required would be enormous and because the amount of information required to express the many-body PMF $F(X)$ is too large.

In this work, we suggest an alternative approach and show that some key information about the supersystem can be imparted to the subsystem in a different and computationally practical way. Here, we will modify the potential energy function that would normally be used for the subsystem alone, by "learning" a bias function on coarse-grained observables of interest, $\langle f_i\rangle$, via a relative entropy based approach. The bias is such that when those observables are sampled by the biased subsystem they will match what would be observed by sampling from $F(X)$.

The idea of adding a bias to incorporate additional information into a system's Hamiltonian comes from work where experimental data is included into a molecular dynamics simulation via either adding a set of linear biases to the system's Hamiltonian

or alternatively using many copies of the system are biased to have an ensemble averaged target observable.[89-90] These techniques have been shown to minimally bias the system such that it samples the target values of the observables.[89] Several techniques now exist to determine a bias to match experimental data on-the-fly,[31, 91] starting with a method called Experiment Directed Simulation (EDS), where the linear restraints are learned through a stochastic gradient descent procedure.[31] Later work demonstrated that this type of EDS bias always decreases (or at least, does not increase) the relative entropy with respect to an ideal target distribution that gives the same desired properties, and so this method can be used to systematically improve multiscale coarse-grained (MS-CG) models that may utilize an imperfect molecular force field. This method of adaptive linear biases is also beginning to demonstrate its utility in other applications, including providing a significant improvement of the static and dynamic properties of an *ab initio* MD (AIMD) water system which is based on a rather inaccurate level of electronic density functional theory, biasing the oxygen-oxygen radial distribution function.[92] A similar method was used to incorporate experimental data (and additionally, experimental errors) to improve the quality of force fields for RNA.[91] The empirical evidence in all cases is that the biasing of one observable tends to improve (or at least not decrease) the quality of others which were not biased, as would be expected for this class of minimal bias techniques.

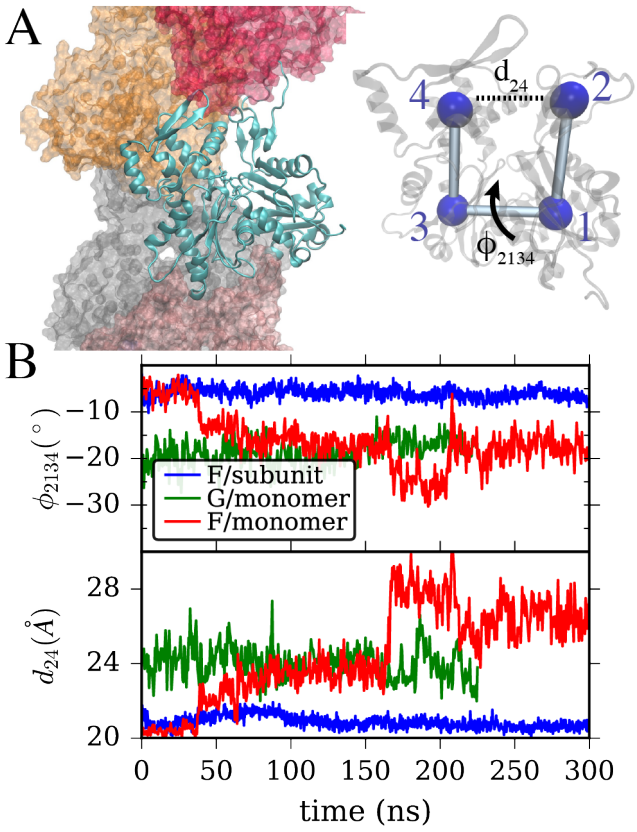

Figure 4-1: (A) Left, a snapshot from an actin filament simulation shows one actin subunit in ribbon style with a bound ATP molecule, surrounded by adjacent subunits. Right, a single actin subunit is overlaid with CG beads at the center of mass of its four major subdomains. Important CVs describing the transition from globular to filamentous conformation are the "cleft distance" from bead 2 to 4, and the "twist" dihedral angle formed by the four subdomains, a rotation around the central "bond" as shown. (B) The values of twist angle and cleft distance are shown for three systems. In blue is a single actin subunit within a filament, in green, a single G-actin in solution starting from its crystal structure, and in red, a single actin in solution starting from the filamentous structure.

In this work, we test whether the methods developed in EDS can learn biasing parameters that constrain a subsystem to behave as if it is embedded in its native (much larger) supersystem environment, recapitulating some of the desired properties from Eqs. 1 and 2 above. We apply these ideas to the protein actin, because it is a challenging biopolymer target with complex structural transitions that have nevertheless been relatively well characterized in MD simulation and experiment by our group and others. In solution, actin exists as a globular domain (G-actin) with a bound ATP molecule[93-94] and adopts a twisted conformation characterized by a dihedral angle of its four sub-domains ~-20.[95] Monomers can assemble to form a non-covalent semi-flexible biopolymer (F-actin), within which each subunit is flattened, adopting a dihedral angle . This flattening is associated with an increase in rate of actin catalyzed hydrolysis of the bound ATP molecule by a factor of $> 10^4$,[86, 96] such that the hydrolysis rate is on the order of ~1 sec^{-1}. The release of the free inorganic phosphate is very slow, occurring on a time scale of ~5 min[97] and makes the actin filament softer and more prone to depolymerization.[98-101] The hydrolysis and phosphate release are crucial processes governing the lifetime and structural properties of actin filaments and cytoskeletal networks in cells, and the molecular mechanisms have been studied using simulations. While it is now relatively standard to simulate with MD a semi-periodic actin filament consisting of 13 subunits (~500,000 atoms when solvated with water), the extensive simulation time needed to study phosphate release by free energy methods and the QM/MM methods required to study the explicit hydrolysis reaction preclude simulating such a large system; instead, only a single actin monomer has been used while being restrained in the filamentous or globular form.[86-87]

Below, we will show that using the EDS approach on a set of CG observables (a combination we term Coarse-grained Directed Simulation, or CGDS) can be used to minimally bias an actin monomer to be in a filamentous-like configuration while maintaining correct fluctuations for the two collective variables that characterize the transition from globular to filamentous actin structure.[94, 102-103] However, given the large and complex system size compared to what has been studied previously, we found that the previous algorithms did not learn EDS biasing parameters fast enough to achieve these goals. Hence a major part of the present work is devoted to improving the algorithms for this type of problem so they can be practicable for similar future applications. These enhancements are first developed on a CG model of actin for speed of testing and development, then demonstrated on fully atomistic systems. All of these algorithms have been implemented and are available for use in the open source sampling library PLUMED2,[104] and the major components are already included in the main release of the PLUMED2 software as an optional module.

The remaining sections of this chapter is organized as follows:

1. Methods: The construction both of atomistic systems and CG systems are described. The CGDS relative entropy minimization algorithm is written in a general framework that encompasses both prior work and our changes to the algorithms.
2. Results: A CG model is used to show how the prior methods can be optimized, and the performance of a first order guess for bias parameters. We then show how the optimization algorithm can be further improved for multiple CVs by transitioning to a simultaneous update of the bias parameters rather than a

stochastic one, in this case. These improved algorithms are then demonstrated to work for a monomeric as well as trimeric actin system which would be appropriate for future free energy simulations.

3. Discussion and Conclusions: Future outlook and ramifications, as well as the challenges encountered are discussed.

4.2 Methods

4.2.1 Molecular dynamics simulations of actin filaments and monomers

G-actin with a bound ATP and a periodic 13 subunit F-actin structure with bound ATP were built and equilibrated at 310 K as described previously[102-103, 105] (~94,000 atoms and ~485,000 atoms respectively), with the structure of ATP-bound actin derived from the crystal structure PDB ID 1NWK[106] and for F-actin from the electron microscopy structure in PDB ID 2ZWH.[95] For the filament, the actin subunit had its nucleotide replaced by an ATP, bound magnesium, and waters within 5 angstroms of the nucleotide from a previously equilibrated monomer simulation. MD simulation of these structures was then performed using GROMACS for ~5 ns to relax the configurations before the data shown below. A third and fourth system were then created by solvating a single actin monomer (~94,000 atoms) and a trimer of actin monomers (~138,000 atoms) from the equilibrated filament structure and relaxing those structures for ~5 ns.

4.2.2 Construction of an elastic network model for an actin monomer

An MD trajectory of the G-actin monomer bound to ATP was used to generate a CG elastic network model of an actin monomer as follows: after equilibrating the structure for 20 ns, the next 50 ns of simulation data were mapped onto a previously characterized

12-bead representation of an actin monomer, with beads 1-4 representing the four major subdomains of actin, seven others comprised by other important sub-regions, and a final bead for the nucleotidea heterogeneous elastic network model (hENM) was then built from this trajectory using a method described elsewhere.[13] In brief, all pairs of beads closer than 100Å were connected by an effective spring whose rest length was given by the average separation in the MD trajectory and whose spring constants are all identical at first. An iterative procedure was then performed that updates the values of the spring constants by an amount proportional to the difference between the normal mode fluctuations along that bond in a given iteration and the fluctuations of that pair distance in the mapped MD simulation (summed over $3N_{beads}$ normal modes). MD simulations of this CG model then reproduce approximately the structural ensemble observed in the original all-atom trajectory.[13] This hENM model was then used to quickly test improvements in the methods described next.

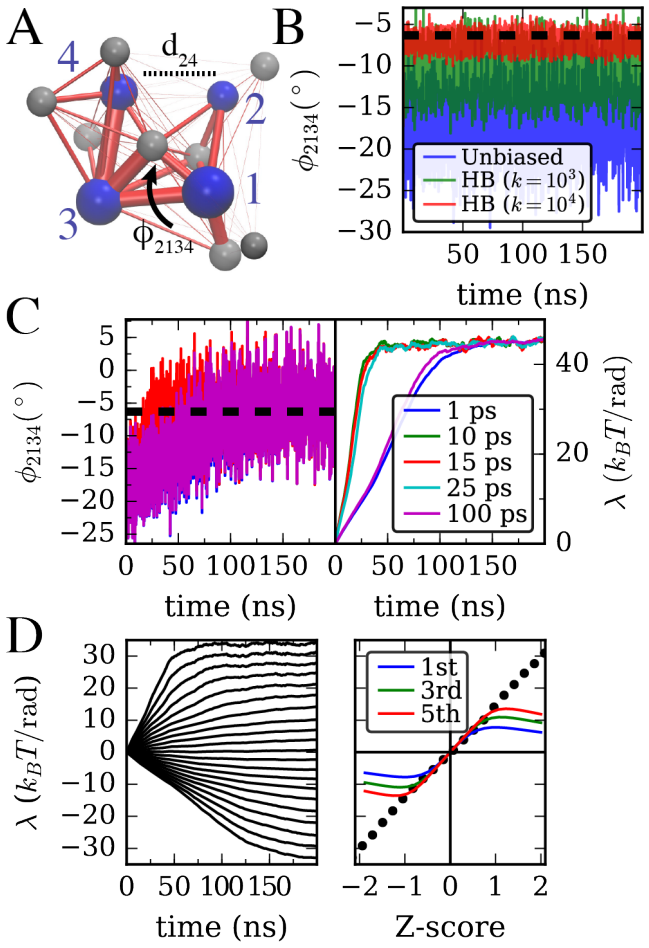

Figure 4-2: (A) Twelve-site hENM of an ATP-bound actin monomer parameterized as described in the main text. The four major subdomains of actin are labeled, and cleft distance and twist angle CVs are defined as in Figure A. (B) Twist angle for an unbiased hENM, as well as with a harmonic bias with force constants 10^3 and 10^4 kJ/mol/rad^2 centered at $\bar{\phi} = -6.3°$ (dashed line). (C) Twist angle evolution as well as biasing parameter using gradient descent algorithm of Ref. 8 is shown for different τ^{avg}. (D) Left, bias parameter as in C (τ^{avg} = 10ps) with target values for ϕ from -25.2º to -9.16º. Right, comparison of final bias parameters on left (dots) with first, third, and fifth order predictions given in the main text. The horizontal axis shows difference of target ϕ from $\phi_{unbiased}$ scaled by the unbiased standard deviation as computed from data in (B).

4.2.3 Relative entropy minimization

In this work, the objective is for our system to evolve under minimally biased dynamics such that average value of particular CG observables of a subsystem of interest match target values obtained via simulations of a larger encompassing supersystem (or alternatively values obtained from experiment). In principle, as in previous work, we can derive the necessary change to the system Hamiltonian (H_0) by minimizing the relative entropy between the distribution normally sampled by that system, $P_0(X)$, and the distribution, $P(X)$, arising from an unknown Hamiltonian (H). The latter unknown Hamiltonian system is subject to constraints on a mean of a set of observables, $\{f_i(X)\}$, which are scalar functions of the configuration of the system and are known properties of the system described by the unknown Hamiltonian. In other words, we want to minimize the functional

$$S\left[P(X)\right] = \int dX P(X) \ln\left(\frac{P(X)}{P_0(X)}\right), \tag{4.3}$$

with the constraints $\int dX\, P(X) = 1$, and $\int dX\, f_i(X) P(X) = \bar{f_i}$. This is formally solved by introducing Lagrange multipliers, $\{\lambda_i\}$:

$$P(X) = P_0(X) e^{\beta\left(\lambda_0 + \sum_i \lambda_i f_i(X)\right)}, \tag{4.4}$$

with $\beta = 1/k_B T$. Using the first normalization constraint to set λ_0, and taking P_0 from the canonical ensemble gives the result:

$$P(X) = P_0(X)e^{\beta\left(\lambda_0 + \sum_i \lambda_i f_i(X)\right)}, \tag{4.5}$$

By comparison with the usual canonical ensemble distribution function, we see here that in order to simulate our system and have it maintain particular target values of our observables, $\{\bar{f}_i\}$, i.e., that are manifest in a simulation (or experiment) for the full and larger encompassing supersystem, we must modify the Hamiltonian H_0 of our system to include a linear term for each observable with (at this point) unknown proportionality constants $\{\lambda_i\}$.

Several papers have offered suggestions for how to determine these parameters in a molecular simulation context. In each case, the idea is to iteratively update the values of $\{\lambda_i\}$ proportional to the difference $\Delta_i(\boldsymbol{\lambda}^t) = \langle f_i(X)\rangle_t - \bar{f}_i$, where the average is taken by sampling for a time τ^{avg} using the Lagrange multipliers from the t^{th} iteration. The rule for updating $\boldsymbol{\lambda} = \{\lambda_i\}$ (bold font indicating vectors, double-bars indicating matrices, dots indicating vector matrix multiplication, other vector arithmetic is element-wise) then takes a form similar to:

$$\lambda^{t+1} = \lambda^t - \eta^t\left(\Delta\left(\lambda^t\right)\cdot\left(\frac{\partial\Delta}{\partial\lambda^t}\right)\right) = \lambda^t - \eta^t\delta^t, \tag{4.6}$$

where $n = 1$ corresponds to gradient descent,[31] $n = 0$ a scheme that only depends on the current distance from the average observables,[91] and $n = -1$ to Newton's method, and the error function being function being minimized is the total squared difference from observed to target CV values, i.e. $\epsilon_t = \sum_i \Delta_i(\boldsymbol{\lambda}^t)^2$.[90] The way that the pre-factor $\boldsymbol{\eta}^t$ is adjusted corresponds to a learning rate rule, that is, how much trust to ascribe to the step size coming from the other λ-dependent terms. For example, Bussi and coworkers use:[91]

$$\eta_i^t = \frac{A_i}{1+\frac{t}{\tau_i}}, \tag{4.7}$$

while White and Voth use:[31, 92]

$$\eta_i^t = \frac{A_i}{\sqrt{\sum_{j=1}^{t}\left(\delta_i^t\right)^2}}. \tag{4.8}$$

[31, 89-91]

For the linear bias, the gradient term is a matrix with entries given by:

$$\left(\frac{\partial \Delta}{\partial \lambda^t}\right)_{ij} = -\left\langle f_i f_j \right\rangle_t + \left\langle f_i \right\rangle_t \left\langle f_j \right\rangle_t . \tag{4.9}$$

so, the gradient is proportional to the covariance of the two observables on iteration *t*:

$$\left(\frac{\partial \Delta}{\partial \lambda^t}\right)_{ij} = -Cov\left(f_i, f_j\right) \equiv \bar{J} . \tag{4.10}$$

In all cases where these methods have been applied in a molecular context, a stochastic procedure was used where one Lagrange multiplier was chosen randomly and adjusted based on this recipe. Below, we will use the learning rate rule of White and Voth, but show that for the type of problems we are interested in, rather than the stochastic gradient descent (SGD) a simultaneous adjustment of all the parameters using the full covariance, either with gradient descent (Covar) or a Levenberg-Marquardt-type algorithm (LM) as suggested previously (see below)[107] can be superior, and prevent the need for tuning of the constant factor A_i for each observable. (Note that in the work of White and Voth and in this implementation, the constant A_i is scaled by the target observable value, so the A_i in this general formulation is replaced by $A_i = 2A_i'/\bar{f}_i$, with A_i' the value set as a RANGE parameter in the current Plumed2 implementation. The bias parameters λ_i always have

units [Energy]/[CV], so in this way, A_i' has units of [Energy]. These are the values reported simply as *A* below.)In the LM algorithm, the step size in Eq. 4.6 is replaced by

$$\delta^t = \left(\left[\bar{J}^T\bar{J} + \gamma^t diag\left(\bar{J}^T\bar{J}\right)\right]\bar{J}^T\right)\cdot\Delta\left(\lambda^t\right), \tag{4.11}$$

where $\mathrm{diag}\left(\bar{\bar{J}}^T\bar{\bar{J}}\right)$ is the purely diagonal matrix having the same diagonal elements as $\bar{\bar{J}}^T\bar{\bar{J}}$, and γ^t is a mixing parameter, and causes the method to behave with some character of Newton's method ($\gamma = 0$) and steepest-descent ($\gamma \gg 1$).[107] We have implemented an adaptive version of the algorithm where we average the total squared error ϵ_t for all CVs over *LM_stride* iterations of the algorithm. This quantity is retained over *m* windows of length *LM_stride*$*\tau^{\mathrm{avg}}$. If the error decreases monotonically for these *m* stages, γ is increased by a multiplicative factor l >1, i.e. $\gamma^{t+1} = l\gamma^t$. If it decreases in each stage, $\gamma^{t+1} = \gamma^t/l$. If it is not monotonic over *m* stages, then γ is not changed.

All algorithmic parameters for each figure are listed in We make a practical note here, that in order to reduce the overhead of biasing via PLUMED, CVs are defined using only two atoms per protein residue (C_α, C_β) and for all-atom simulations we use the built-in multiple time stepping procedure of Bussi and coworkers,[108] with the bias algorithm performed every-other MD time step (STRIDE=2).

Fig / System	Algorithm	Parameters
Fig. **4-**2B / hENM	HB	$k_\phi = 10^3, 10^4 \frac{\text{kJ}}{\text{mol rad}^2}, \bar{\phi} = -6.32°$
Fig. **4-**2C / hENM	SGD	$A = 10k_BT, \bar{\phi} = -6.32, \tau^{\text{avg}}$ = from 1-100 ps
Fig. **4-**2D / hENM	SGD	$A = 10k_BT$, τ^{avg} = 10 ps, $\bar{\phi}$:-14.99.16° to -22.95.2° ($\langle\phi\rangle \approx -17°, \sigma_\phi \approx 4°$)
Fig. **4-**3A (top) / hENM	SGD	$\tau^{\text{avg}} = 10\text{ps}, \text{A} = 10\text{k}_\text{B}\text{T}, \bar{\phi} = -6.32°, \bar{d} = 20.62\text{Å}$
Fig. **4-**3A (top) / hENM	Covar	See above, but $\text{A} = 1\text{k}_\text{B}\text{T}$
Fig. **4-**3A (bottom) / hENM	SGD	$\tau^{\text{avg}} = 100\text{ps}, \text{A} = 10\text{k}_\text{B}\text{T},$ $\bar{\phi} = -6.32°, \overline{\phi^2} = 40.86^{(°)^2}, \bar{d} = 20.62\text{Å}, \overline{d^2} = 425.03\text{Å}^2$
Fig. **4-**3A (top) / hENM	Covar	See above, but $\text{A} = 1\text{k}_\text{B}\text{T}$
Fig. **4-**3B / hENM	Covar	$\tau^{\text{avg}} = 100\text{ps}, \text{A} = 1\text{k}_\text{B}\text{T},$ $\bar{\phi} = -6.32°, \overline{\phi^2} = 40.86°^2, \bar{d} = 20.62\text{Å}, \overline{d^2} = 425.03\text{Å}^2$
	LM	See above, with $\gamma = 0.1, 0.01$
	Adaptive LM	See above, $\gamma^0 = 0.1,\ m = 3, l = 1.2, LM\_stride = 10$
Fig. **4-**4A (top), 4B / G-monomer	HB	$k_d = k_\phi = 5000 \frac{\text{kcal}}{\text{mol rad}^2}, \bar{\phi} = -6.32°, \bar{d} = 20.62\text{Å}$
	Adaptive LM	$\tau^{\text{avg}} = 100\text{ps}, \text{A} = 1\text{k}_\text{B}\text{T}, \quad \gamma^0 = 0.01$ $\bar{\phi} = -6.32°, \overline{\phi^2} = 40.86^{(°)^2}, \bar{d} = 20.62\text{Å}, \overline{d^2} = 425.03\text{Å}^2,$ $m = 3, l = 1.2, LM\_stride = 100$
Fig. **4-**4A (bottom) / F-monomer	Adaptive LM	See above, except $\gamma^0 = 0.1$
	Frozen bias (after 80 ns of above)	$\lambda_\phi = -10.2 \frac{\text{kcal}}{\text{mol rad}}, \lambda_{\phi^2} = 484.0 \frac{\text{kcal}}{\text{mol rad}^2},$ $\lambda_d = 2.2 \frac{\text{kcal}}{\text{mol Å}}, \lambda_{d^2} = 6.63 \frac{\text{kcal}}{\text{mol Å}^2}$
Fig. **4-**5B,C	LM	$\tau^{\text{avg}} = 100\text{ps}, \text{A} = 1\text{k}_\text{B}\text{T}, \quad \gamma^0 = 0.01$ $\overline{\phi_{A1}} = \overline{\phi_{A2}} = \overline{\phi_{A3}} = -6.32°,$ $\overline{d_{A1}} = \overline{d_{A2}} = \overline{d_{A3}} = 20.6\ \text{Å}$

Table 4-1: Parameters used in CGDS algorithm for determining the fit parameters.

4.3 Results

4.3.1 Actin monomers in solution are more twisted and open, with larger fluctuations than subunits in a filament

As described in the methods section, MD simulations of an actin filament and actin monomers from two different starting configurations were performed. The difference between the G-actin and F-actin structure can be well characterized by defining the cleft distance (d) and twist angle (ϕ) as shown in A.[94, 102] In this study, CG bead positions used for analysis and for biasing are defined by computing the center of mass of the C_α and C_β in each actin subdomain. Using these definitions, we computed the value of actin monomer twist angle and cleft distances in our MD simulations. In B, we show that, as in previous simulations and experimental studies, an actin subunit in a filament is flat ($\phi \geq -10°$) and closed ($d \lessapprox 21$ Å) [94-95]while a G-actin with bound ATP is twisted, ($\phi \lessapprox -20°$) and more open ($d \gtrapprox 25$ Å).[95, 103] Interestingly, in our simulation of an ATP-bound monomer starting from a filamentous configuration, the monomer readily adopts a G-actin like structure over the course of a relatively short MD simulation, suggesting that the twisted/open structure has a lower free-energy outside of a filamentous context, as expected based on known actin biology.[94] Finally, we observe that the fluctuations of these two CVs for an actin monomer in solution are larger than in a filament, where allosteric coupling to adjacent actin subunits constraints the range of conformations (see.

4.3.2 Linear bias method can out-perform harmonic bias for matching target CV value

Unconstrained simulations of our hENM test system show a ϕ angle oscillating around the atomistic G-actin value from Applying a harmonic bias (HB) to the hENM model of the form $\frac{1}{2}k_\phi\left(\phi-\bar{\phi}\right)^2$ with spring constant $k_\phi = 10^3$ kJ/mol/rad^2 and even $k_\phi = 10^4$ kJ/mol/rad^2 centered at $\bar{\phi} = -6.3°$ moves the sampled value of ϕ closer to but not all the way to this target value. As discussed previously, applying this bias also substantially changes the size of fluctuations of this observable.[89] In contrast, shows that when using the adaptive algorithm of Ref., the target average is achieved exactly, although it takes some time to learn the bias parameter.

4.3.3 Adaptive algorithm has optimal averaging time for 1 CV

The adaptive algorithms discussed in the Methods section sample observable values over a time τ^{avg} before adjusting the bias parameter based on a learning rule. In we show in simulations of the hENM actin model that the convergence of the gradient descent algorithm can be very sensitive to the value of τ^{avg}, and appears to have an optimal value, in this case approximately 10-15 ps. Very short averaging windows produce a poor average of the variance of the observable used in the update rule, and very long averaging time produces a better average, but too long is spent on this process. An idea of how to set this can come from computing the autocorrelation function (ACF) for the CV in an unbiased simulation. For the twist angle, we estimate the autocorrelation time (when the ACF decays to 1/*e*) at ~ 4 ps. Hence to get a good estimate for the average and variance of this CV used in the algorithm, it seems that at least 2 autocorrelation times are needed. The speed of convergence can also be adjusted by changing the constant *A* in the learning

rule (in this case all simulations used A = 10 k_BT), but we have found that having too large of an initial value of A can cause the bias parameters to overshoot, hence it is better to choose good values for τ^{avg} first, as this has a non-linear effect on the learning time.

4.3.4 Derivation and validity of a linear response approximation to bias parameters

If the target value for a CV is close to the value in an unbiased simulation, then we know that λ_i for that CV will be small. Hence, we can expand the exponential as a Taylor series around $\lambda_i = 0$. We will illustrate this derivation first for a single CV. From Eq. 4-5, we can see that when the correct Lagrange multiplier has converged for a CV f':

$$\overline{f}' = \left\langle f' \right\rangle_\lambda = \frac{\int dX f'(X) e^{-\beta\left(H_0(X)+\lambda f'(X)\right)}}{\int dX e^{-\beta\left(H_0(X)+\lambda f'(X)\right)}}. \tag{4.12}$$

Without loss of generality, we can consider the CV $f \equiv f' - \overline{f}'$, with target value $\overline{f} = 0$. Hence for this new CV, the left-hand side of this equation is zero and the denominator on the right-hand side, as a constant, does not affect the equality. We can then expand the exponential in the numerator and write:

$$0 = \int dX f(X) e^{-\beta H_0(X)} \left(1 - \beta\lambda f(X) + \frac{(-\beta\lambda)^2}{2} f^2(X) + \ldots \right). \tag{4.13}$$

Dividing both sides of the equation by the unbiased partition function $Z = \int dX e^{-\beta H_0(X)}$, we can express each term as an average that can be obtained in the unbiased simulation. This gives an equation to be solved for the unknown constant λ.

$$0 = \left\langle f \right\rangle - \beta\lambda \left\langle f^2 \right\rangle + \frac{(-\beta\lambda)^2}{2} \left\langle f^3 \right\rangle + \cdots = \sum_{j=0}^{\infty} \frac{(-\beta\lambda)^j \left\langle f^{j+1} \right\rangle}{j!}. \tag{4.14}$$

If the averages $\langle f^j \rangle$ can be computed accurately from an unbiased simulation, then this equation can be truncated at some power and solved for λ. At first order, this equation is trivially solved:

$$\lambda = k_B T \frac{\langle f \rangle}{\langle f^2 \rangle} = k_B T \frac{\langle f - \bar{f}' \rangle}{\left\langle \left(f' - \bar{f}' \right)^2 \right\rangle}. \tag{4.15}$$

For N CVs $\{f_i\}$ centered at the target value, we get N equations:

$$0 = \int dX f_i(X) e^{-\beta H_0(X)} \prod_{k=1}^{N} \sum_{j=0}^{\infty} \frac{\left(-\beta \lambda_k\right)^j \langle f_k^j \rangle}{j!}. \tag{4.16}$$

At first order, we get the multi-dimensional equivalent expression to equation 15:

$$\lambda = k_B T \operatorname{cov}\left(f_i, f_j\right)^{-1} \cdot \langle f \rangle. \tag{4.17}$$

This solution is equivalent to taking a single step of Newton's method (see, equation 10 and 11).

These results are tested on the hENM model in. In we vary the target value for the twist angle in an actin monomer from above to below the average ϕ. We compute the moments $\langle (\phi - \bar{\phi})^n \rangle$ and solve Eq. 4-14 numerically truncating from first to fifth order in λ. We find for the distribution of ϕ generated by our hENM model, the even-ordered solutions can be purely imaginary. However, the odd ordered equations always have a real solution and these are shown on the right as a function of how far the target is from the unbiased mean, scaled by the standard deviation, $Z = (\bar{\phi} - \langle \phi \rangle)/\sqrt{\langle (\phi - \langle \phi \rangle)^2 \rangle}$.

As expected, for target values close to the unbiased average, the bias parameter is very small. We observe that the first order approximation breaks down near $Z = \pm 0.5$ and the

fifth order solution breaks down around $Z = \pm 1$. The linearity of the learned bias parameters in this range is likely a consequence of applying this method to a purely harmonic test system (although the cost of twisting ϕ is not perfectly quadradic).

Given the simplicity of computing the first order solution (even for *N* CVs), we suggest that future practitioners of these relative entropy minimization algorithms start with this as an initial guess. Since it arises from a linear response approximation, this amount of bias is unlikely to cause an irreversible change in the system when activated. However, in the rest of the data below, this is not done so as to show the full process of the learning algorithm.

4.3.5 Simultaneous update of bias parameters outperforms stochastic gradient descent

To fulfill our goal outlined in the introduction, we need to bias the actin monomer to behave as if it is in a filament. To do this, we will attempt to simultaneously bias four CVs: the twist angle and cleft distance, as well as their second moments (fluctuations). As discussed before, the optimal τ^{avg} for a CV is related to that CV's auto-correlation time. The problem that can be encountered when going to multiple CVs is that there can be a large timescale separation. This forces a user of the method to use a τ^{avg} as big as needed for the slowest-to-average CV. In the stochastic algorithm, only one CV is updated at a time, and so it is possible to have many long averaging periods in a row where only the faster averaging CV's bias parameter is updated. In (top), we first show in simulations of the hENM model that the SGD algorithm is effective for biasing the average cleft distance and twist angle. However, in order to bias these parameters and their second moments, a longer averaging time is needed, making convergence much

slower. We show in Figure 4-3A(bottom) that in this case, using the full covariance matrix as in Eq. 4-10 greatly accelerates convergence (note that this is after choosing a 10*x* lower learning rate parameter A for the covariance calculation).

4.3.6 Levenberg-Marquardt algorithm greatly improves on gradient descent

Since we were no longer using the stochastic version of the optimization algorithm, we modified our learning algorithm to include the LM update rule.[107] In we compare results of LM to the covariance version of our algorithm. We see that the LM algorithm with $\gamma = 0.01$ accelerates convergence by another factor of >10x (~50 vs ~500 ns) over the covariance gradient descent. This allows us to easily see the difference in speed between the two algorithms. This also demonstrates that when using the more "intelligent" step sizes of the LM algorithm, we no longer find it necessary to tune A to get fast convergence. We also show one simulation using the adaptive algorithm described above. In this case, we see that the adaptive LM algorithm converges at about the same speed as the algorithm with its fixed initial γ. However, we note that this method may be useful in atomistic contexts (such as discussed in the following sections), where it could be advantageous to for this this confidence parameter γ to change as the atomistic system movies from one state to another.

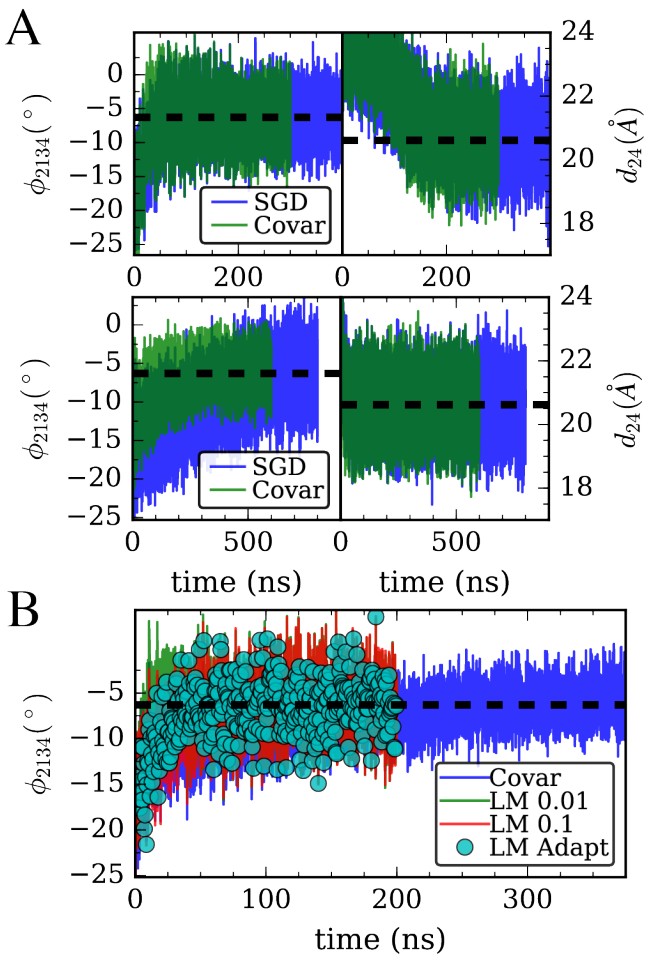

Figure 4-3. (A) Simulation of hENM model using stochastic gradient descent (blue) and full covariance matrix (green) to bias the two shown CVs as well as their variance, with otherwise identical algorithmic parameters, (B) The full covariance method is compared to the Levenberg-Marquardt (LM) algorithm with $\gamma = 0.1, \gamma = 0.01,$ and the adaptive algorithm with starting $\gamma = 0.1$.

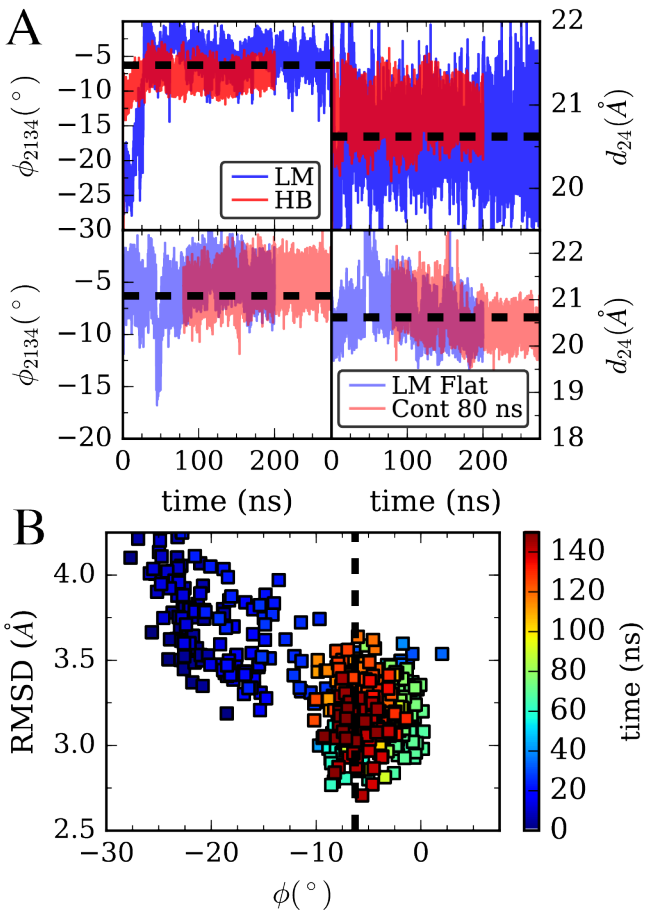

Figure 4-4. (A) Top, Adaptive Levenberg-Marquardt (LM) algorithm matching 4 CVs: cleft distance, twist angle, and their variances is compared to harmonic bias on angle and distance with large spring constants on both. Data is for all-atom MD simulations of the G/monomer system in Figure 4-1. Bottom, in blue, the LM algorithm is performed on an actin monomer starting from a filament structure (F/monomer in Figure 4-1). In red, the bias parameters at time 80 ns are fixed and a separate simulation is run using this learned bias. **(B)** Comparison of the structure of the G/monomer in the LM trajectory from (A,top) with a filament subunit by backbone RMSD. Color shows progress along the trajectory in (A).

4.3.7 Linear bias on monomer CVs is effective as a restraint, matches structure in filament

Having now developed and implemented improved learning algorithms, we sought to test these methods on the all-atom MD simulation of G-actin and F-actin monomers. The four CVs, cleft distance and angle, as well as their second moments, were targeted to match the values from a single actin monomer in the 13-mer simulation shown in B (see). The LM algorithm of all various flavors were found to converge in reasonable amounts of simulation time, with the amount of time required to match all CVs depending upon the particular parameters. In the upper panels of A, we show a simulation using the adaptive LM that matches the target CVs in ~100ns, with a small error in all 4 CVs (see. We note that given the moderately large expense of simulating this system, we did not try to optimize the parameters used much beyond what was learned from the simpler hENM, besides increasing τ^{avg} to account for the slower fluctuations in the atomistic system. Nevertheless, the algorithms showed themselves to be robust, converging eventually in all trials of the LM algorithm with large enough τ^{avg}. The LM results in are compared to a harmonically biased simulation with a very large spring constant applied to both the twist angle and cleft distance CVs. Despite being ~10*x* larger than what was used previously in a QM/MM application,[86-87] the harmonic bias (starting from the G-actin structure) fails to match the target twist angle. In, we compare the structure of this biased G-actin monomer to a filament subunit configuration. We see that as the algorithm progresses, the G-actin structure (which is already similar to an F-actin monomer structure) converges to a structure that has a lower RMSD to the target, although RMSD was not an observable explicitly biased. We note that previous simulation studies have

shown that actin filaments are heterogeneous,[101] and the RMSD of an actin subunit in a filament compared to another or compared to itself later in an MD simulation is typically in this range (for the last 100ns of the simulation in, the subunit-subunit RMSD is 3.35 ± 0.30 Å). Hence, we cannot expect to achieve a closer match than what is seen in B without explicitly biasing RMSD.

In the lower panels of the adaptive LM algorithm is used on a simulation of an actin monomer starting from a flat, filamentous like structure. In this case, it can be seen that around 50 ns, the structure begins to twist and open, as was seen in the unbiased simulation in. Remarkably, the bias algorithm is able to adapt to this change in observable and return the configuration to close to the target value. In other trials using the stochastic gradient descent and covariance gradient descent, the convergence was not fast enough to prevent the structure from twisting, and then it took a much longer time for the system to return to a flattened structure. We note that during this 200ns simulation, convergence is not yet achieved, meaning that there is greater error here/slower convergence to the target CV values than starting from the twisted structure. This seems to be a general trend, and we will consider this idea further in the Discussion section. Finally, in these lower panels we show the result of freezing the bias that has been learned at 80ns from the flat-monomer (before convergence) and show that the simulation approximately maintains the CV values from that starting time.

Fig / System + method	$\langle d_{24} \rangle$ (Å)	$\langle d_{24}^2 \rangle$ (Å^2)	$\langle \phi_{2134} \rangle$ (°)	$\langle \phi_{2134}^2 \rangle$ (°^{2})
Fig. 4-1B / F-actin (1 subunit)	**20.62 (0.0%)**	**425.03 (0.0%)**	**-6.32 (-0.0%)**	**40.86 (0.0%)**
Fig. **4-**1B / G-monomer	23.82 (15.5%)	568.15 (33.7%)	-16.93 (168.0%)	290.70 (611.4%)
Fig. **4-**1B / F-monomer	26.52 (28.6%)	704.00 (65.6%)	-17.37 (175.0%)	307.00 (651.3%)
Fig. **4-**2B / hENM (no bias)	23.95 (16.2%)	574.24 (35.1%)	-17.44 (176.1%)	317.18 (676.2%)
Fig. **4-**2B / hENM HB K=1000	23.79 (15.4%)	566.72 (33.3%)	-10.54 (66.8%)	116.27 (184.5%)
Fig. **4-**2B / hENM HB K=10000	23.82 (15.5%)	568.04 (33.6%)	-6.95 (9.9%)	49.07 (20.1%)
Fig. **4-**2C / hENM SGD τ^{avg} = 15 ps	23.88 (15.8%)	570.98 (34.3%)	-6.30 (-0.2%)	54.97 (34.5%)
Fig. **4-**3A / hENM SGD 2 CV	20.62 (-0.0%)	425.72 (0.2%)	-6.48 (2.6%)	52.02 (27.3%)
Fig. **4-**3A / hENM Covar 2 CV	20.65 (0.2%)	427.28 (0.5%)	-6.53 (3.4%)	52.65 (28.8%)
Fig. **4-**3A / hENM SGD 4 CV	20.60 (-0.1%)	425.03 (-0.0%)	-6.50 (2.8%)	50.99 (24.8%)
Fig. **4-**3A&B / hENM Covar 4 CV	20.60 (-0.1%)	424.88 (-0.0%)	-5.99 (-5.2%)	40.03 (-2.0%)
Fig. 4-3B / hENM LM $\boldsymbol{\gamma = 0.1}$	20.62 (0.0%)	425.95 (0.2%)	-6.04 (-4.3%)	44.52 (8.9%)
Fig. 4-3B / hENM LM $\boldsymbol{\gamma = 0.01}$	20.63 (0.0%)	426.02 (0.2%)	-6.07 (-4.0%)	44.63 (9.2%)
Fig. 4-3B / hENM LM Adapt $\boldsymbol{\gamma^0 = 0.1}$	20.62 (0.0%)	425.94 (0.2%)	-6.02 (-4.7%)	43.97 (7.6%)
Fig. **4-**4A / G-monomer HB 2 CV	20.87 (1.3%)	435.78 (2.5%)	-7.37 (16.7%)	55.38 (35.5%)
Fig. **4-**4A / G-monomer LM Adapt 4 CV	20.55 (-0.3%)	422.51 (-0.6%)	-6.04 (-4.4%)	38.03 (-6.9%)
Fig. **4-**4A / F-monomer LM Adapt 4 CV	20.44 (-0.8%)	418.08 (-1.6%)	-5.11 (-19.1%)	28.01 (-31.5%)

Table 4-2: Observed values for collective variables parameters**.** Quantities are computed for the final 50ns shown in each figure. Percentages are comparison with respect to a single actin monomer, data on the first line of the table (bold). Biased parameters are underlined.

Fig. **4-**4A / F-monomer Fixed Bias 4 CV	20.32 (-1.4%)	412.95 (-2.8%)	-4.37 (-30.9%)	20.56 (-49.7%)
Fig. **4-**5 / Trimer-A1 LM Adapt	20.32 (-1.4%)	412.90 (-2.9%)	-6.21 (-1.7%)	40.78 (-0.2%)
Fig. **4-**5 / Trimer-A2 LM Adapt	20.53 (-0.4%)	421.68 (-0.8%)	-7.03 (11.3%)	51.62 (26.3%)
Fig. **4-**5 / Trimer-A3 LM Adapt	20.53 (-0.4%)	421.39 (-0.9%)	-6.75 (6.8%)	47.31 (15.8%)

Table 4-2 continuted

4.3.8 Biasing CVs in a larger subsystem is an effective alternative constrained moiety

When considering the problem of biasing a subsystem to represent aspects of its behavior within a larger supersystem, in general there will be many possible subsystems of different sizes to consider using. In most cases, it would be advantageous to choose the smallest system size possible, as we have done by choosing a single actin subunit in a filament. However, larger systems afford the advantage of representing a more native-like context; for example, including more protein-protein interfaces. To demonstrate this idea for actin, we tested our algorithms for an actin trimer which was previously used in a metadynamics study[100] of the nucleotide effect on the conformational states of actin (). In, we show data from a simulation where the cleft distance and dihedral angle in all three actin subunits are biased, but in this case only the mean values for these six CVs are set. The algorithm is able to find six parameters for which all the CVs closely match their target values (see. Interestingly, due to the allosteric coupling between monomers, variances of these six CVs are much closer to the full filament values than in the case of an unbiased monomer. This is reflected in where the dihedral angles of an unbiased trimer simulation are histogrammed and compared to the filament value, with the unbiased dihedral distributions having only ~35% larger standard deviations than an actin subunit in a filament (and the distance distribution being even closer, ~20% different, not

shown). When the bias is applied, then, the peak of the distribution shifts as desired while leaving the rest of the distribution approximately the same. The result is only ~10% error in the mean of twist angle and <15% error in the second moment (with distances having <3% error in both). Although a larger system, we believe this kind of tradeoff between size and additional molecular context could be appropriate in some future applications of our GCDS methodology.

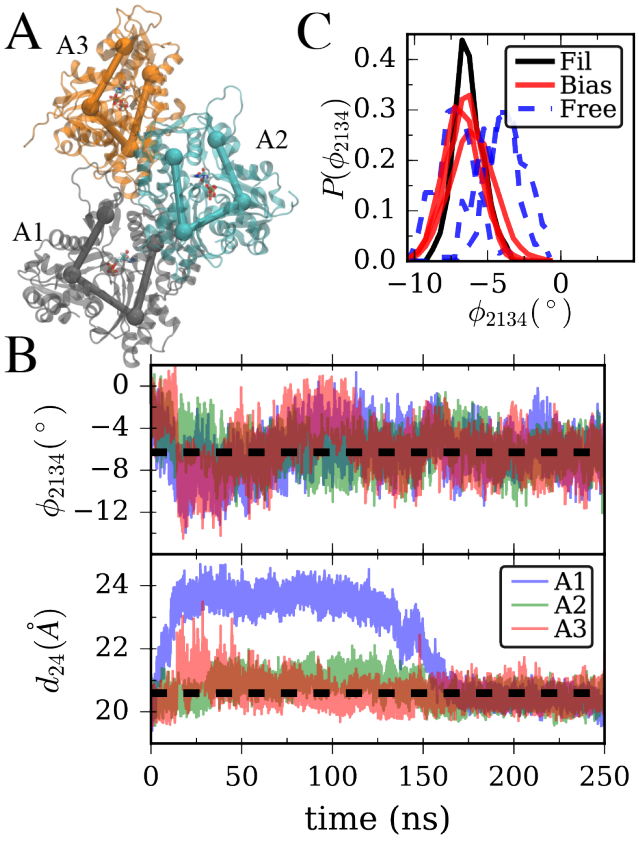

Figure 4-5. (A) Illustration of all-atom three actin sub-filament with CG subdomains from Figure A overlaid. (B) Twist angle and cleft distance for each of the subunits in (A) during LM bias simulation. (C) Observed distribution of twist angles in the final 50ns of an unbiased 100ns simulation of the structure in (A) (dashed line) vs. the final 50ns of the biased simulation with data plotted in (B) and the final 50 ns of filament data from Figure B.

4.4 Discussion and Conclusions

In this work, we present the idea that linear biases learned by a relative entropy minimization scheme can be used to restrain a molecular system in such a way that it retains some information about its behavior within a larger macromolecular "supersystem" context. In particular, we show that in the case of actin, we can bias two

CG observables and their second moments, with the result being that an actin monomer can adopt and maintain a filament-like conformation with native-like fluctuations in these observables. Although the CGDS algorithm takes some time as compared to simply applying a harmonic bias, these methods can achieve a closer match for the distribution of the target observables. We have also made a number of algorithmic improvements which for this application greatly reduce the amount of sampling needed to learn the biasing parameters, down to a very reasonable amount of simulation time (<100ns), and produce systems that would be appropriate for use in subsequent free energy or reactive (QM/MM or otherwise) MD simulations where a small subsystem is required but including the effects of larger scale fluctuations are expected to affect the results.

In previous studies from our group, the stochastic gradient descent algorithm was sufficient to learn biasing parameters in a very short amount of simulation time.[31, 92] We note that the context we are presenting here is very different, where within the simulation the observables to be biased depend on the positions of a few CG observables. In the previous studies, the systems of interest were isotropic liquids, and the CVs of interest (averages over radial distribution functions) depend on the pairwise distance between each molecule. This produces a self-averaging such that the effect of changing the biasing parameter can be sampled over many environments simultaneously. In the protein context, the relaxation time of the observable is much longer and moreover it is likely to be sampled only over a single copy of the system. Hence, we suggest that (1) starting from a linear response approximation to the bias parameters, (2) optimizing sampling time, and (3) using methods such as the Levenberg-Marquardt algorithm that take advantage of covariance and try to make "smart" step sizes based on that information, are

all important steps that should be applied in this (CGDS) context. Although we suspect these three steps are likely useful in the prior cases, in practice the learning time for previous applications was not a bottleneck.

Previously, there was some concern in the literature that using the full covariance matrix with correlated observables might result in an optimization problem that was non-convex and might not converge.[89] Although in this work we have biased both the first and second moment of our CVs, which are correlated, we have not found this to be a problem in practice. Another concern with using the full covariance matrix and a Newton-like method with correlated observables is that it might be singular and hence produce divergent step sizes in the iterative algorithm. As noted previously, the Levenberg-Marquardt term proportional to γ (Eq. 4-10) is specifically designed to avoid this problem.[107]

Finally, we return to one challenge encountered during our CGDS studies, which is that these adaptive algorithms are not well tailored to biasing a system that starts close to a target configuration in a state that is metastable on simulation time scales. An idealized free energy surface in such a case is likely to look something like the illustration in We in fact know this for an actin monomer from previous and ongoing umbrella sampling and metadynamics simulations using these CVs to obtain the free energy landscape for actin flattening. If the starting state of the system is a local free energy minimum, such as for an ATP-bound actin monomer in a flattened configuration, then the initial estimate for the Lagrange multiplier on that CV will be close to zero until fluctuations begin to cause the system to drift towards a deeper minimum. This is exactly what is observed in. On the other-hand, starting in another state is less desirable for two reasons: (1) starting from the

lower free energy minimum, the Lagrange multiplier estimate may be very large at first, and then it may take some time once the system is near the target state for the Lagrange multiplier to return to its fixed point, and (2) during this process, there is no guarantee that the simulation will find the target structure as the values of the CVs are improved (although we demonstrate in B that this was not an issue for the case of G-actin). This is a challenge for which we do not yet have a complete solution. Yet, we have found in our experience that biasing the second moment of the target CV in addition to the mean when starting from the higher free-energy state goes a long way towards solving this problem. We believe that this is because, as illustrated in, even if the minimum for the subsystem has the same mean as the target system, it is likely to be much less constrained, and hence will have a much wider basin in any coarse-grained CV such as the ones considered in this work.

The present paper does not claim to provide the final word in the development of CG "guided" or "directed" methods to bias all-atom MD simulations to "feel" some effects of being in a larger macromolecular supersystem. However, we assert that the CGDS method developed herein is an important first step in that direction. We will certainly improve and extend this approach in the future, and we encourage other researchers to contribute to this effort.

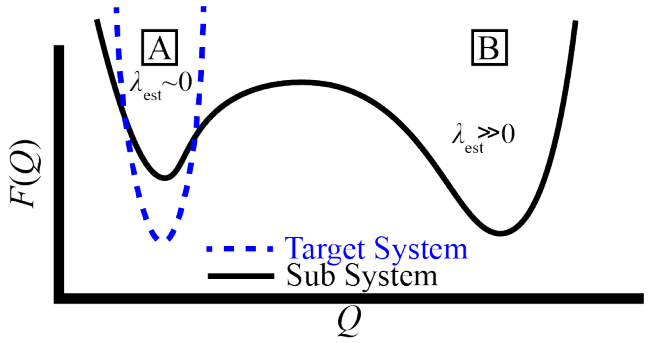

Figure 4-6. A structure of a target system (in this study, an actin filament) is likely to be known from experiment, and as such is in a relatively deep local free-energy minimum. Hence, the observed values for a CV (Q) are likely to be normally distributed around a single value (with a roughly harmonic potential of mean force $F(Q)$). When a sub-structure such as an actin monomer is removed to solution, the starting structure (A) will likely still be near a local free energy minimum, however there may be alternative lower free-energy configurations (B). The initially-estimated Lagrange multipliers needed to have the subsystem stay in state A will depend on whether the system starts in state A or B.

Chapter 5

Reactive Corase-grained Molecular Dynamics

5.1 Introduction

The study of chemical reaction is essential in the understanding of many liquids (e.g., electrolytes) and other materials. Reactions are known to influence properties such as diffusion, as in acidic aqueous solutions where the Grotthuss mechanism shuttles the hydrated excess proton between multiple water molecules.[109] In electrolyte solutions, charged ions can induce reactions in the electrolyte that break it down, while in homogeneous catalysis the aqueous environment can greatly influence or even participate in the pertinent chemical reactions.

In molecular dynamics (MD) simulation with standard empirical forcefields, a fixed bonding topology makes the simulation of phenomenon involving chemical reactivity impossible to study. The necessity of reactive features in MD simulations has lead, e.g., to the development of the multistate empirical valence bond (MS-EVB) method.[110-112] The theoretical basis for the MS-EVB model describes the bond-breaking and forming by a complex that couples multiple bonding "states", typically at least a product and reactant state, which are described with different bonding topologies and force-fields, inspired by the original EVB method as described by Weiss and Warshel.[30] The MS-EVB method has allowed, e.g., for the simulation of proton transport in many systems such as proton exchange membranes,[113] the influenza A M2 proton channel,[114] the proton pump cytochrome c oxidase, near platinum surfaces,[115] and at the water liquid-vapor interface. In more recent years, the MS-EVB method is being replaced by the more fundamental multiscale reactive molecular dynamics (MS-RMD) approach[116] in which

the interaction in the model are derived from quantum mechanical calculations via a variational force-matching approach.

While the MS-EVB and MS-RMD methods have addressed a number of molecular-scale chemically reactive problems, the large length and time scales associated with many important chemical problems (e.g., energy storage materials, membranes, etc.) make the use of (all-atom) reactive MD methods far more challenging and much less informative. One approach to overcoming this length and time scale limitation in molecular simulation is the use of coarse-grained (CG) modeling,[117] in which select degrees of freedom are eliminated from of the system. CG models have been used to investigate a number of phenomena such as membranes, membrane-protein interactions, and self-assembly of large multiprotein complexes

The extension of CG methods to the chemically reactive regime seems to be a natural one as the MS-RMD method has been case as a method for "coarse-graining away" electronic structure,[118] while certain CG methods have been extended to include multiple internal quantum-like states.[119] However, although one can imagine taking a more empirical and *ad hoc* approach for reactive CG modeling as in the MARTINI model,[120] to date none of the "bottom up" coarse-grained methodologies such as multi-scale coarse-graining (MS-CG),[12, 35, 121-122] iterative Boltzmann inversion (IBI),[123] and relative entropy minimization (REM),[124] which develop a CG model from the underlying molecular scale interactions, have been extended to include chemical reactions with dynamic bonding topologies. This is the focus of the present work, with the MS-CG method as the cornerstone of the new reactive CG methodology, although IBI or REM could also be adopted.

We first note that much of the previous work on reactions in low-resolution models has been carried out with dissipative particle dynamics (DPD) models, which creates a model where particles interact with conservative, dissipative, and fluctuating forces.[125] DPD incorporates reactive information by having multiple bonding configurations possible for particles, and the transitions between these particles is dictated by the probabilities of reactions when particles are close enough to one another.[126-129] The probability of reaction is typically estimated by relating the reaction probability to the rate of the reaction. This approach does not take into account the underlying molecular-scale nature of the reaction and so it must rely on approximations if the rate of the reaction is not known *a priori*. Furthermore, only reactions that happen much quicker than the timestep of the simulation can be realistically simulated. This DPD approach is also not parameterized in a multiscale manner that can directly account for the input from an AA simulation.

In this short chapter, an extension of the MS-CG method is presented called reactive multiscale coarse-graining (rMS-CG). In this approach, one can dynamically change the CG bonding topology during the progress of a CG MD simulation if one so desires. Moreover, by taking into account the AA potential of mean force (PMF) for the chemical reaction, two CG models (e.g., MS-CG) can coupled together such that they are able to reproduce the reactive AA PMF to a reasonable, but with (potentially) many fewer degrees of freedom than the fully AA case. Similar to previous reactive MD methods, an EVB-like approach is utilized where the diagonal elements of a two-state (or more than two if desired) Hamiltonian represents non-reactive CG models that are made independently of one another. Each of these diagonal CG models represents a product or

reactant of the reaction in question. The off-diagonal elements then couple the two CG states, i.e., provide a mechanism by which reactions between these two states can occur and have a bonding topology change at the CG level. Importantly, the coupling between the two states is a function of only a the CG collective variable (CV) that describes the progress of the reaction and is calculated in a way that captures the difference between the non-reactive CG models and the reactive AA PMF when mapped on to the CG coordinates. The result is a CG simulation that can reproduce the AA reactive process but using many fewer degrees of freedom than the atomistic simulation. This paper represents the first step towards introducting chemical reactivity into CG modeling and simulation.

5.2 Methods

To develop a reactive CG model with a dynamic bonding topology, the product and reactant states are treated as discrete states with separate force-fields and interaction potential energy surfaces, labeled V_{11} and V_{22}. In this current development, we are limited to treating only two CG bonding topologies for reactant and product, although the method will be generalized in the future to include additional states with corresponding bonding topologies as in the MS-EVB approach at the AA level. Each CG bonding topology can be attributed to a state with a "basis set" and the entire system can then be described using a quantum state-like linear combination of these states, the coefficients of which are determined by solving the eigenvalue problem,

$$\begin{pmatrix} V_{11}\left(\mathbf{R}^N\right) & V_{12}\left(\mathbf{R}^N\right) \\ V_{12}\left(\mathbf{R}^N\right) & V_{22}\left(\mathbf{R}^N\right) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \tag{5.1}$$

Where V_{11} and V_{22} correspond to the CG interaction potentials that would describe a non-reactive system. V_{12} corresponds to the coupling interaction potential between the reactant and product states. E is the lowest energy eigen solution of the 2 x 2 matrix system. In rMS-CG, V_{11} and V_{22} are determined from a non-reactive trajectory.

If there exists a nontrivial solution to the characteristic equation above, then the secular equation can be solved for the lowest energy eigenvalue

$$\begin{vmatrix} V_{11}\left(\mathbf{R}^N\right)-E & V_{12}\left(\mathbf{R}^N\right) \\ V_{12}\left(\mathbf{R}^N\right) & V_{22}\left(\mathbf{R}^N\right)-E \end{vmatrix} = 0 \tag{5.2}$$

Which can be expressed for the coupling term

$$V_{12}\left(\mathbf{R}^N\right) = \left[V_{11}\left(\mathbf{R}^N\right)-E\right]\left[V_{22}\left(\mathbf{R}^N\right)-E\right] \tag{5.3}$$

Next, we make the assumption that the interaction potential near the transition state is approximately equal to the free energy. As such, we assume that E is equal to the free energy along the reaction coordinate of the mapped atomistic system, that V_{11} and V_{22} are equal to free energy of the non-reactive CG system along the reaction coordinate, $F_{11}\left(\mathbf{Q}\right)$ and $F_{22}\left(\mathbf{Q}\right)$, and that V_{12} is a function along that reaction coordinate. This results in the equation

$$V_{12}\left(\mathbf{Q}\right)^2 = \left[F_{11}\left(\mathbf{Q}\right)-F_{AA}\left(\mathbf{Q}\right)\right]\left[F_{22}\left(\mathbf{Q}\right)-F_{AA}\left(\mathbf{Q}\right)\right], \tag{5.4}$$

where $F_{AA}\left(\mathbf{Q}\right)$ is determined directly from the atomistic simulations. By including the off-diagonal coupling between states in a way that is determined directly from the AA

PMF, the CG effective free energy surface of the rMS-CG model can reproduce the AA PMF in terms of CG CVs.

5.3 Models and Simulations

In order to demonstrate the rMS-CG methodology, an S_N2 reaction between 1-chlorobutane and a chloride ion was simulated. Initial configurations of a box containing 1 molecule of 1-chlorobutane, 1 chloride ion, 1 cesium ion, and 1000 methanol molecules were generated using Packmol.[130] The methanol, 1-chlorobutane molecules and ions were parameterized using the OPLS all-atom force-field.[47] For the non-reactive model, Morse potential parameters for the reactive bond between carbon and chlorine were calculated using MP2 calculations on NWChem.[131] The bond length of the carbon chlorine bond was extended to 0.28 nm from 0.18 nm in order to decrease the free energy barrier. All MD simulations were performed in LAMMPS.[73] The system was simulated with a timestep of 1 fs and used non-bonded Lennard-Jones interactions with a radial cutoff of 1 nm. Particle-particle particle-mesh (PPPM) was used to treat the long-range electrostatics.[74] The system was equilibrated for 100 ps at constant *NPT* at 1 atm and 300 K, and then simulated for an additional 100 ps at constant *NVT* at 300 K.[75]

Umbrella sampling runs were then performed from the equilibrated configuration. Umbrella sampling was performed using the PLUMED 2 plug-in with a custom bias CV generated using MATHEVAL library.[132] The bias CV used is given by

$$\mathbf{q} = \frac{\left(\mathbf{r}_{Cl} + \mathbf{r}_{Cl^-}\right)}{2} - \mathbf{r}_{Me}, \tag{5.5}$$

where $\mathbf{r}_{Me}$ is the center of mass of the reactive carbon and the two hydrogens bonded to it, which inspired the reactive CV used in the MS-EVB3 model.[110] It should be noted that

the $\mathbf{r}_{Me}$ vector used is a vector from a center of mass of a grouping of particles instead of just one particle. Each umbrella was run for 100 ps in order to equilibrate the bias, then statistics were collected every 250 fs for 2.5 ns.

CG models were created using a mapping of one-site per methanol molecule, one site per cesium ion, one site per chloride ion, one site for the reactive carbon and bonded hydrogens, one site for the bonded chlorine, and one site for the n-propyl group. Non-reactive CG models were parameterized from the non-reactive AA trajectory using the MS-CG methodology.[12, 35, 133] Since the bias CV used to calculate the free energies was in terms of coordinates resolvable in both the CG and AA models, the biased trajectories simply with the bias force removed was used to calculate the CG interactions. The pairwise interactions were calculated using sixth order spline basis function with a cutoff of 1.0 nm and basis resolution of 0.04 nm.

In order to calculate the reactive bonded interaction between the reactive carbon and chorine, additional AA umbrella sampling runs were performed with a bias CV of distance between the chlorine and the center of mass of the reactive carbon and bonded hydrogens. Bonded interactions were calculated with sixth order splines and basis set resolution of 0.008 nm while angle interactions were calculated using sixth order spline and basis set resolution of 5 degrees. Bonded interactions were then fit to harmonic interactions about the prominent mode for the reactive carbon-propyl bond and Morse interaction for the reactive bond. Angle interactions were also fit to a harmonic interaction about the prominent mode. An equilibration run of 1,000,000 CG timesteps was then performed using a mapped configuration from an AA run as a starting configuration. Umbrella sampling runs of the CG models for 5,000,00 CG timesteps per

umbrella were performed in order to get the non-reactive potential terms needed to calculate coupling. A similar bias CV to the one used in the AA, where the position of the reactive carbon CG site replaces the position of the reactive carbon, was used.

All AA and rMS-CG reactive runs were performed in a modified version of LAMMPS that was designed to perform reactive MD simulations.[134] A reactive AA model was generated from the non-reactive model by parameterizing the coupling term given by

$$A\left(R_{ClCl},\mathbf{q}\right)=\exp\left(-\gamma\mathbf{q}^2\right)\bullet\left\{1+P\exp\left[-k\left(R_{ClCl}-D_{ClCl}\right)\right]\right\}, \tag{5.6}$$

where $\mathbf{q}=\frac{\left(\mathbf{r}_{Cl}+\mathbf{r}_{Cl^-}\right)}{2}-\mathbf{r}_C$ and $R_{ClCl}=\left|r_{Cl}-r_{Cl^-}\right|$. Given the similarity between proton transport and the S_N2 reaction, parameters similar to the MS-EVB3 model were used where D_{ClCl} was adjusted to take into account the larger bond LJ sigma of the chloride ion compared to the water oxygen.[110] Reactive runs were created using the same equilibration procedure described above for the non-reactive runs. In order to for the umbrella sampling runs to more easily reach a steady state position near the equilibrium position of the bias, the reactive AA and CG simulations were started either in the reactant or product state, depending on which side of the transition state the equilibrium position of the umbrella was set to. Umbrella sampling runs were then performed using the same bias CV as the non-reactive umbrella sampling, equilibrating for 100 ps and then collecting statistics every 250 fs for 2.5 ns. Umbrella equilibrium positions and spring constants for the reactive runs are given in the supplemental information. A reactive coupling was calculated using equation 4, then was fit to a Gaussian. An rMS-CG run was then performed using umbrella sampling with the same bias CV as the non-reactive

CG runs. Each umbrella was equilibrated for 1,00,000 CG timesteps and then statistics were collected every 250 CG timesteps for 5,000,000 CG timesteps. All PMFs in are calculated using the weighted histogram analysis method (WHAM).[135]

5.4 Results and Discussion

In order to create an rMS-CG model , one must calculate a non-reactive CG model that describes the non-reactive AA system. If this is done in a bottom-up way, one could run an unbiased AA MD run and use the trajectory to parameterize the force-field of the CG model. However, this is often impractical in practice because some component of the system is poorly sampled, such as solute-solute interactions. In the case of the test system, interactions between the chloride ion and the 1-chlorobutane are a highly important part of the model, but complete sampling is impractical because these components of the system spend a large amount of time further than the interaction cutoff. To overcome this problem, umbrella sampling runs were performed such that solute particles were within the interaction cutoff range. Normally this can only be accomplished by reweighting each frame according to the bias applied to the configuration, however, it has been shown by Dama and Sinitskiy that biased trajectories could be used in MS-CG without any reweighting.[119] Using this, CG models of the solute molecules were constructed efficiently using umbrella sampling.

In the computational literature of S_N2 reactions, the typical reaction progress CV is given by the difference of the distance between each chlorine and reactive carbon,

$$\xi = r_1 - r_2, \tag{5.7}$$

where $r_1 = \left|\mathbf{r}_C - \mathbf{r}_{Cl}\right|$ and $r_2 = \left|\mathbf{r}_C - \mathbf{r}_{Cl^-}\right|$. While this reaction progress CV is typically sufficient for describing AA reactions, it is inadequate in describing the CG reaction

coordinate. In the AA simulations, the CV defined in using equation 6 as the reactive CV results in a backside attack because of the space excluded by the hydrogens off of the reactive carbon. However, since the hydrogens are mapped into the site with the reactive carbon, equation 7 is no longer sufficient to describe the progress of the reaction. In order to circumvent this problem, a new reaction progress CV defined by equation 6 was used. This CV is best described intuitively as the distance of the reactive carbon from the midpoint of the two chlorine atoms. The primary advantage of this reaction progress CV is that it takes into account the angular nature of the backside attack, which is necessary to properly describe the S_N2 reaction.

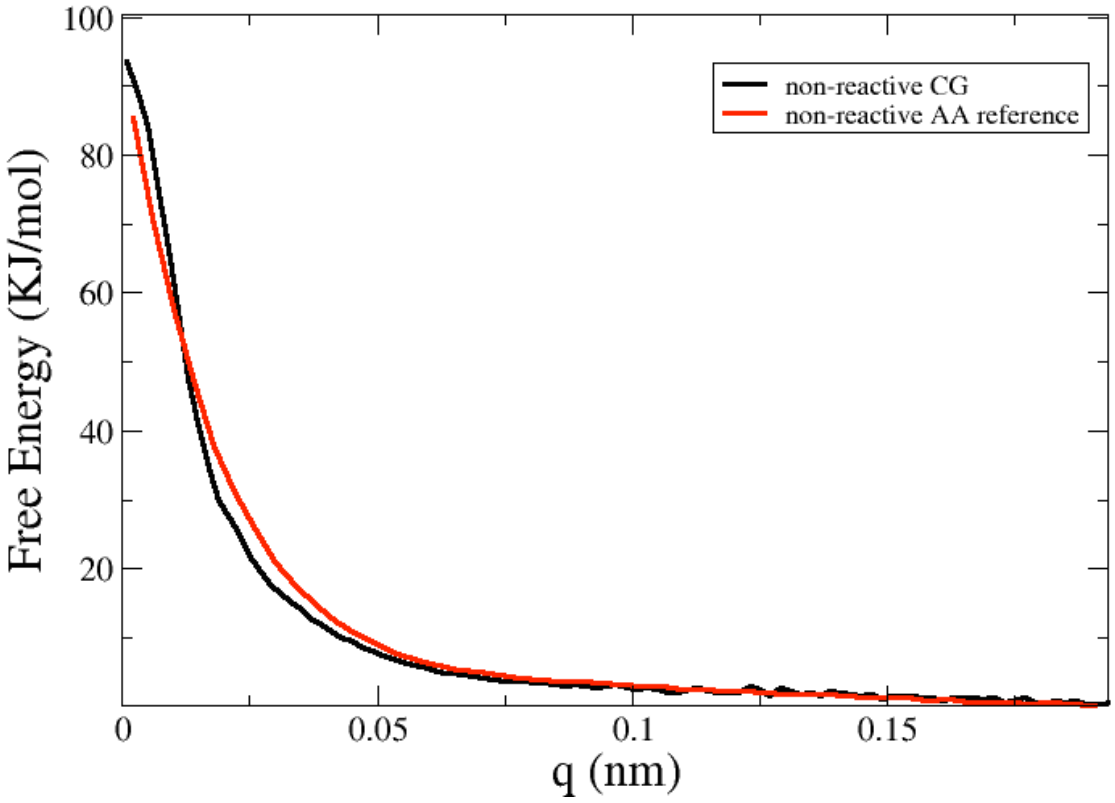

Figure 5-1: Comparison of the PMF of the non-reactive CG and non-reactive AA models along the CV that will serve as the reactive CV in the rMS-CG simulation.

A free energy along the reaction coordinate for the reactive and nonreactive models of the reactive system is shown in figure 1. Since the reaction is symmetric, the free energy for the V_{22} state was calculated by mirroring the free energy of V_{11} at q=0. The free energy of the non-reactive model was unable to be sampled with umbrella sampling below 0.002, so the free energy was linearly extrapolated beyond that point. A comparison of the free energy between the non-reactive AA and CG model is shown in figure 5-1. While there is some slight deviation in the free energy near q=0, the CG model has nearly the same hard wall distance where the free energy starts to more rapidly increase. Also, the free energy at q=0 are close to each other.

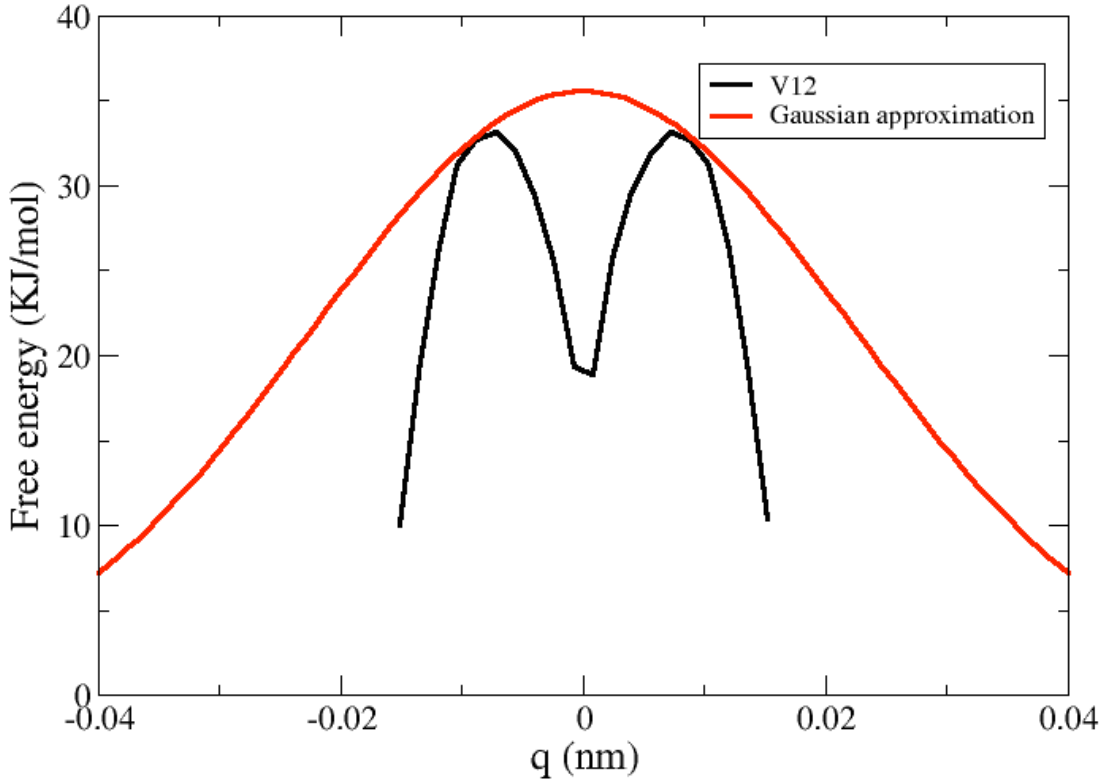

Figure 5-2: A plot of the off-diagonal coupling calculated by equation 4 and the Gaussian approximation used in the rMS-CG simulation. While the Gaussian approximation has much higher coupling than the calculated coupling where q is greater than 0.015 nm, the large free energy difference in the diagonal states make the coupling effectively zero even with the finite coupling provided by the Gaussian approximation.

The non-reactive CG free energy, the mirrored CG free energy, and the reactive AA free energy were used to calculate the off-diagonal coupling for the reactive CG model. The off-diagonal coupling was then fit to a Gaussian. While the actual coupling term does not closely resemble a Gaussian, care was taken to assure that the energy at the two humps and the energy at the Gaussian at these points were the same. Capturing the correct energy at the humps is important since the complex does not spend a significant

amount of time at the transition state at q=0, but does transition between each side of the peak. While there is a much larger coupling at q>0.2 nm than in the calculated coupling, the extremely large free energies difference of the diagonal states at this point allows for almost no coupling, even with a finite coupling term between the two states. By including the coupling, the resultant CG free energy very well reproduces the reference AA free energy. By using the coupling, the CG model is able to reproduce the AA free energy even though the non-reactive CG models that are used to define the diagonal states are not able to exactly reproduce the non-reactive AA free energies.

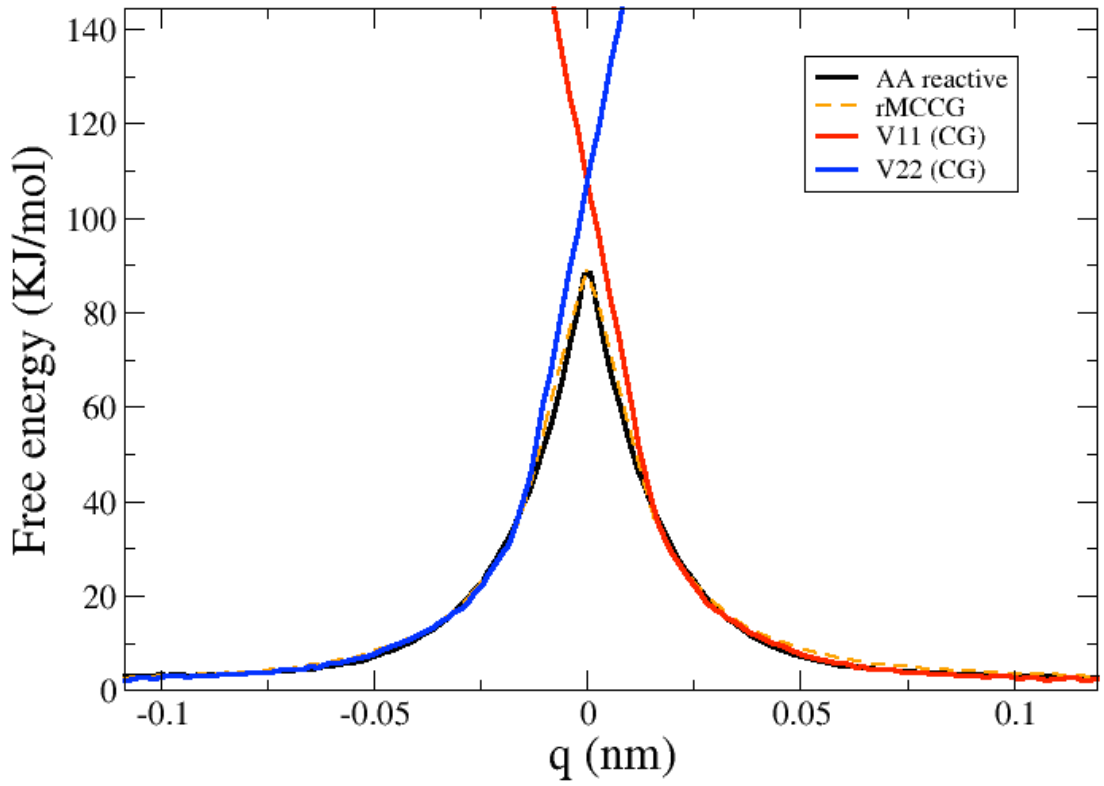

Figure 5-3: Comparison of the rMS-CG model to the AA reference data, along with the PMFs of the diagonal non-reactive states. The rMS-CG model is able to correctly model the barrier height of the AA model, even though the non-reactive MS-CG model does not perfectly agree with the AA model.

5.5 Conclusion

A new method is introduced in this work that allows for CG grained simulations to have dynamic bonding topologies in a way that reproduces the AA PMF. Systems that undergo reactions cannot be captured by a CG model without a dynamic bonding topology. Although there exists DPD methods that can switch between bonding topologies, this method is not guided by AA data and cannot have a mixture of multiple bonding topologies. Our new method is able to use AA-guided off-diagonal coupling, the $V_{12}^{AA/CG}$ term in equation 4 to allow the AA system to directly influence the CG system.

An important strength in our new method is the fact that the off-diagonal coupling term is a function of the CVs that describes the transition process. When a constant off-diagonal coupling term is used, the system is unable to capture the correct location of the transition barrier. Using a coupling term that is a function of the transition CVs not only allows the system to reproduce the location of the transition barrier, but also leads to more realistic descriptions of the transition process. Additionally, as the method is applied to other systems, more complicated CVs can be used to take into account other influences on structural transformations, such as environmental interactions. As long as a corresponding set of CG CVs can be found that clearly distinguish between the transitions, the major reduction in computational effort can be taken advantage of to investigate new emergent behavior.

Chapter 6

Comparative Study of CG model of DOPC Optimized by Relative Entropy Minimization and Mulitscale Coarse-Graining

6.1 Introduction

Cells are a well-known structural unit of living organism at the microscopic level. An important component of the cell is the cellular membrane, which gives the cell structures and regulates diffusion of ions and other molecules from the interior and exterior and vice versa.[136-138] Thus the study of cell membranes has been of great experimental and computational interest to understand their role in drug delivery and other biological processes.[4-5] As nature's "liquid crystal", the structure and function of theses membranes is often dictated by molecular fluctuations, which can translate to large-scale macroscopic properties.[139] These molecular fluctuations are difficult to probe experimentally, which has lead to the study of membranes with the use molecular simulation, namely molecular dynamics (MD). While the ever-increasing power of computer hardware has made it possible for computers to simulate ever larger systems and longer times with atomic resolution, modern computational resources are still unable to simulate the necessary length and time scales necessary to understand processes important to the behavior of cellular membranes such as angulations and phase-transitions.

This has lead to the development of coarse-grained (CG) models to simulate cellular membranes at times and lengths of biological interest. One important component of membranes that receives much attention is the phospholipid, as it typically makes up a majority of cellular membranes and is known to form bilayers. A phospholipid is an

amphiphilic molecule with a hydrophilic head of charged phosphorus moieties and hydrophobic tail comprised of long hydrocarbon chains. While the interplay between water and lipids play an important part in determining the ultimate structure of the bilayer, many CG models of lipids resort to solvent-free models, as the simulation of water, even at the CG level, still takes a majority of the computational cost of a simulation.

A wide variety of CG lipid models exist and can be broken down into two main types: top-down and bottom-up models. Top-down models aim to reproduce macroscopic behavior of a system and are then used to infer microscopic qualities. Top-down model can be further broken down into phenomenological models and experimental models. Phenomenological models aim to create a model that captures large-scale behavior of a lipid bilayer using a model that is as minimal as possible. These are usually built using only a few beads (3-5) and do not have any chemical specificity. Some phenomenological models include the Cooke/Deserno,[140] Brannigan/Brown,[141] and Schmid/Lenz[142] models. Experimental top-down models aim to create a model that reproduces certain macroscopic observables relating to a lipid bilayer. These models are typically much more detailed than phenomenological models, containing many more beads with additional chemical specificity. Some experimental models include Dry MARTINI[143] and Shinoda lipids.[144] Bottom-up models aim to reproduce microscopic behavior of a system and are then used to infer more macroscopic quantities. These models are typically parameterized from a more detailed computer simulation. As such, they are often difficult to parameterize since one must have high quality reference trajectories covering all relevant states of the system. Thus, only a few types of lipids have been parameterized

and less extensively used than top-down models. Bottom-up models may come in a variety of resolutions but the most prominent models are ones that contain 10-15 beads, such as Wang/Deserno,[145] and Sodt/Head-Gordon,[146] while only a few models exist at lower resolutions.[147-148]

While top-down models are much more commonly used to study membrane properties, they must be parameterized by trial and error. Furthermore, there is not a rigorous theory that connects top-down models to statistical mechanics, which puts their explanatory and predictive power into question. Bottom-up models, however, do have rigorous theories that connect them to statistical mechanics and methodologies exist for systematically creating such models from reference data.[12] Two bottom-up techniques that have received theoretical attention are relative entropy minimization[11] (REM) and Multiscale Coarse-graining[63] (MS-CG). Both of these methods attempt to create interactions between CG sites that will create trajectories that are consistent with a mapped atomistic trajectory, but attempt to do this in different ways. REM aims to create a model that captures correlations of the order of the basis set being used, while MS-CG aims to create a model that will reproduce the total force of a mapped model as closely as possible within a given basis set. In the limit of an infinite basis set and perfect sampling, it has been shown that these methods will produce the same model, as REM and MS-CG fit the average and the gradient of the information content model, respectively. Realistic CG models, however, are built with imperfect basis sets and limited sampling, so MS-CG and REM will give different solutions in practice. It has been hypothesized that REM will produce models that get averages of quantities more accurately while MS-CG will create models that better capture the curvature and fluctuations of models.[69] MS-CG also

attempts to fit a physical quantity, forces, instead of just simply matching distribution functions, which make it less likely to create a direct interaction between sites where one shouldn't exist. Earlier work on MS-CG has also suggested that it may better capture higher order correlations than other distribution based methods, such as REM, even with a limited basis set. The combination of these analyses have resulted in dictum which pronounces that methods like REM give complete fit of the reduced statistics while MS-CG gives a reduced fit to the complete statistics.

Despite the multitude of theoretical research comparing the two methods, little research has been done comparing models of complex systems using these optimization approaches. Studies that typically compare the two methods look at toy systems or simple liquids. While building CG models of these systems can be instructive in illuminating differences between methods, there are often few other observables to be compared besides ones being used for optimizing the model. Thus, to get a better understanding of the capabilities of the two methods, a comparison of a complex models, built using the same basis set, are compared by looking at secondary characteristic of the models. The present work aims to compare two CG models of DOPC, one built by REM and another built by MS-CG, and compare the secondary characteristics to better understand the two methods.

6.2 Models and Simulations

An all-atom (AA) simulation of DOPC used as reference data for the CG models was simulated with GROMACS.[149] An initial configuration of 1152 DOPC lipids in a box of 45,000 water molecules and 0.15 M concentration of potassium chloride was generated using the CHARMM-GUI membrane builder.[150] A simulation in the NPT

ensemble was done using a Nose-Hoover thermostat[75] set to 300 K with time constant of 1 ps and a semiisotropic Parrinello-Rahman barostat[151] set to 1 atm with atime constant of 5 ps. Hydrogen atoms were constrained using LINCS constraints[152] and electrostatics were calculated using particle-mesh Ewald summation. An integration time step of 2 fs was used. The CHARMM36 force field was used for lipid interactions and the TIP3P model was used for water.[19] Force-switching was done starting at 1.0 nm to a cut-off of 1.2 nm. After 300 ns equilibration under the conditions described above, statistics were captured under an NVT ensemble every picosecond for 100 ns, where the average volume over the last 100 ns was used and particles were rescaled from the last frame of the NPT run to match the average volume.

All lipids were mapped using a six-site mapping, detailed in figure 6-1, with water being mapped out of the system. This resolution is between the typically used highly coarse-grained MS-CG or phenomenological models and higher-resolution experimental top-down models. This resolution also allows for the distinguishing between the two tails of the lipid, which is important for describing some features. In both CG models, pairwise interaction potentials were used for intermolecular interaction and with a third order spline basis function. A basis set resolution of 0.01 nm and 0.03 nm was used for the MS-CG and REM models, respectively. Bonded and angled interactions were modeled using a third order spline basis function with a resolution of 0.005 nm for bonded terms and 0.5 degrees for angled terms.

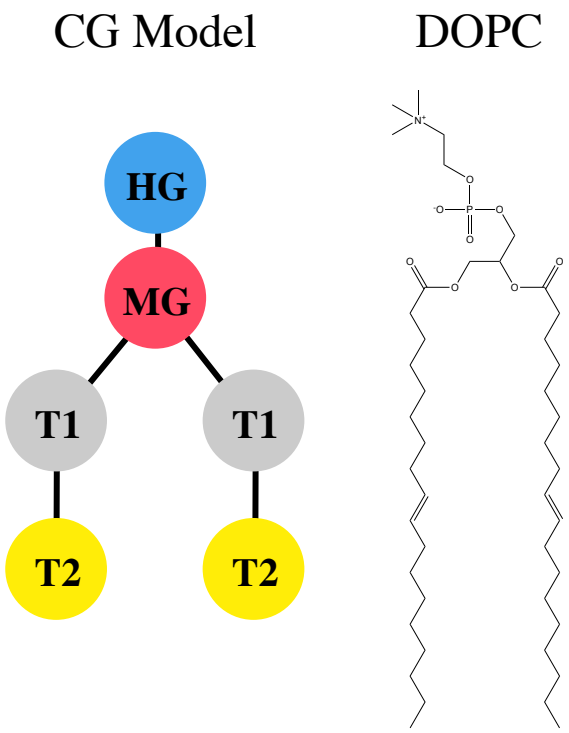

Figure 6-1: Description of mapping for DOPC model. A resolution for the coarse-grained model was chosen such that both tail could be resolved.

MS-CG forces were fit using the standard procedure described previously in the literature.[153] As described previously in the literature, REM optimization is done by calculating the derivative of the relative entropy with respect to the fitting parameters on an initial guess CG force field. These derivatives are calculated and then used to determine a new CG force field that is used to calculate new derivatives of the relative entropy.[34] This process is continued until convergence of the parameters is attained to some tolerance. The MS-CG interactions were used as the initial guess iteration for the REM optimization. Due to the low stability intermediate models determined by REM, a subset of 3 interactions, chosen randomly and changed every iteration, were fit instead of modifying all interactions. Once the subset of interactions was updated for 50 iterations,

all interactions were updated for up to 150 iterations until a relative tolerance of 0.05 was reached. This cycle was repeated one more time. All CG simulations were run in LAMMPS.[73] Intermediate CG simulations were run for 500,000 CG time steps, collecting statistics for every 500 time steps. Both the MS-CG and terminal REM models were run for 5,000,000 CG time steps and with statistics collected every 500 time steps.

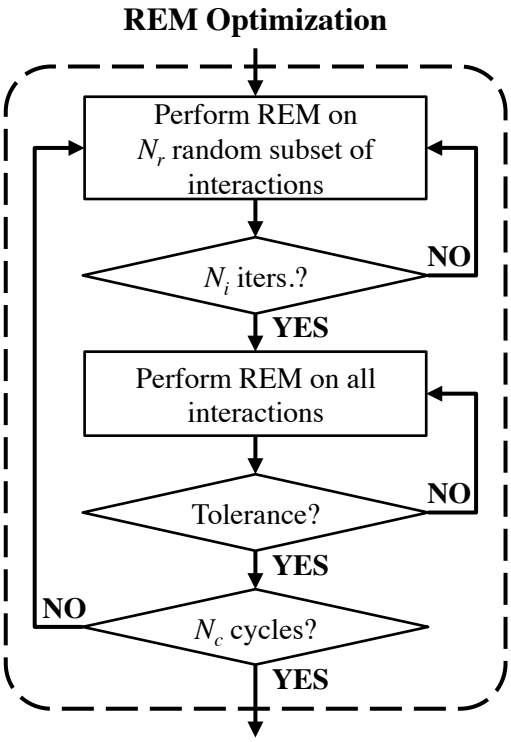

Figure 6-2: Schematic of REM optimization process used. A subset of the interactions was fit initially followed by a fitting of all the REM interactions. This was done so that intermediate REM models were more stable.

6.3 Results

It is informative to begin our analysis by assessing the similarities between the effective CG interactions predicted by MS-CG and REM. For simplicity, we will restrict our discussion to the 4 possible self-interactions, as seen in Figure 6-3, which represent a

subset of the 10 types of non-bonded interactions. One notable trend that is consistent between the two models is the relative partitioning of attractive and repulsive interactions. In general, the HG beads are repulsive while most of the lipid attraction is associated with MG, followed closely by weak T1 and T2 attractions. This trend is reminiscent of many well-known phenomenological implicit-solvent CG lipid models, which often mimic the hydrophobic effect with attractive tail and repulsive head interactions In fact, the model proposed by Brannigan, Philips, and Brown bears the greatest resemblance,[141] in which the interface CG beads (similar to MG) are designed to impart the strongest attraction, thereby maintaining the interfacial tension of the bilayer. Later, we will discuss the resultant behavior of the MS-CG and REM models, which will give credence to certain aspects intuited by previous phenomenological CG lipid models.

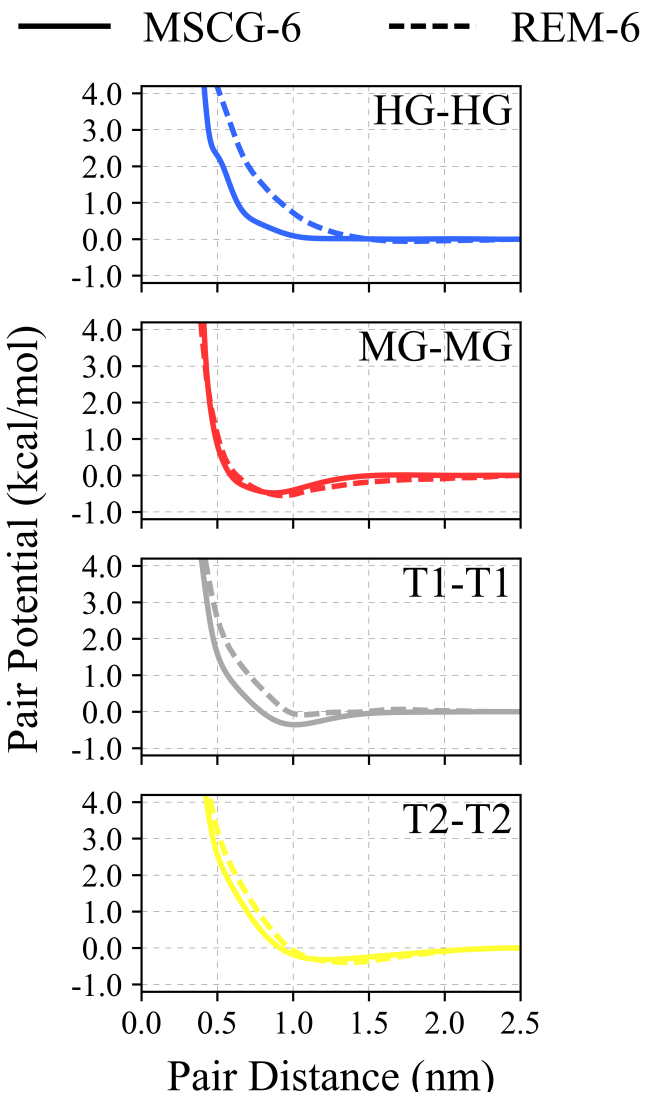

Figure 6-3: Comparison of potentials optimized via REM and MS-CG. REM predicts interactions that are more repulsive that MS-CG.

There are a few key differences between the MS-CG and REM models (shown in Figure 6-3) that are important to mention. First, the REM method predicts greater attraction by the MG beads compared to that of MS-CG, as evident by both the slightly deeper and longer-range nature of the MG-MG potential. In turn, and seemingly to compensate, the T1-T1 and T2-T2 interactions are noticeably weaker in the REM case, as

seen by their shallower interaction profiles. Finally, the REM model also predicts a softer but longer-range repulsion from the HG-HG interactions; repulsion begins around a separation distance of 1.5 nm in the REM case compared to around 1.0 nm in the MS-CG case.

We next assess the structural differences in equilibrated lipid bilayers using the two CG models. Lipid bilayers, by nature, are structurally anisotropic as lateral (i.e. in-plane) lipid packing should be distinguished from normal (i.e. out-of-plane) packing into two leaflets. As such, we compare the lateral and normal lipid number densities in Figure 6-4 and restrict ourselves to the 4 pair correlations chosen before; here, the results from MS-CG and REM models are compared against the number density profile from the mapped all-atom (AA) reference trajectories.

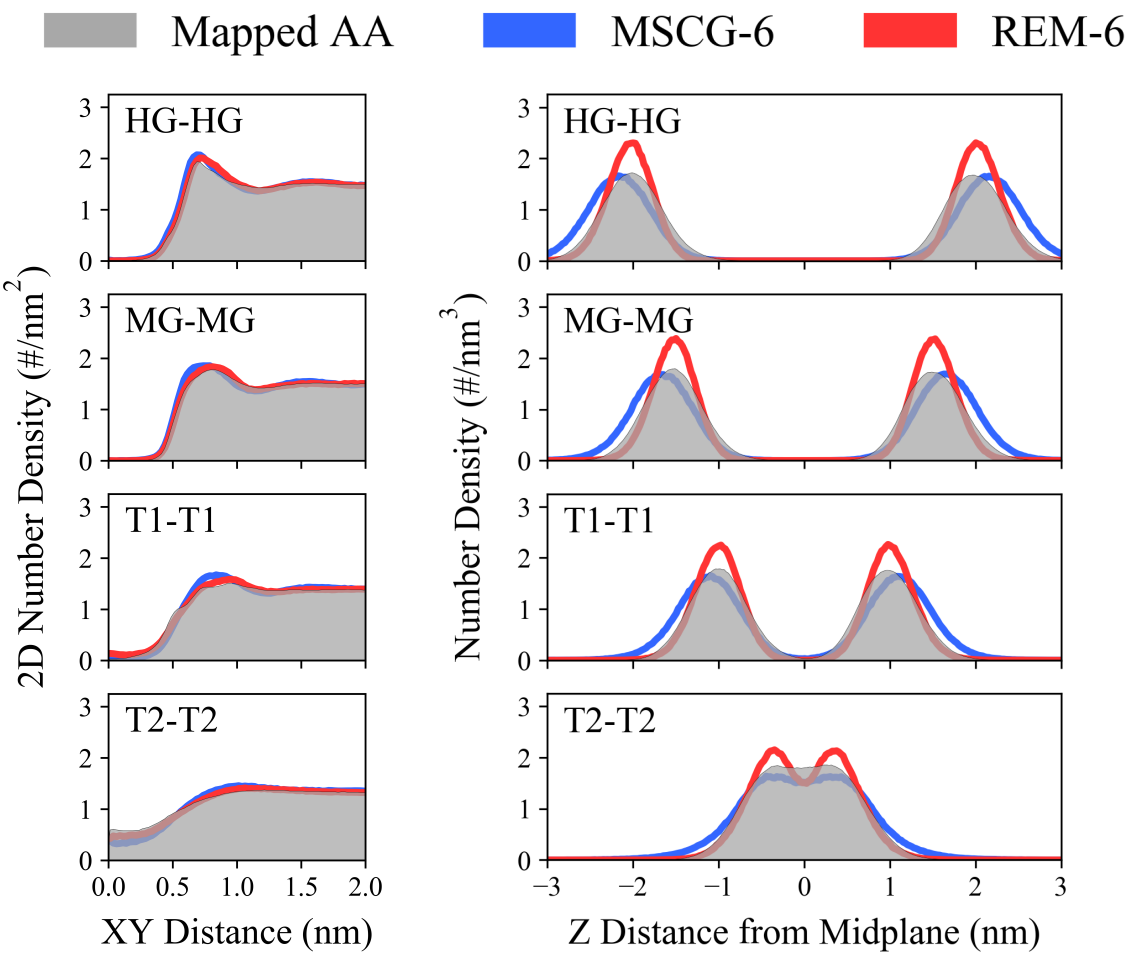

Figure 6-4: Graphs of the 2D number density in plane with the bilayer and 1D number density perpendicular to the bilayer. In the perpendicular number density, REM gets the correct average value of the distribution, while MS-CG matches the shape of the distribution more correctly.

The primary difference between the MS-CG and REM results can be seen by comparing how well each model recapitulates the reference data in the lateral and normal directions. In the lateral direction (left of Figure 6-4), both models show good agreement with the AA data. However, the first density peaks predicted by the MS-CG model, especially the MG and T1 profiles, tend to be slightly narrower than that of the REM

model, while also positioned at a shorter distance than that of the AA data. On the other hand, the MS-CG profiles in the normal direction (right of Figure 4), exhibit peaks that are positioned slightly wider than that of the AA data, while maintaining a similar spread. Here, the positions of the REM peaks are consistent with that of the AA data, but tend to be far narrower. We therefore find that both models exhibit over-structuring, which can be further seen when comparing the larger T1-T2 (S_{TT}) and HG-MG (S_{HM}) order parameters for the two CG models (Table 6-1) to that of the AA data. However, the over-structuring itself is anisotropic, such that the MS-CG model is more tightly bound in the lateral direction while the REM model is more tightly bound in the normal direction. Consequently, as seen in Table 6-1, the MS-CG and REM models have better agreement with the area per lipid and bilayer thickness, respectively, when compared to the AA data. It should also be noted the MS-CG model will form a bilayer from random configuration of lipids, while the REM model will not.

Property	MSCG-6	REM-6	AA	Exp.
APL (nm^2)	0.66	n/a	0.67	0.72[a]
Bilayer Thickness (nm)	4.4	4.1	4.0	4.48[a]
S_{TT} [S_{HM}]	0.65 [0.62]	0.63 [0.64]	0.56 [0.53]	n/a
Bending Modulus (k_bT)	65.4 ± 5.4	279.8 ± 17.1	28.8[b]	18.3[a]

Table 6-1: Summary of secondary properties of the DOPC models. REM is better able to reproduce properties relating to the averages of distribution functions, but MS-CG does a much better job of capturing the properties not directly related to the distribution function.

On a final note, we compare the membrane bending modulus using the two CG models, which we calculate from the low-frequency fluctuation spectra as described by Brandt et al. The bending modulus is an indication of the stiffness of the flexural modes of the bilayer, which are critical for membrane-mediated protein interactions. Our analysis in Table 6-1 suggests that both MS-CG and REM models are far stiffer (especially in the latter case) than that of previous experimental and theoretical observations. We expect that the origin of this stiffness is from the absence of the entropy-driven hydrophobic effect, i.e. due to the lack of solvent, which is a primary driver for lipid self-assembly. To compensate, both CG models seem to introduce explicit attraction, which is largely seen in the MG-MG interaction and, to a lesser extent, the T1-T1 and T2-T2 interactions (Figure 6-3). Thus, the effective CG interactions successfully stabilize the bilayer, but also introduce kinetic barriers that prevent perturbations, such as the aforementioned flexural modes. Therefore, we suggest that future work should consider including a means to represent solvent, thereby increasing the ability to express the elusive hydrophobic effect.

6.4 Discussion

In this section we will discuss how the two models of DOPC can be rationalized with current theoretical understandings of the differences between MS-CG and REM. Firstly, it should be noted that the RDF is not used as a basis of comparison of the two models. While REM is minimizing the relative entropy to determine the final REM model, this is equivalent to finding a CG potential that produces an RDF that will match the AA mapped RDF. MS-CG attempts to find a CG potential whose gradients reproduce

the forces on a mapped atomistic trajectory. Thus, using the RDF as a basis of comparison will always unfairly favor REM, as it is used as the fitting target within the methodology.

In place of the RDF, a comparison of the lateral and normal number density is performed between the two models. In the lateral comparison, REM seems to match the mapped AA number density closer than MS-CG. This is likely due to fact that the lateral number density is the major contributor to the RDF. Investigating the normal number density better illuminates the differences between the two models. As mentioned earlier, REM gets the averages correct with the wrong distribution shape while MS-CG get the correct distribution shape with the wrong averages. This can be rationalized by considering how each method fits the information content of the model. As described in Rudzinski and Noid, REM attempts to match the averages of the information content, while MS-CG attempts to fit gradients of the information content.[69] By examining the normal number density, it can be seen that REM creates a model that has distributions with the correct averages while the curvature of the distribution from the MS-CG model is closer to the mapped trajectory. Given the theoretical understanding of the two methods, it would seem that trend would be true for other properties related to distribution functions in models of systems. Other properties of the system can be rationalized by the averages versus curvature analysis of the two methods. The REM model better captures the atomistic bilayer thickness than the MS-CG model. Bilayer thickness can be thought of as an averaging of the head-head normal distributions, which it was already discussed that REM does a better job of capturing.

Properties that cannot be rationalized by the differences between how the two distributions are fit can be rationalized by how the two methods take into account many body interactions. In Noid et al.,[68] an analysis of the distribution matching and MS-CG within the YGB framework was done to explore how each captures the three-body correlations. As a distribution matching method, REM ignores higher-order correlations in order to capture the two-body correlation as well as possible if only a two-body basis set is used. This negligence of the three-body correlations in REM models may result in a model that overfits the two-body correlations at the expense of other properties of the system. MS-CG does take three-body correlations into account when the model is built. This fact can be used to understand why the MS-CG model better captures the bending modulus than REM model, as bending modulus depends on a membranes ability to induce curvature, which can only be explained with at least three points. Furthermore, proper aggregation behavior depends heavily on many-body interactions being properly captured, which is what is done by MS-CG model. Thus, while incorporating a higher order basis into the model would improve both the MS-CG and REM models, it seems that it would be required in the REM model to properly model these properties, which depend highly on three-body correlations. But MS-CG seems to capture them surprising well even with only a pair basis set.

As a possible future direction, it would seem that a method that incorporates both some aspects of both the REM and MS-CG methodologies might result in the best models being formed. This could be accomplished by only optimizing some basis function using REM from a MS-CG starting point. Since the curvature of the distribution seems to highly depend on the form of the repulsive part of pair interactions, it would

make sense to model the repulsive part of a pair interaction and model the attractive parts with REM. This is similar to what is done to improve coordination numbers by Experiment Directed simulation. Another future direction that will be necessary to proper modeling of lipid models is the incorporation of solvent effects in some way. An interaction that acts perpendicular to lipid membrane seems to be the natural choice for such an interaction, but calculating the normal vector for the bilayer on the fly is a computationally expensive task. Furthermore, it is unclear how this type of interaction should behave when a lipid is not within a bilayer.

Chapter 7

Conclusion and Future Directions

7.1 Introduction

The research presented in this thesis is aimed at presenting new CG methods for optimizing force-fields in order for CG models to fulfill their potential as tools for exploring longer length and time scales. In Chapter 3, the maximum entropy term that creates agreement between simulation and experiment is casted in the relative entropy formalism. In Chapter 4, the maximum entropy term was applied to coarse-grained variables and improved optimization techniques were explored. In Chapter 5, a matrix Hamiltonian based on EVB was implemented to allow for reactivity within CG simulations. In Chapter 6, we compare the MS-CG and REM optimization methods when applied to a DOPC lipid model. Based upon the work in this thesis, there are a number of future directions for research related to coarse-graining.

7.2 Future Direction

One of the most interesting applications of the results of the research presented in this thesis is the possibility of creating mixed top-down and bottom-up CG model using the EDS methodology. A direct connection between the bias form used in the EDS method and the relative entropy formalism was established in this thesis. It was shown that the EDS bias reduces the relative entropy between the model system and a hypothetical perfect model that perfectly agrees with experiment. Thus, the EDS method provides a systematic way of creating a top-down model of a given system. Furthermore, EDS does not restrict the basis set to Lennard-Jones functions, as is the case in the

MARTINI CG model and many other top-down models. This provides the possibility for a much more expressive model and one that possibly even include complex many-body interactions. These interactions could be mapped back down to a pairwise basis set with MS-CG, if desired. This would provide a way to close the cap between the two philosophies of building CG models.

One possible application of the CGDS method is to study the ATP hydrolysis. ATP hydrolysis has been studying using the harmonic basis functions to restrain a g-actin monomer into the conformation of f-actin, which was shown to decrease the barrier of the reaction. It has been shown in this thesis that the harmonic bias will create agreement between the second moment of the observable in question, even though it is typically used to create agreement between the first moment of the observable. This result seems to indicate that the prediction of the reaction barrier would be better than in previous studies.

Reactive methodologies seem to have a strong potential for application in electrolyte systems or other systems where the long time accumulations of products affect the properties of the system. While it would be very expensive to a high-resolution model of a system with many reactions happening, a CG model could be parameterized from one reaction at high resolution. The long time accumulation of the products could then modeled with a reactive CG model, which is important in the study of electrolyte breakdown in lithium ion batteries.

The comparison of the MS-CG and REM methodologies has indicated that the ideal approach for making bottom-up models may be to combine the two methodologies. A way this could be accomplished is by only optimizing some basis function using REM

from a MS-CG starting point. Since the curvature of the distribution seems to highly depend on the form of the repulsive part of pair interactions, it would make sense to model the repulsive part of a pair interaction and model the attractive parts with REM. This is similar to what is done to improve coordination numbers by EDS. This might be able to create models that get the correct distribution averages and curvature within the same model.

7.3 Remaining Challenges

Even though the methods presented give way of making CG models that combine top-down and bottom methods, one still needs to find a way of determining which observables are the most important to be fit. As indicated earlier, relative entropy based methods are known to overfit to data given to the model. While EDS uses a functional form that minimally biases the model, it is likely that observables will be correlated and that some will need to be weighted more heavily than others. Furthermore, the representability problem may make it difficult to fit observables not related to structural properties.

The methodology for determining the reactive coupling in rMS-CG seems to be more related to Boltzmann inversion, as it attempts to relate a PMF to a potential energy along a given collective variable. An iterative approach could be adopted to make it more like iterative Boltzmann inversion. Given the comparison between MS-CG and REM in chapter 6, it would seem that a very interesting area of research would be to find a MS-CG style fit of this reactive PMF. One could argue that this was done in the original MS-RMD paper, which does take advantage of force-matching residual, but this approach was iterative. It would be advantageous to find a method that could give a coupling

without the need for iteration. However, this would require a simulator to know the eigenvector of each state while the fit is being performed, which is not possible given the representability problem. If one were able to determine a way to find the coupling constant of the CG model during the fit, this could provide a promising methodology for producing reactive coarse-grained models.

7.4 Final Thoughts

It is my hope that the work in this thesis will inspire others to work on hybrid methods that incorporate different methodologies and philosophies. I feel that there is often an adversarial mindset method development where discoverers and proponents of one method are only interested in showing their method is better than others. I believe that I have shown that some methods are better at capturing some properties of a system while other methods are better at capturing other properties. I believe this sentiment to be true in other aspects of science as well, not just CG modeling.

Bibliography

[1] Newton, I., *Mathematical Principles of Natural Philosphy* Daniel Adee: New York.

[2] Pauli, W., "Uber das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik," Z. Phys. **36,** 5 (1926).

[3] Pauling, L., "The Nature of the Chemical Bond," J. Am. Chem. Soc.ß **53,** 4 (1931).

[4] Borhani, D. W.; Shaw, D. E., "The Future of Molecular Dynamics Simulations in Drug Discovery," J. Comput. Aided Mol. Des. **26,** 1 (2012).

[5] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A., "Role of Molecular Dynamics and Related Method in Drug Discovery," J. Med. Chem. **59,** (2016).

[6] Hyunh, L.; Neale, C.; R., P.; Allen, C., "Computational Approaches to the Rational Design of Nanoemulsions, Polymeric Micelles, and Dendrimers for Drug Delivery," Nanomedicine **8,** 1 (2012).

[7] Li, Y.; Tang, S.; Abberton, B. C.; Kroger, M.; Burkhart, C.; Jiang, B.; Papakonstantopoulos, G. J.; Poldneff, M.; Liu, W. K., "A Predictive Multiscale Computational Framework for Viscoelastic Properties of Linear Polymers," **53,** (2012).

[8] Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O., "The High-Throughput Highway to Computational Materials Design," Nat. Mater. **12,** (2013).

[9] Voth, G. A., *Coarse-Graining of Condensed Phase and Biomolecular Systems* CRC Press: Boca Raton, 2009.

[10] Muller-Plathe, F., "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back," ChemPhysChem **3,** 9 (2002).

[11] Shell, M. S., "The Relative Entropy is Fundamental to Multiscale and Iverse Thermodynamic Problems," J. Chem. Phys. **129,** (2008).

[12] Noid, W. G.; Chu, J.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C., "The Multiscale Coarse-Graining Method. I. A Rigorous Bridge Between Atomistic and Coarse-Grained models," J. Chem. Phys. **128,** (2008).

[13] Lyman, E.; Pfaendtner, J.; Voth, G. A., "Systematic Multiscale parameterization of Heterogenous elestic network models of proteins," Biophys. J. **95,** 9 (2088).

[14] Zhang, Z.; Lu, L.; Noid, W. G.; Krishna, V.; Pfaendtner, J.; Voth, G. A., "A systematic Methodology for Defining Coarse-grained Sites in Large Biomolecules " Biophys. J. **95,** 11 (2008).

[15] Rudzinski, J. F.; Noid, W. G., "Investigation of Coarse-Grained Mappings via an Iterative Generalized Yvon-Born-Green Mehtod," J. Phys. Chem. B **118,** 28 (2014).

[16] Cao, Z.; Voth, G. A., "The multiscale coarse-graining method. XI. Accurate interactions based on the centers of charge of coarse-grained sites," J. Chem. Phys. **143,** (2015).

[17] Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A., "On the Representability Problem and the Physical Meaning of Coarse-Grained Models," J. Chem. Phys. **145,** (2016).

[18] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," J. Comput. Chem. **4,** 2 (1983).

[19] Huang, J.; Mackerell, A., "CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR data," J. Comput. Chem. **34,** 25 (2013).

[20] Cornell, W. D.; Cleplk, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," J. Am. Chem. Soc. **117,** 19 (1995).

[21] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., "ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters for ff99SB," J. Chem. Theory Comput. **11,** 8 (2015).

[22] Barker, J. A.; Fisher, R. A.; Watts, R. O., "Liquid Argon: Monte Carlo and Molecular Dynamics Calculation," Mol. Phys. **21,** 4 (1971).

[23] Donnamaria, M. C.; Maranon, J.; Howard, E. I.; Fantoni, A.; Grigera, J. R., "The infleunce of charge Calculation on Molecular Dynamics Simulation of Adenine in Water " Mol. Simul. **18,** 1 (1996).

[24] Marrink, S. J. R.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H., "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulationsq," J. Phys. Chem. B **111,** (2007).

[25] Izvekov, S.; Voth, G. A., "Multiscale Coarse Graining of Liquid-State Systems," J. Chem. Phys. **123,** (2005).

[26] Chandler, D., *Introduction to Modern Statistical Thermodynamics* Oxford University Press: New York 1987.

[27] Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids* Oxford University Press: Oxford, 1989.

[28] Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications* 2nd ed.; Academica Press: San Diego, 2002.

[29] Noid, W. G., "Perspective: Coarse-Grained Models for Biomolecular Systems," J. Chem. Phys. **139,** (2013).

[30] Warshel, A.; Weiss, R. M., "An empirical valence bond approach for comparing reactions in solutions and in enzymes," **102,** 20 (1980).

[31] White, A. D.; Voth, G. A., "Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data," J. Chem. Theory Comput. **10,** 8 (2014).

[32] Jaynes, E. T., "Information Theory and Statistical Mechanics," Phys. Rev. **106,** 4 (1957).

[33] Chaimovich, A.; Shell, M. S., "Coarse-graining errors and numerical optimization using a relative entropy framework," J. Chem. Phys. **134,** 9 (2011).

[34] Carmichael, S. P.; Shell, M. S., "A New Multiscale Algorithm and Its Application to Coarse-Grainede Peptide Models for Self-Assembly," J. Phys. Chem. B **166,** (2012).

[35] Noid, W. G.; Liu, P.; Wang, Y.; Chu, J. W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A., "The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models," J. Chem. Phys. **128,** 24 (2008).

[36] Das, A.; Andersen, H. C., "The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation," J. Chem. Phys. **131,** (2009).

[37] Krishna, V.; Noid, W. G.; Voth, G. A., "The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures," J. Chem. Phys. **131,** (2009).

[38] Das, A.; Andersen, H. C., "The multiscale coarse-graining method. V. Isothermal-isobaric ensemble," J. Chem. Phys. **132,** (2010).

[39] Larini, L.; Lu, L.; Voth, G. A., "The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials," J. Chem. Phys. **132,** (2010).

[40] Lu, L.; Voth, G. A., "The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials," J. Chem. Phys. **134,** (2011).

[41] Das, A.; Andersen, H. C., "The multiscale coarse-graining method. VIII. Multiresolution hierarchical basis functions and basis function selection in the construction of coarse-grained force fields," J. Chem. Phys. **136,** (2012).

[42] Das, A.; Andersen, H. C., "The multiscale coarse-graining method. IX. A general method for construction of three body coarse-grained force fields," J. Chem. Phys. **136,** (2012).

[43] Das, A.; Lu, L.; Andersen, H. C.; Voth, G. A., "The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems," J. Chem. Phys. **136,** 19 (2012).

[44] Duncan, J. L.; Kelly, R. A.; Nivellini, G. D.; Tullini, F., "The Empirical General Harmonic Force Field of Ethane," J Mol Spectrosc **98,** 1 (1983).

[45] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E., "Systematic validation of protein force fields against experimental data," PLOS ONE **7,** 2 (2012).

[46] Wang, L. P.; Martinez, T. J.; Pande, V. S., "Building Force Fields: An Automatic, Systematic, and Reproducible Approach," J. Phys. Chem. Lett. **5,** 11 (2014).

[47] Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," J. Am. Chem. Soc. **118,** 45 (1996).

[48] Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A., "Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: a new method for force-matching," J. Chem. Phys. **120,** 23 (2004).

[49] Masia, M.; Probst, M.; Rey, R., "Ethylene carbonate-Li+: A theoretical study of structural and vibrational properties in gas and liquid phases," J. Phys. Chem. B **108,** 6 (2004).

[50] Jorn, R.; Kumar, R.; Abraham, D. P.; Voth, G. A., "Atomistic Modeling of the Electrode-Electrolyte Interface in Li-Ion Energy Storage Systems: Electrolyte Structuring," J. Phys. Chem. C **117,** 8 (2013).

[51] Best, R. B.; Vendruscolo, M., "Determination of protein structures consistent with NMR order parameters," J. Am. Chem. Soc. **126,** 26 (2004).

[52] Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K., "Combining Experiments and Simulations Using the Maximum Entropy Principle," PloS Comput. Biol. **10,** 2 (2014).

[53] Islam, S. M.; Stein, R. A.; McHaourab, H. S.; Roux, B., "Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme," J. Phys. Chem. B **117,** 17 (2013).

[54] Roux, B.; Islam, S. M., "Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy," J. Phys. Chem. B **117,** 17 (2013).

[55] Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M., "Simultaneous determination of protein structure and dynamics," Nature **433,** 7022 (2005).

[56] Rozycki, B.; Kim, Y. C.; Hummer, G., "SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions," Structure **19,** 1 (2011).

[57] De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M., "Characterization of the Interdomain Motions in Hen Lysozyme Using Residual Dipolar Couplings as Replica-Averaged Structural Restraints in Molecular Dynamics Simulations," Biochemistry **52,** (2013).

[58] Roux, B.; Weare, J., "On the statistical equivalance of restrained-ensemble simulations with the maximuym entropy method," J. Chem. Phys. **138,** (2013).

[59] Pitera, J. W.; Chodera, J. D., "On the Use of Experimental Observations to Bias Simulated Ensembles," J. Chem. Theory Comput. **8,** 10 (2012).

[60] Duchi, J.; Hazan, E.; Singer, Y., "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," J. Mach. Learn. Res. **12,** (2011).

[61] Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C. L.; Rodriguez-Ropero, F.; van der Vegt, N. F. A., "Systematic coarse-graining methods for soft matter simulations - a review," Soft Matter **9,** 7 (2013).

[62] Saunders, M. G.; Voth, G. A., "Coarse-graining methods for computational biology," Annu. Rev. Biophys. **42,** (2013).

[63] Izvekov, S.; Voth, G. A., "A multiscale coarse-graining method for biomolecular systems," J. Phys. Chem. B **109,** 7 (2005).

[64] Murtola, T.; Falck, E.; Karttunen, M.; Vattulainen, I., "Coarse-Grained Model for Phospholipid/Cholesterol Bilayer Employing Inverse Monte Carlo with Thermodynamic Contraints," J. Chem. Phys. **126,** (2007).

[65] Lyubartsev, A. P.; Laaksonen, A., "Calculation of Effective Interaction Potentials form Radial Distribution Functions," Phys. Rev. E **52,** (1995).

[66] Mullinax, J. W.; Noid, W. G., "A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials," J. Phys. Chem. C **114,** 12 (2010).

[67] Bilionis, I.; Zabaras, N., "A stochastic optimization approach to coarse-graining using a relative-entropy framework," J. Chem. Phys. **138,** 4 (2013).

[68] Noid, W. G.; Chu, J.; Ayton, G. S.; Voth, G. A., "Multiscale Coarse-Graining and Structural Correlations: Connections to Liquid-State Theory," J. Chem. Phys. B. **111,** (2007).

[69] Rudzinski, J. F.; Noid, W. G., "Coarse-graining entropy, forces, and structures," J. Chem. Phys. **135,** 21 (2011).

[70] Cover, T. M.; Thomas, J. A., Entropy, Relative Entropy, and Mutual Information. In *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, New Jersey, 2006; pp 13-56.

[71] Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., "The MARTINI Coarse-Grained Force Field: Extension to Proteins," J. Chem. Theory Comp. **4,** (2008).

[72] de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J., "Improved Parameters for the Martini Coarse-Grained Protein Force Field," J. Chem. Theory Comp. **9,** (2012).

[73] Plimpton, S., "Fast Parallel Algorithms for Short-Range Molecular-Dynamics," J. Comput. Phys. **117,** 1 (1995).

[74] Eastwood, J. W., "Optimal Particle-Mesh Algorithms," J. Comput. Phys. **18,** 1 (1975).

[75] Hoover, W. G., "Canonical Dynamics - Equilibrium Phase-Space Distributions," Phys. Rev. A **31,** 3 (1985).

[76] Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M., "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," Comput Phys Commun **180,** 10 (2009).

[77] Iannuzzi, M.; Laio, A.; Parrinello, M., "Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics," Phys. Rev. Lett. **90,** 23 (2003).

[78] Yuet, P. K.; Blankschtein, D., "Molecular Dynamics Simulations Study of Water Surfaces: Comparison of Flexible Water Models," J. Phys. Chem. B **114,** (2010).

[79] White, A. D.; Dama, J. F.; Voth, G. A., "Designing Free Energy Surfaces That Match Experimental Data with Metadynamics," J. Chem. Theory Comput. **11,** 6 (2015).

[80] K, X., "Nonaqueous Liquid Electrolytes for Lithium-Based Rechargeable Batteries," Chem. Rev. **104,** (2004).

[81] Winter, M., "The Solid Electrolyte Interpahse-The Most Important and Least Understood Solid Electrolyte in Rechargable Li Ion Batteries," Z. Phys. Chem. **223,** 10-11 (2009).

[82] Verma, P.; Maire, P.; Novak, P., "A review of the features and analyses of the solid electrolyte interphase in Li-ion batteries," Electrochim. Acta **55,** 22 (2010).

[83] Wang, Y. X.; Nakamura, S.; Ue, M.; Balbuena, P. B., "Theoritical Studies to Understand Surface Chemistry on Carbon Anodes for Lithium Ion Batteries: Reaction Mechanisms of Ethylene Carbonate," J. Am. Chem. Soc. **123,** 47 (2001).

[84] Tasaki, K., "Solvent decompositions and physical properties of decomposition compounds in Li-ion battery electrolytes studied by DFT calculations and molecular dynamics simulations," J. Phys. Chem. B **109,** 7 (2005).

[85] Smith, J. C.; Roux, B., "Eppur si Muove! The 2013 Nobel Prize in Chemistry," Structure **21,** 12 (2013).

[86] McCullagh, M.; Saunders, M. G., "Unraveling the Mystery of ATP Hydrolysis in Actin Filaments," J. Am. Chem. Soc **136,** 37 (2014).

[87] Sun, R.; Sode, O.; Dama, J. F.; Voth, G. A., "Simulating Protein Medicated Hydrolysis of ATP and Other Nucleoside Triphosphates by Combining QM/MM Molecular Ddynamics with Advances in Metadynamics," J. Chem. Theory Comput. **13,** 5 (2017).

[88] Bahr, L.; Rader, A. J., "Coarse-graind Normal Mode Analysis in Structural Biology," Curr. Opin. Struct. Biol **15,** 5 (2005).

[89] Pitera, J.; Chodera, J. D., "On the Use of Experimental Observations to Bias Simulated Ensembles," J. Chem. Theory Comput. **8,** 10 (2012).

[90] Roux, B.; Weare, J., "On the Statistical Equivalence of Restrained-Ensemble Simulations witht he Maximum Entropy Method," J. Chem. Phys. **138,** 8 (2013).

[91] Cearsi, A.; Gil-Ley, A.; Bussi, G., "Combining Simulations and Solution Experiments as a Paradigm for RNA Force Fields Refinement," J. Chem. Theory Comput. **12,** 12 (2016).

[92] White, A. D.; Kinght, C.; Hocky, G. M.; Voth, G. A., "Communication: Improved Ab Initio Molecular Dynamics by Minimally Biasing with Experimental Data," J. Chem. Phys. **146,** 4 (2017).

[93] Pollard, T. D.; Cooper, J. A., "Actin, a Central Player in Cell Shape and Movement," Science **326,** 5957 (2009).

[94] Dominguez, R.; Holmes, K. C., "Actin Structure and Function," Annu. Rev. Biophys. **40,** (2011).

[95] Oda, T.; Iwasa, M.; Aihara, T.; Maeda, Y.; Narita, A., "The Nature of the Globular- to Fibrous-Actin Transition," Nature **457,** 7228 (2009).

[96] Blanchoin, L.; Pollard, T. D., "Hydrolysis of ATP by Polymerized Actin Depends on the Bound Divalent Cation but Not Profilin," Biochemistry **41,** 2 (2002).

[97] Melki, R.; Fievez, S.; Carlier, M. F., "Continous Monitoring of Pi Release Following Nucleotide Hydrolysis in Actin and Tubulin Assembly Using 2-Amino-6-mercapto-7-methylpurine Ribonucleoside and Purine-nucleoside Phosphorylase as an Enzyme-lined Assay," Biochemistry **65,** 37 (1996).

[98] Chu, J. W.; Voth, G. A., "Allostery of Actin Filaments: Molecular Dynamics Simulations and Coarse-grained Analysis," Proc. Natl. Acad. Sci. U. S. A. **65,** 37 (2005).

[99] Pfaendtner, J.; Lyman, E.; Pollard, T. D.; Voth, G. A., "Structure and Dynamics of the Actin Filament," J. Mol. Biol. **396,** 2 (2010).

[100] Pfaendtner, J.; Branduardi, D.; Parrinello, M.; Pollard, T. D.; Voth, G. A., "Nucleotide-dependent Confromational States of Actin," Proc. Natl. Acad. Sci. U. S. A. **106,** 31 (2009).

[101] Fan, J.; Saunders, M. G.; Voth, G. A., "Coarse-graining provides Insights on the Essential Nature of Hetreogenity in Actin Filaments," Biophys. J. **106,** 31 (2012).

[102] Saunders, M. G.; Tempkin, J.; Weare, J.; Dinner, A. R.; Roux, B.; Voth, G. A., "Nucleotide Regulation of the Structure and Dynamics of G-actin," Biophys. J. **106,** 8 (2014).

[103] Saunders, M. G.; Voth, G. A., "Comparison Between Actin Filament Models: Coarse-graining Reveals Essential Differences," Structure **20,** 4 (2012).

[104] Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camiloni, C.; Bussi, G., "PLUMED 2: New Feathers for an Old Bird," Comput. Phys. Commun. **185,** 2 (2014).

[105] Hocky, G. M.; Baker, J.; Bradley, M.; Sinitskiy, A. V.; De la Cruz, E. M.; Voth, G. A., "Cations Stiffen Actin Filaments by Adhering a Key Structural Element to Adjacent Subunits," J. Phys. Chem. B. **120,** 2 (2016).

[106] Graceffa, P.; Dominguez, R., "Crystal Structure of Monomeric Actin in the ATP state. Strucural Element to Adjacent Subunits," J.Biol. Chem. **278,** 36 (2003).

[107] Stinis, P., "A Maximum Likelihood Algorithm for the Estimation and Renormalization of Exponential Densities," J. Comput. Phys. **202,** 2 (2005).

[108] Ferrarotti, M. J.; Bottaro, S.; Perez-Villa, A.; Bussi, G., "Accurate Multiple Time Step in Biased Molecular Simulations," J. Chem. Thoery Comput. **11,** 1 (2015).

[109] Knight, C.; Voth, G. A., "The Curious Case of the Hydrated Proton," **45,** 1 (2012).

[110] Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A., "An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport," **112,** 2 (2008).

[111] Schmitt, U. W.; Voth, G. A., "Multistate Empirical Valence Bond Model for Proton Transport in Water," **102,** 29 (1998).

[112] Voth, G. A., "Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems," **39,** 2 (2006).

[113] Savage, J.; Voth, G. A., "Proton Solvation and Transport in Realistic Proton Exchange Membrane Morphologies," **120,** 6 (2016).

[114] Chen, H.; Wu, Y.; Voth, G. A., "Proton Transport Behavior through the Influenza A M2 Channel: Insights from Molecular Simulation," **93,** 10 (2007).

[115] Cao, Z.; Kumar, R.; Peng, Y.; Voth, G. A., "Hydrated Proton Structure and Diffusion at Platinum Surfaces," **119,** 26 (2015).

[116] Knight, C.; Lindberg, G. E.; Voth, G. A., "Multiscale reactive molecular dynamics," **137,** 22 (2012).

[117] Saunders, M. G.; Voth, G. A., "Coarse-graining of multiprotein assemblies," **22,** 2 (2012).

[118] Knight, C.; Voth, G. A., "Coarse-graining away electronic structure: a rigorous route to accurate condensed phase interaction potentials," **110,** 9-10 (2012).

[119] Dama, J. F.; Sinitskiy, A. V.; McCullagh, M.; Weare, J.; Roux, B.; Dinner, A. R.; Voth, G. A., "The Theory of Ultra-Coarse-Graining. 1. General Principles," **9,** 5 (2013).

[120] Marrink, S. J.; Tieleman, D. P., "Perspective on the Martini model," **42,** 16 (2013).

[121] Sergei, I.; Gregory, A. V., "Multiscale coarse graining of liquid-state systems," **123,** 13 (2005).

[122] Izvekov, S.; Voth, G. A., "Solvent-Free Lipid Bilayer Model Using Multiscale Coarse-Graining," **113,** 13 (2009).

[123] Lyubartsev, A. P.; Laaksonen, A., "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," **52,** 4 (1995).

[124] Shell, M. S., "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," **129,** 14 (2008).

[125] Liu, M. B.; Liu, G. R.; Zhou, L. W.; Chang, J. Z., "Dissipative Particle Dynamics (DPD): An Overview and Recent Developments," **22,** 4 (2015).

[126] Sato, M.; Matsuoka, S.; Sloot, P. M.; Albada, G. D. v.; Dongarra, J.; Liu, H.; Li, H.; Lu, Z.-Y., "Proceedings of the International Conference on Computational Science, ICCS 2011Incorporating Chemical Reactions in Dissipative Particle Dynamics Simulations," **4,** (2011).

[127] Lísal, M.; Brennan, J. K.; Smith, W. R., "Mesoscale simulation of polymer reaction equilibrium: Combining dissipative particle dynamics with reaction ensemble Monte Carlo. I. Polydispersed polymer systems," **125,** 16 (2006).

[128] Lísal, M.; Brennan, J. K.; Smith, W. R., "Mesoscale simulation of polymer reaction equilibrium: Combining dissipative particle dynamics with reaction ensemble Monte Carlo. II. Supramolecular diblock copolymers," **130,** 10 (2009).

[129] Farah, K.; Karimi-Varzaneh, H. A.; Müller-Plathe, F.; Böhm, M. C., "Reactive Molecular Dynamics with Material-Specific Coarse-Grained Potentials: Growth of Polystyrene Chains from Styrene Monomers," **114,** 43 (2010).

[130] Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M., "PACKMOL: A package for building initial configurations for molecular dynamics simulations," **30,** 13 (2009).

[131] Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A., "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," **181,** 9 (2010).

[132] Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M., "PLUMED: A portable plugin for free-energy calculations with molecular dynamics," Comput Phys Commun **180,** 10 (2009).

[133] Lu, L.; Voth, G. A., The Multiscale Coarse-Graining Method. In *Adv. Chem. Phys.*, Rice, S. A.; Dinner, A., Eds. Wiley-Interscience: New York, 2012; Vol. 149.

[134] Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A., "Computationally Efficient Multiconfigurational Reactive Molecular Dynamics," **8,** 12 (2012).

[135] Grossfield, A., "WHAM: the weighted histogram analysis method,"

[136] *Structure and Dynamics of Membranes I. From Cells to Vesicles* 1st ed.; Elsevier: North Holland, 1995.

[137] Alberts, B.; Johnson, A.; Lewis, L.; Raff, M.; Roberts, K.; Walter, P., *Molecular Biology of the Cell* 4th ed.; Garland Science: New York, 2002.

[138] Karp, G., *Cell and Molecular Biology: Concepts and Experiments* 5th ed.; John Wiley & Sons: Hoboken, NJ, 2007.

[139] Philips, M. C.; Williams, R. M.; Chapman, D., "On the Nature of Hydrocarbon Chain Motions in Lipid Liquid Crystals," Chem. Phys. Lipids **3,** 3 (1969).

[140] Cooke, I. R.; Kremer, K.; Deserno, M., "Tunable Generic Model for Fluid Bilayer Membranes," Phys. Rev. E **72,** (2005).

[141] Brannigan, G.; Philips, P. F.; Brown, F. L. H., "Flexible Lipid Bilayers in Implicit Solvent," Phys. Rev. E **72,** (2005).

[142] Lenz, O.; Schmid, F., "A Simple Computer Model for Liquid Lipid Bilayers," J. Mol. Liq. **117,** 1 (2005).

[143] Arnarex, C.; Uusitalo, J. J.; Masman, M. F.; Ingolfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periole, X.; de Vries, A. H.; Marrink, S. J., "Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent," J. Chem. Theory Comput. **11,** 1 (2015).

[144] Shinoda, W.; DeVane, R.; Klein, M. L., "Zwitterionic Lipid Assemblies: Molecular Dynamics Studies of Monolayers, Bilayers, and Vesicles Using a New Coarse Grain Force Field," J. Phys. Chem. B **114,** 20 (2010).

[145] Wang, Z.; Deserno, M., "Systematic Implicit Solvent Corase-Graining of Bilayer Membranes: Lipid and Phase Transferability of the Force Field," New J. Phys. **12,** 095004 (2010).

[146] Sodt, A. J.; Head-Gordon, T., "An Implicit Solvent Coarse-Grained Lipid Model with Correct Stress Profile," J. Chem. Phys. **132,** (2010).

[147] Srivastava, A.; Voth, G. A., "Hybrid Approach for Highly Coarse-Grained Lipid Bilayer Models," J. Chem. Theory Comput. **9,** 1 (2013).

[148] Srivastava, A.; Voth, G. A., "Solvent-Free, Highly Coarse-Grained Models for Charged lipid Systems," J. Chem. Theory Comput. **10,** 10 (2014).

[149] Abraham, M. J.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J. C.; Hess, B.; Lindahl, E., "High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers," SoftwareX **95,** 9 (2015).

[150] Jo, S.; Kim, T.; Iyer, V. G.; Im, W., "CHARMM-GUI: a Web-based Graphical User Iterface for CHARMM," J. Comput. Chem. **29,** 11 (2008).

[151] Parrinello, M.; Rahman, A., "Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method," J. Appl. Phys. **52,** (1981).

[152] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., "LINCS: A Linear Constraint Solver for Molecular Simulations," J. Comput. Chem. **18,** (1998).

[153] Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A., "Efficient, Regularized, and Scalable Algorithms for Multiscale Coarse-Graining," J. Chem. Theory Comput. **6,** 3 (2010).