THE UNIVERSITY OF CHICAGO

ADVANCED COARSE-GRAINED MODELS OF LIPID AND LIQUID SYSTEMS ANALYZED FROM THE PERSPECTIVE OF ACCURACY AND PRACTICALITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

TIMOTHY DAVID LOOSE

CHICAGO, ILLINOIS

JUNE 2023

Dedicated to my wife Maliha, who believed in me even when I didn't.

“Particle man, particle man
Doing the things a particle can
What's he like? It's not important
Particle man

Is he a dot, or is he a speck?
When he's underwater does he get wet?
Or does the water get him instead?
Nobody knows, Particle man”

-They Might Be Giants

Table of Contents

List of Figures

List of Tables

Acknowledgements

They say raising a child takes a village, I now understand that this applies to graduating a Ph.D. student as well. There are so many people at the University of Chicago and beyond who made it possible for me to get through my time in graduate school with my sanity mostly intact.

First are my friends and family: Mom and Dad, who always believed in me and encouraged my curiosity from a young age. I attribute my nerdiness and wonder at the natural world to them both. My sister Lizz made sure I wasn't too nerdy which is equally important. She has also been a role model to me, showing me that I can choose my own path in this world. Mary Beth showed me what true strength and wisdom are, though it took a while for me to internalize those lessons. I always try to think about what you would tell me whenever I have a problem. I also thank all my friends at the university, especially Adam, Brad, Christian, James, and Josh. I'll never forget our time in the Kent basement grading lab reports and fumbling our way through stat mech problem sets. My friends from the University of Texas: Arthur, James, Nick, Nikita, and Mia all helped me get through the pandemic. I can't wait to run more games of Pathfinder for them all soon. Lastly, I owe all of this to my wife Maliha. She has given me so much strength, solace, and joy over these years without which this would not be possible. I also thank my in-laws, Nadeem, Nuzhat, Amna and Haroon, who have shown me so much support through their prayers and encouragement.

I have so many mentors and co-workers to acknowledge as well. First, my advisor Gregory A. Voth, who gave me a lot of latitude in how I approached my projects. I believe that this has made me a better, more confident scientist. I'm grateful as well for all the mentorship and grant writing I was able to do in the lab as well as it made me a much more well-rounded person and gave me plenty to do when the science wasn't working. Thomas Dannenhoffer-Lafage showed me

the ropes in my first summer. Aleksander Durumeric pushed me to explore machine learning, which proved to be a fruitful area of research and gave me two chapters for this dissertation. Alex Pak helped me through the toughest projects I had was a constant sounding board for new ideas. Patrick Sahrmann, my frequent collaborator, inspired me to brush up on my math skills. He has also taken on the mantle of improving CG lipid models and incorporating machine learning into the group's methods. I'm certain he will go far.

Finally, I thank the National Science Foundation for funding through the Graduate Research Fellowship Program, as well as the National Institute of Health. I'm grateful to have been so well funded throughout the process without needing to teach. I also acknowledge the RCC and XSEDE/ACCESS for giving me access to high quality computational resources which allowed me to carry out all the research contained in this dissertation. It's been very fun to play around with some of the largest supercomputers in the world.

Abstract

In the world of molecular simulation, there is a large gap between the capabilities of all-atom molecular dynamics and many biophysical systems that are of interest. Coarse-grained (CG) molecular dynamics addresses this gap by increasing the timescales and system sizes which are accessible to molecular simulation. It does so by reducing the complexity of the all-atom system to a reduced number of degrees of freedom known as CG beads or sites. These sites represent a collection of individual atoms and are often chosen via chemical intuition. Parameterizing the interactions between these sites is done in one of two ways. The first: top-down CG, involves selecting experimental observables and hand tuning the force field to match these targets. This process is laborious and is impossible to do systematically without extensive knowledge of force field development. On the other hand, bottom-up CG addresses these issues by algorithmically parameterizing force fields to minimize loss functions with respect to reference all-atom data which has been mapped to the CG resolution. In this way, bottom-up CG models can in theory be generated for arbitrary systems if one has access to short reference trajectories and an appropriate CG algorithm. In practice though, bottom-up CG becomes more and more difficult as one attempts to apply it to more complicated systems- particularly when many-body correlations, anisotropy, or large numbers of model parameters are in play. Unfortunately, these problematic aspects are all but a certainty when generating CG models of relevant biophysical processes, such as membrane remodeling or protein assembly. Thus, it is imperative that the limitations of bottom-up CG are understood so that better methods can be developed to address them.

This work is broken up into two main sections. Chapters 2 and 3 analyze bottom-up and top-down CG lipid bilayers at a resolution of 4 heavy atoms to one CG bead. In both cases CG models tend to fail to reproduce thermodynamic properties of the bilayer without explicit

temperature dependence. In addition, the bottom-up CG lipid models require extensive and complicated optimization schemes which limit their practicality. These issues stem from solvent effects and the inherent anisotropy of lipid membranes, and are tricky to address without advanced methodologies, such as the implementation of semi-explicit solvent virtual particles. Ultimately, these models fail to meet the accuracy and expressiveness of CG lipids at lower resolutions which suggests that the higher CG resolution is inappropriate for such systems. The second section of this work pertains to applying machine learning (ML) methods to CG modeling. In chapters 4 and 5 it is demonstrated that for liquid systems with significant many-body and nuclear quantum effects deep neural networks generate much more accurate models. Normally, neural networks bring this increase in accuracy at the cost of integration speed and significant data requirements. In fact, these models in many cases run slower than the corresponding all-atom simulations. However, if a path integral representation of the system is required, these ML based force fields are not only faster, but are essentially just as accurate. In the case of classical MD to CG resolution, this is not the case, but equivariant neural networks can be applied which significantly reduce the amount of training data needed to produce an accurate model. In fact, a single frame of MD data is sufficient to generate a stable CG model of water with an equivariant neural network, which is two orders of magnitude lower than the requirements of non-equivariant networks. Overall, these projects demonstrate that bottom-up CG modeling remains difficult for complex systems, but recent advancements in machine learning and traditional CG methods provide a path towards more accurate and practical CG models.

Chapter 1: Introduction

The study of statistical mechanics is fundamental to the understanding of biology at the molecular level. Cells are composed of huge numbers of molecules, the aggregate behavior of which gives rise to the emergent complex behavior we know as life.[1] As an example, all cells are bound by a lipid membrane which is composed many billions of atoms within hundreds of separate molecular species, including but not limited to proteins and lipids, which can form a wide range of morphologies and localized structures in order to regulate cellular signaling and motion.[2–4] Chemical knowledge of these molecules and how each individual species interacts with one another can be obtained via application of quantum mechanical principles, and macroscopic behavior of the bilayer can be probed via experiment. Statistical mechanics can bridge the gap between these two regimes and provide a theory to predict how microscopic interactions give rise to macroscopic behavior, thus connecting physics with biology.

Molecular dynamics (MD) provides a valuable tool to probe details of molecular systems that are inaccessible, either due to temporal or spatial resolution constraints, to study via experimental techniques.[5] By simulating the motions of individual atoms using pairwise force fields, the static and dynamic properties of the system can be interrogated at a computational cost far below that of *ab-initio* methods while retaining a great deal of accuracy. Furthermore, MD is readily extendable in a variety of ways. Thermostats and barostats can enable the simulation of systems outside the microcanonical ensemble.[6,7] Quantum mechanical detail can be incorporated into the force field to provide a middle ground between electronic structure and traditional MD with respect to efficiency and accuracy.[8,9] Atomic representations can be extended to account for polarizability and other state changes. This is all to say that there are few classes of molecular

systems that cannot be studied with MD. However, MD has found itself to be particularly useful for the study of biological processes. Due to the large number of atoms that modern hardware and MD software packages can handle, and the relative lack of quantum effects present in such processes, MD has shown tremendous usefulness to support experimental research, as well as to make predictions of its own. These simulations can provide insight into the fundamental statistical mechanics that drive life at a molecular scale, and allow for the calculation simple structural properties such as radial distribution functions all the way up to macroscopic thermodynamic properties such as entropies, and free energy landscapes, as well as dynamic properties such as diffusion and rate constants which cannot be studied via methods such as Monte Carlo sampling.[10] The direct connection of data gathered from molecular dynamics trajectories to statistical mechanics provides rich ground for the development of biophysical theories which frequently inform our understanding of disease and drug discovery.

Despite this power, there are still many biophysical processes that remain out of reach to all-atom (AA) MD even with the advantage of advanced hardware such as GPUs, which can greatly accelerate MD integration. Currently, we are limited to simulations on the temporal scale of microseconds. In contrast, the nature of protein folding, a key process which underpins essentially all cellular functions, is easily 3 orders of magnitude longer.[11] On the spatial side, MD is practically limited to simulations of 100 million to 1 billion atoms[12], while many physiologically relevant systems, such as large membrane patches, can easily eclipse this scale when accounting for the associated proteins and solvent molecules. Of course, there are specialized hardware designed with all-atom MD in mind such as Anton which can push beyond these boundaries, but they are by far the exception rather than the rule.[13] So while atomistic MD simulations have tremendous potential, they only covers a fraction of the cases which scientists desire to study.

There are two main approaches to solving this issue. The first, enhanced sampling (ES), involves accelerating the rate at which interesting molecular configurations are sampled via biases added to the Hamiltonian.[14] The second method, which is the main topic of the present work, is coarse-graining (CG). CG simulation increases the size and timescales accessible to molecular dynamics through a reduction of the number of particles simulated accomplished by mapping groups of interactions into CG sites as shown in Figure 1.1.[15–19] These will be referred to interchangeably as sites or beads throughout this document. The reduction of the atomic system to a lower resolution benefits MD sampling in three ways. First, the resulting simulation has fewer beads than the corresponding atomistic system has atoms, which naturally speeds up integration. Second, entire

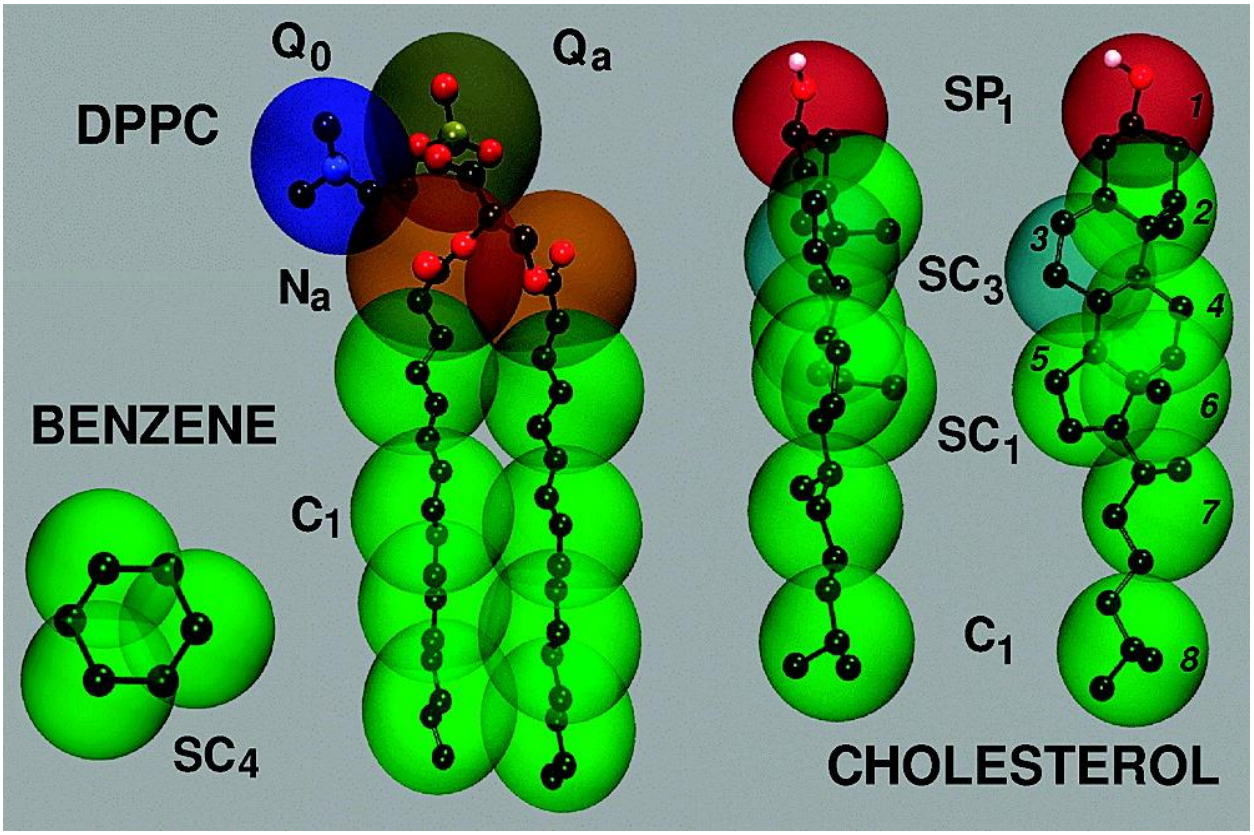

Figure 1.1: CG mappings for several lipids and benzene. Heavy atoms are represented as small spheres and bonds are represented with sticks. Each large sphere represents a single CG site. In the case of this mapping atoms can correspond to a single bead, or multiple, as is shown by the overlap in the CG beads for benzene. Each CG site is assigned a type which determines its interactions in the CG force field. From [17].

portions of the system can be coarse-grained out entirely, leading to a drastic reduction in complexity. For example, it is common practice to ignore the bulk solvent surrounding the molecules one finds more interesting, which can reduce the overall number of sites by up to 90 percent. Lastly, the CG force field is much simpler than the corresponding AA one, which leads to a smoother free energy landscape that is sampled over. This allows for the timestep of the simulation to be increased, potentially by an order of magnitude or more, without leading to energy drift.

When developing a CG model, there are several choices which must be made. The first is the mapping itself. There are nearly limitless ways to convert an AA system to a CG one, depending on the desired resolution. This choice is often made with considerations to overall simulation speed, as well as chemical intuition, as individual chemical moieties are often grouped into their own sites. The second choice to be made is the method by which the resulting CG force field is parameterized. There are again two main types of CG models that can be chosen. First is top-down CG, which sees the development of the force field as an attempt to match certain observed properties of the system of interest, such as oil/water partition coefficients. In this regime, it is often experimental data which motivates the force field, and potentials are essentially selected by hand. A common choice for top-down CG modeling is the Martini force field.[17] The second type of CG model is the bottom-up CG model. In this case, the force field is developed as an attempt to algorithmically match the CG system to the behavior of the corresponding atomistic system when mapped to that resolution, as shown in Figure 1.2.[20] Thus, AA trajectories are motivating the force field, which is then generated systematically to minimize a loss function such

as the relative entropy also known as the Kullback-Leibler divergence or the residual forces on each particle.[21–23]

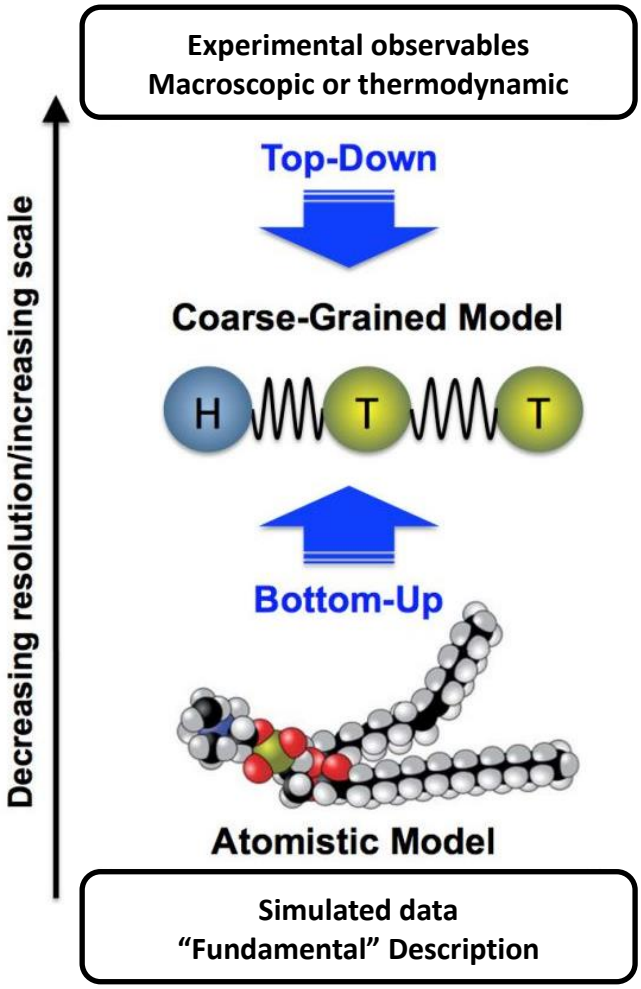

Figure 1.2: The difference between top-down and bottom-up CG methodology. Both methods can be used to generate a CG model with identical bonding topology, but with different behaviors. Adapted from [20].

Both top-down and bottom-up CG have advantages. Top-down CG force fields such as Martini are extensive, and a huge number of molecules can be simulated without any work on the CG modeler's side.[17,24] Martini can simulate systems ranging from glycosylated proteins[25] and lipids[26] to ionic liquids.[27] On the other hand, bottom-up CG models utilize AA reference data to optimize force fields, which when used properly creates a rigorous statistical mechanical bridge between resolutions. In addition, these models can be created systematically, with nothing more

than a CG algorithm and a reference MD trajectory being necessary to make a model. This contrasts with top-down CG models, which require knowledge how the force field was generated as well as extensive testing to extend the models if the system of interest is not represented.

CG modeling, like any other scientific endeavor worthy of study, is not without its own challenges. Namely, finding an acceptable balance between accuracy and speed is non-trivial for systems that are not simple liquids without significant multi-body correlations and often involves application of advanced techniques. For bottom-up CG, this problem is exacerbated by sample data that is required to parameterize the model. Acquiring sufficient sampling must not involve so much simulation that it decreases the overall speed of calculating a result, but for many systems such as lipid bilayers, large amounts of sampling are required.[28] This "data-hunger" is especially difficult in machine learning based CG models. Many modern neural network (NN) based models require microseconds or even milliseconds of reference data, even for small peptides.[29] The choice of algorithm used to generate the model is also important, as certain systems respond much better to certain algorithms than others.

For bottom-up CG, there are several parameterization schemes which are widely used in the field today. For the purposes of this work, there are three which are particularly important. The multiscale coarse-graining method (MSCG), also known as force matching, seeks to solve the CG PMF by minimizing the residual forces between the mapped atomistic reference data and the CG model at those configurations. Typically, MSCG models account only for pairwise interactions, with splines being the functional form of choice in cases where expressive potentials are required, and Lennard-Jones potentials in cases where simpler potentials are sufficient. Relative entropy minimization (REM) on the other hand seeks to match the equilibrium structural distributions of the system. The CG model is evaluated by how well its radial distribution functions match up with

those of the mapped atomistic reference data via the KL divergence, and the CG potential is iteratively updated in response to the gradients of this divergence. The last class of CG models which will be discussed are those parameterized using machine learning, specifically those in which a deep neural network (DNN) is used as the force field itself. While this is a broad class of methods, there are some similarities shared by all which will be discussed. First, these methods inherently incorporate many-body effects, as the individual unit of force evaluation is a single atom and every atom within a certain cutoff radius, all of which are passed into the network simultaneously, as opposed to pairwise based methods where each atomic force is described as a sum of forces between pairs of atoms.

Even once a well-suited mapping and methodology are selected, there are several problems inherent to CG models that must be accounted for. The first is the issue of transferability. Most bottom-up CG methodologies attempt to fit the CG potential as an approximation to the many-body potential of mean force (PMF) of the mapped atomistic representation. That is to say that this PMF is conditioned on the configurations of the atomistic force field which map to the appropriate CG configuration. This PMF is thus dependent on the thermodynamics state point that the system resides in, and because the CG model lacks the detail and thus entropic representation of the atomistic force field, is not necessarily accurate outside of this state.[30] The second problem is that of representability, meaning the CG model's capacity to accurately represent all observables of the system it is based on.[31] For some observables, CG models are correct. For example, in the case of a CG protein model mapped to the alpha-carbon resolution, the positions of the alpha carbons are trivially representable. However, there are many observables for which this fails. For example, the calculation of pressure through the virial equation depends on the dynamical variables of every atom in the system. By mapping these atoms to CG beads by their centers of mass, most of this

information is lost, and thus the pressure calculated from a CG simulation has no bearing on the pressure of the underlying atomistic system. Another example of this is entropy. CG models inherently cannot account for entropy above the CG resolution, such as the rotational entropy of molecules mapped onto single point particles, or the configurational entropy from to internal vibrational modes.[32] While it is possible to alter the CG optimization scheme to properly estimate specific physical observables in the CG regime, this is not always practical nor is it a universal solution.[33]

There is one final piece of the CG puzzle which is often ignored or forgotten by CG method developers: the practicality of using these methods in real applications. Fundamentally a CG model needs to be practically applied to be useful, after all these models are tools to replace AA-MD, a method which is extremely robust and has decades of pedigree to recommend it. For a CG model to be worthwhile, it must be acceptably accurate when compared to the alternative AA simulations which they replace, while also offering a faster overall solution. This speed necessarily includes time taken to generate reference data and build models, which could always just be spent on more AA-MD simulations. Thus at least for bottom-up CG methods, the models start from behind, and data collection and parameterization are often slower than they may first appear.

The theme of this work is to describe recent advancements and challenges in the generation of CG models, particularly those of lipid and liquid systems. The lipid systems focus on high resolution CG mappings of roughly 4 heavy atoms to a single CG site. While many CG models of lipids of this resolution have been in the past, recent advances in both hardware and software have allowed for a modern re-evaluation of these models, and how both bottom-up and top-down CG models perform in this regime. The liquid studies focus on systems for which traditional pairwise

CG force fields tend to fail due to multi-body correlations. In this case, machine learning provides an attractive alternative but often at the cost of slower integration or expensive data requirements.

This work is broken down into four main chapters. The first chapter is an analysis of the Martini 3.0 top-down CG model from a statistical mechanics perspective. Two CG lipid bilayers are simulated and compared with identical systems simulated using the AA CHARMM36 force field. The first is a binary mixture of 70 % 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and 30 % 1,2-dioleoyl-sn-glycero-3-phospho-L-serine (DOPS) to probe Martini's capacity to properly represent the thermodynamics and structural properties of mixed charged and uncharged lipids. The second is a 50 % DOPC and 50 % 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) lipid membrane to probe the same properties in a mixed bilayer composed of saturated and unsaturated lipids. While binary lipid mixtures demonstrate complex behavior,[34,35] Martini treats separate lipid species as much more similar than the physical reality, and is thus unable to capture the behavior of such membranes.

Namely, the DOPC/DOPS mixture contains interspecies interactions at the interface of the membrane which exhibit altered area per lipid due to charge condensation.[36] The DOPC/DPPC bilayer on the other hand exhibits a mixture of unsaturated and saturated tails, which changes the tail ordering and membrane flexibility. While previous studies have examined the effects of cholesterol in a DOPC Martini bilayer using the Martini 2 force field, these bilayers without cholesterol exhibit much more subtle changes from the single species forms.

I demonstrate that Martini 3 models suffer from the same transferability issue just as Martini 2 and bottom-up membranes do, whereby simulation of these models outside the specific thermodynamic state point from which they were parameterized leads to unexpected failures, even when transferability is considered during parameterization. Analysis of Martini 3.0 binary lipid

mixtures shows that these top-down models fail to properly partition their CG PMFs into enthalpic components and temperature dependent entropic components. Perspective is given on the fact that this transferability issue is ever present unless explicit temperature dependence is implemented.

In addition to thermodynamic properties, the structural material properties of each membrane are analyzed through the bending modulus k_c, a measure of how flexible the membrane is, calculated via the height fluctuation spectrum of the membrane. CG lipids often exhibit problematic deviations from the atomistic bending moduli due to a variety of reasons. In the case of Martini 2 membranes containing cholesterol, the membranes are in relative agreement with CHARMM lipids with respect to the bending modulus but deviate in increased fluctuations in the higher frequency modes of the fluctuation spectrum. In the case of Martini 3 lipids without cholesterol, a much larger deviation is uncovered. Cholesterol tends to increase the bending modulus of membranes where they are present,[37] but it is apparent that Martini membranes have a relatively constant bending modulus. For the two systems studied, Martini consistently overestimates the bending stiffness.

The third chapter describes work on high resolution bottom-up lipid models. Here 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) is used as a test case for a variety of CG methodologies. Five models are presented. The first, is a simple force matched bilayer with fully implicit solvent denoted IS-MSCG. This model demonstrates that force-based methods fail when sampling is low, even if that low sampling is limited to less impactful interactions. In the case of IS-MSCG, the head-tail interactions are poorly sampled even when much more sampling is applied to the system, leading to models that require hand tuning to function. Next, I show that REM performs better at this resolution with the IS-REM model but is still not sufficient to obtain a model

that matches the bending modulus of the atomistic system, especially when compared to similar models at lower CG resolutions.

These issues can be addressed by the incorporation of semi-explicit virtual solvent particles, CG sites which represent free solvent particles associated with the surfaces of the membrane yet have no linear mapping to atoms in the fine-grained resolution. These beads add minimal computational cost yet stabilize the membrane via more explicit representation of the hydrophobic effect which drives membrane assembly. This model is referred to as VS-REM and exhibits far more accurate bending modulus than the previous models and forms a stable bilayer without necessary hand tuning and with accurate structural properties. I then apply the virtual particle forces obtained from this model to the reference data used in force matching to generate the VS-MSCG model. Curiously, this model fails for the same reasons the IS-MSCG model does, suggesting that virtual sites cannot overcome the sampling issues that higher resolution CG lipids are prone to.

I demonstrate that temperature transferability can be obtained in such bottom-up models via short reparameterization of VS-REM using reference data at different temperatures, named the TT-VS-REM model. This allows for models which properly partition entropy and enthalpy in the PMFs of lateral association. Lastly, I discuss issues with lipid models at this resolution specifically, where the increased complexity of the system causes pathological failures in these membranes when simulated for long enough. While this could likely be addressed via more sampling of the atomistic system, it is not desirable to do so as the higher resolution doesn't offer enough of a benefit when compared to lower resolution models.

Chapter 4 covers a methodology for generating CG models using DNN-based force fields with minimal sampling. As previously discussed, most DNN based CG models suffer from heavy

data requirements which puts them at a severe disadvantage over more traditional methods, which typically require sampling on the order of picoseconds or nanoseconds. These issues remain true for less complicated systems such as liquids. It is not useful to have a CG model that requires more sampling to generate than it itself would end up generating and thus many of these DNN based models struggle to find use in the current hardware landscape. Thus, I introduce a DNN based CG force field which uses equivariant neural networks. These networks embed the local environment of a CG bead in such a way that imposes force equivariance onto the system, a feature which other DNNs learn via seeing many training examples. This method produces CG models which not only surpass force matching models in terms of accuracy due to the inherent many-body accounting that DNN force fields do, but it does so using orders of magnitude less sampling than non-equivariant networks.

I show that single site water, a system which is typically very difficult to coarse-grain, can be captured very well by several DNN based methods which were designed to match forces from electronic structure data onto atomistic resolutions. However, the methods without explicit symmetry equivariance perform much worse in the low data limit. Equivariant DNNs can produce highly accurate models using only a fraction of the training data.

The fifth and last main chapter discusses machine learned centroid molecular dynamics (ML-CMD), another DNN based method for the learning of effective centroid forces in centroid molecular dynamics (CMD). This method is analogous to coarse-graining in that the full ring polymer representation of a quantum particle based on Feynman Path Integrals (PI) is mapped down to the centroid resolution, and machine learning is used to predict the forces on the centroid. In this case the centroid is the CG site, much as a CG site would typically be matched to the center of mass of the constituent atoms. Unlike typical coarse graining though, ML-CMD accurately

reproduces the dynamics of the PI system since the path centroid is the most classical-like variable upon which dynamics can be accurately computed.

I demonstrate ML-CMD via two common test systems, a single-site para-hydrogen model at its triple point, and liquid water at room temperature. Both models produce highly accurate structural properties of the PIMD data they were trained from. Even more impressive is ML-CMD's ability to accurately capture the dynamics of the systems. The PIMD data used to train the model does not guarantee dynamical fidelity, yet the centroid mapping leads to proper diffusion coefficients and velocity autocorrelation functions for ML-CMD. The method is much faster than traditional CMD, in which the centroid forces are calculated repeatedly on the fly by averaging over the non-centroid forces. By front loading all of the centroid force calculations into the ML training scheme, longer ML-CMD simulations become even more efficient compared to traditional CMD, as there is practically no redundancy in the number of non-centroid to centroid force calculations. ML-CMD gives results which are fundamentally as accurate as CMD and other path integral based methods, while integrating orders of magnitude faster. Special care is taken to demonstrate that this method not only integrates faster but is also faster to use overall accounting for training data acquisition, model training, and integration together. This speedup is significant for the calculation of simple diffusion coefficients and would only get larger for use cases which require more simulation, as is common when designing a CG model.

Finally, I present my perspective on the field of CG modeling, and areas of research that I believe have been ignored by the CG modeling community at large. The first of these is the practicality of creating bottom-up CG models. Many method development papers involve lengthy optimization schemes which are difficult to justify for users who are not familiar with the many ins and outs of CG model development. I posit that this issue could be somewhat addressed with

a larger focus on application than method development, where large quantities of CG models could be generated and tested. This would provide the field with a better understanding of what does and does not work and a better set of heuristics for generating more complex models. It would also allow bottom-up methods to be used “out of the box”. Top-down models such as Martini are very easy to use as almost any system can be simulated without additional parameterization. If bottom-up modelers worked towards generating libraries of models, many common use cases could be covered without requiring anyone to generate anything new. In fact, while CG modelers are adept at analyzing the accuracy of their models and methods, it is often practicality which drives users towards certain models, and pre-generated models are by far the most practical way to use a CG model. Bottom-up CG methods must offer a corresponding increase in accuracy and predictive power if they require users to conduct reference simulations, learn a new codebase and parameterize their own model.

With respect to lipid bilayers, I argue that there is a “quality trough” at higher CG resolutions. The higher the resolution of the CG model, the more information must be built into the model not only to make the model consistent with the atomistic data, but also to make the model functional at all. At high resolutions there are many more points of failure which can occur, which can truly only be resolved by spending more time on generating training data or incorporating advanced CG methodology, undercutting the effectiveness of the model from the beginning. Beyond this, these higher resolution bottom-up CG lipids are quite impractical to generate. The methods discussed in chapter 3 take a significant amount of time and expertise to apply effectively, to the ultimate detriment of the unwary CG modeler who wants to get a simulation running quickly. I discuss speed as a key element of a CG model and argue that the fact

that higher resolution lipid models tend to fail is positive. Lower resolution models truly have little downside in this regard, in that they are the most accurate and the fastest bottom-up CG lipids.

Finally, I discuss machine learning (ML) based CG methodologies and give my perspective on their usefulness now and in the future. While many issues can be resolved via incorporation of fundamental physics into DNN models (such as symmetry equivariance), there is still one large issue that DNN based force fields must overcome: speed. Neural networks may be faster to integrate than quantum calculations, making them highly desirable for QM/MM or AIMD based methods; they are far slower than simple pairwise basis sets, where CG models typically find themselves. If there is to be an efficient ML CG model, it must rely on future technological advances. These advances may be in the form of specialized hardware which can evaluate neural networks far faster, or it may come in the form of new algorithms to simplify the calculations. Despite this, I am optimistic that ML based CG methods will continue to produce the most accurate and expressive CG models and may be standard in the future. Additionally, ML still has many use cases in CG modeling, particularly in the generation of more accurate pairwise force fields and in the evaluation and backmapping of existing CG models.

Chapter 2: The Lens of Statistical Mechanics Reveals Why a Different Martini Can Still Give You a Hangover

Abstract:

The Martini 3.0 force field, which was parameterized to better capture temperature transferability in top-down coarse-grained models, is analyzed through the lens of statistical mechanics to benchmark how well it represents thermodynamic and material properties of the underlying atomistic representation. To that end the potentials of mean force for lateral association in Martini 3.0 binary lipid bilayers are decomposed into entropic and enthalpic components and compared to those of corresponding atomistic bilayers which have been mapped onto equivalent coarse-grained sites. This is accomplished by applying the reversible work theorem to lateral pair correlation functions between coarse-grained lipid beads taken at a range of different temperatures. These entropy-enthalpy decompositions provide a metric by which the underlying statistical mechanical properties of Martini can be interrogated as they relate to temperature transferability. Overall, Martini 3.0 fails to properly partition entropy and enthalpy despite changes to the force field from the Martini 2.0 version to address the transferability problem. The fact that these issues remain unresolved is a clear demonstration that to create a truly transferable model, there must be explicit temperature dependence involved, as a coarse-grained model parameterized without such

dependence is inherently state dependent. In addition to the entropy-enthalpy decompositions, Martini 3.0 also fails to produce lipid bilayers with the proper bending modulus, with each model tested being consistently less flexible than corresponding all-atom membranes.

2.1 Introduction:

Despite the continual technological advancements in computing power, the simulation of many large, relevant biological processes remains tantalizingly out of grasp. For example, the maturation of HIV-1 viral particles occurs on the timescale of minutes, far outside the reach of all-atom molecular dynamics (AA-MD)[38] despite the fact that atomistic simulations of the entire viral have been performed.[39] For this reason, there is an ever-present need to enhance the speed at which MD simulations calculate a result. The two major paths for accomplishing this are enhanced sampling, in which computation is sped up by weighting the statistics of the system of interest in favor of sampling the desired phenomena, and coarse-graining (CG), in which the entire system is simplified by collapsing groups of atoms into individual CG "beads".

CG models can also be divided into two categories: bottom-up and top-down. Bottom-up CG involves parameterizing interactions between beads based on all-atom molecular dynamics data algorithmically by minimizing the difference between the CG model and the mapped atomistic data.[15,19] Top-down CG on the other hand involves parameterizing each interaction by hand in order to match certain macroscopic target properties.[17,40–42] Each methodology has advantages: top-down CG models are easy to implement due to mature model libraries, but bottom-up CG models are often easier to extend, as anyone can systematically generate their own model of a given system provided they have sufficient AA sampling. In addition, bottom-up CG models

have a rigorous correspondence with the atomistic data.[43] This is advantageous as CG models tend to be used as a replacement for AA MD models. By guaranteeing at least a theoretical connection between the CG model and the AA model which it replaces, bottom-up CG provides a basis upon which to judge the CG model.

Martini, developed by Marrink and co in 2007[17], is a top-down CG model parameterized via matching partitioning free energies of various compounds between polar and non-polar phases. The result was a simple and effective set of CG sites mapped to a 4-heavy atom to 1 bead resolution that could be combined to represent a variety of different molecules, which was quickly extended to include small molecules, proteins, lipids and carbohydrates.[44,45] This model is notable for its treatment of the bulk solvent as large beads which represent a collection of 4 water molecules. These solvent beads tend to freeze at or around room temperature, a problem that was resolved via smaller anti-freeze particles that disrupt nucleation points for Martini ice. Later, a "Dry" Martini model was introduced which avoids representing solvent altogether, further enhancing the efficiency of the model at the cost of accuracy.[46] Solvent free CG models such as this are common both in top-down and bottom-up CG due to the fact that in nearly all biophysical systems, the majority of the simulation is composed of water molecules. Most recently, Martini 3.0 was released, which addressed previous concerns about temperature dependence as well as greatly expanding the overall number of possible interaction types.[24]

Martini is a go-to CG model for several reason. It has a massive library of molecules, easy to use tools and a 4 heavy atom to 1 CG bead resolution which retains a large amount of chemical specificity while simultaneously offering a significant speedup over all-atom simulations, especially when utilizing the solvent-free "dry" Martini models. This pre-generated library of molecule types makes using Martini follow a very similar workflow to using other AA force fields.

Additionally, the reduction in degrees of freedom brings with it a smoother free energy surface upon which the system evolves, speeding up diffusion and thus sampling beyond a simple uplift in integration rate.[47] Many complex systems have been built and studied within the Martini framework, including a 63 component asymmetric plasma membrane model[48] and more recently, an entire "cell",[49] both of which would be simply impossible to accomplish via bottom-up CG given state of the art methods.[50,51]

However, there are several disadvantages to using the Martini force field. The 4 to 1 mapping scheme employed still leads to issues of chemical specificity. Notably there is a degeneracy of molecules that map to certain Martini topologies. For example, both 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC) and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) map to the same Martini molecule. For this reason, the Martini model lacks a certain level of rigorous connection to the underlying atomistic representation and leads to the conclusion that Martini is a semi-phenomenological model rather than a model which seeks to represent the chemical properties of each of its molecules. Issues stemming from the 4 to 1 mapping scheme also crop up when focusing on Martini water. As previously mentioned, each Martini water bead represents 4 water molecules, which causes inaccurate freezing behavior to occur near room temperature, as well as issues pertaining to diffusion and the hydration of solute molecules, all of which are critically important to the study of biophysical systems.[52–55]

Due to its popularity, it is important to understand Martini's limitations as a model. Not only does this allow the scientific community to better interpret results from Martini simulations, but it also helps guide its use towards cases in which it is sufficiently accurate, as well as help guide the development of more accurate top-down CG models in the future. Recently, an analysis of Martini 2.0 and dry Martini lipids were performed which suggests that the models are limited

not only in structural bilayer properties, but more critically in the thermodynamic decomposition of their potentials of mean force. Specifically, in the case of a 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC)/Cholesterol bilayer, Martini 2.0 failed to properly partition the entropy and enthalpy of the CG potential of mean force for lateral association for the head groups as well as the glycerol beads.[56]

As of the time that this data was collected, Martini 3.0 cholesterol is unavailable, and thus this paper focuses on bilayers without cholesterol as a direct comparison is impossible. However, as lipid bilayers are an excellent case study for the association of long amphipathic molecules, which are very similar to the parameterization methodology used for Martini in the first place (water/octanol partitioning) we chose to retain the focus of this work on lipid bilayers. Instead of simulating a pure DOPC bilayer as a comparison, simulations of two binary lipid mixtures were simulated to study the thermodynamic properties of Martini interactions. The first system is a 70:30 mixture of DOPC and 1,2-dioleoyl-sn-glycero-3-phospho-L-serine (DOPS). Analysis of this mixture provides insight into Martini's ability to partition charged and uncharged lipids. In this case, the heterogeneity of the bilayer is localized within the head groups of the membrane. The second system selected is a 50:50 mixture of DOPC and 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), to study Martini's ability to capture saturated lipid tails interacting with unsaturated ones. While the prior study focused its analysis purely on the interactions between lipid head and glycerol groups, the current work extends this treatment to the analysis of the lipid tails as well. It is expected that in general Martini 3.0 will capture lateral association of lipid tails more accurately than the head groups, though currently there is no study which directly analyzes Martini's capacity in this regard from a statistical mechanical standpoint.

In this work we analyze Martini 3.0 from the perspective of statistical mechanics, using the two membrane systems to probe how Martini lipids capture temperature transferability and higher order structural properties, namely the bending modulus k_c. First, the theory connecting the RDF to the decomposition of entropy and enthalpy in CG PMFs is discussed. Then the connections between simulated lipid bilayers and continuum models is explained in order to motivate the methodology behind our calculations of bending modulus. We then describe the computational details of the work, including details necessary to reproduce each molecular dynamics simulation performed, and details justifying the method by which we calculated the RDFs used in the entropy-enthalpy decompositions. Results for the entropy-enthalpy decompositions and bending moduli for each system are then presented, followed by analysis of how these results show that the latest release of Martini still cannot capture thermodynamic transferability without an explicit temperature dependence built into the model.

2.2 Theory

There is an inherent state dependence incurred when creating a molecular model at a CG resolution. This dependence arises due to the loss of entropy due to mapping.[30,57] As each CG configuration corresponds to a number of different atomistic configurations, the act of mapping conditions pins the CG model to a specific thermodynamic state point.[30] While a large portion of simulations are performed at a single common thermodynamic state point (that is 298 K and 1 atm), simulating a CG model outside of this state point (or whichever point the model was designed for) can lead to unphysical behavior as the PMF of the system will fail to correlate with the conditioned PMF of the corresponding system.[32]

Statistical mechanics offers a quantitative way to probe these PMFs via entropy-enthalpy decompositions. By the reversible work theorem, the radial distribution function can be connected to the PMF of association between two CG sites.

$$g(r) = e^{-\beta w(r)} \tag{2.1}$$

In equation 2.1, β is equal to $1/k_B T$, g(r) refers to the RDF, while w(r) is the value of the PMF at some distance R. As a PMF is a free energy, it in turn can be expressed as a sum of two components, enthalpy (ΔH), and entropy multiplied by negative temperature, $-T\Delta S$.

$$w(r) = \Delta H - T\Delta S \tag{2.2}$$

For each value of r, the values of ΔH and ΔS can be estimated via linear regression of the PMF with respect to temperature, where the intercept represents enthalpy, and the slope corresponds to the negative of the entropic component.[58]

For CG models such as Martini, this analysis provides insight into how the missing configurational mapping entropy is accounted for in the model.[59] As Martini 3.0 was developed with an eye on temperature transferability, this work should be able to quantify these improvements in lipid systems. It will provide a clear indication as well of where the model fails (i.e. head vs tail groups) and how such models can be improved to increase their transferability. We hypothesize that as Martini was parameterized with hydrophilic/hydrophobic partitioning in mind, the tail groups of Martini lipids should perform better in this regard than the head groups, as lipid tails provide a similar chemical environment to octanol. However, this is dependent on Martini's ability to generalize from a system in which the hydrophobes are shorter and have little

to no lateral organization to a lipid bilayer in which there is far more ordering present. It is possible that errors in the organization of the head groups, to which the tails are attached, could limit Martini lipid's ability to associate correctly.

While entropy-enthalpy decompositions offer a statistical mechanical insight into the thermodynamics and transferability of the Martini model, there is also the problem of representability in CG models. Representability refers to a CG models capacity to properly represent quantifiable observables from the AA regime. There are many observables which are patently impossible to correctly calculate for any CG model, such as entropy, due to the loss of entropy incurred when mapping as discussed earlier. In addition, many observables are trivially representable for CG models, such as the location of alpha carbons in a CG protein which is mapped down to an alpha-carbon resolution.[32] Previous study of Martini 2.0 lipids have used the bending modulus as a test for Martini representability, as well as its accuracy as a whole in its capacity to properly capture equilibrium structural properties of lipid bilayers.

The bending modulus of a material refers to its propensity to bend when subjected to a strain normal to the surface of the material itself. While this is easy to calculate for rigid objects as merely the ratio of stress to strain, it is more complicated for flexible sheet like objects such as lipid membranes. A common way to calculate this property for simulated bilayers is via the height fluctuation spectrum. This spectrum is calculated via a discrete Fourier transform of the midplane of the bilayer:[60,61]

$$u(q) = \frac{1}{2N}\sum_{k=1}^{N}\sum_{j}^{2} z_{j,k} e^{-iq\cdot r_{j,k}} \qquad (2.3)$$

where *u* is the Fourier coefficient, *q* is the two-dimensional reciprocal space vector, *N* is the number of lipids, k and j count lipids and leaflets respectively, *z* is the z coordinate of a given lipid and r is the x, y coordinate of a given lipid.

This spectrum is connected to the height fluctuation spectrum in Canham-Helfrich continuum models of lipid bilayers.[62,63] For such a system, the free energy can be related to the bending modulus via

$$E(u\{x,y\}) = \frac{1}{2}\int [k_c(\Delta u)^2 + \gamma(\nabla u)^2]dx\,dy \tag{2.4}$$

where E is the free energy, k_c is the bending modulus and γ is the area compressibility of the membrane. A continuous Fourier transform, and removal of harmonic Fourier modes can relate the Fourier modes and the bending modulus like so

$$\langle A|u(q)|^2\rangle = \frac{k_bT}{k_cq^4 + yq^2} \tag{2.5}$$

where A is the area of the lipid patch at that moment, and $1/k_bT$ is the typical thermodynamic beta. In the long wavelength regime of this equation and assuming no surface tension, the bending modulus can be estimated. This function relies on a large enough lipid patch to properly sample low wavelength modes, as it is only at long ranges that molecular systems of lipids begin to share behavior with continuum models. This methodology has been applied to both AA and CG models alike, and the methodology discussed by Brandt *et al.* is used in the present work for calculations.[60] It is important to note here that comparisons of AA and CG bending moduli rely on proper representation of the height fluctuation spectrum. In previous efforts, it is shown that for Martini mappings, the position of the phosphorus atom in lipid heads is represented well by the position of Martini mapped phosphate groups[56]. This makes sense, as the phosphate bead is made up of this

phosphorus atom and the four covalently bonded oxygens. As the phosphate moiety is tetrahedral and thus the phosphorus atom occupies the center of mass of the entire moiety, it should be no surprise that the average phosphate bead z coordinate is an effective metric to calculate the midplane of the bilayers. As noted, this may not be the case for all beads, so the phosphate bead for the mapped AA data was used here as a direct comparison with Martini 3.0.

2.3 Methods:

In total, 4 systems were simulated to gather data necessary for entropy-enthalpy decompositions and bending modulus calculations. For each lipid mixture, an atomistic simulation was performed using the CHARMM36M force field in GROMACS.[64,65] For each system, 4 separate simulations were carried out, at 290, 305, 320 and 335 K in order to sample a variety of temperatures for entropy enthalpy decomposition calculations. Each system consisted of 1152 lipid molecules and was first annealed to the appropriate temperature before being simulated for 200 ns under NPT conditions. Each trajectory was stripped of water and then split into two trajectories, one for each leaflet for the calculation of entropy enthalpy decompositions, justifications for which are below. For the bending modulus calculations, the 305 K split trajectories were combined to obtain a full bilayer. The AA simulations were then mapped to Martini resolution using a center of mass mapping scheme in order to more directly compare the CHARMM and AA models.[66,67]

For the CG models, the process was largely the same. The Martini 3.0 force field was utilized[24], with each system run at 290, 305, 320 and 335 K. However, each simulation was stopped after 100 ns of NPT integration. It was found that due to Martini's faster lipid diffusion, less sampling was required to obtain converged RDFs.

The membrane patches simulated in this study were larger than those used in the analysis of Martini 2. In that paper, lipid patches containing 268 DOPC molecules and 70 cholesterol molecules were used, which necessitated much longer simulations in order to obtain RDFs that converged to the degree necessary to calculate accurate entropy-enthalpy decompositions.[56] All simulations were prepared using the CHARMM-GUI membrane builder browser based tool.[68,69]

xy-RDFs:

To properly study the thermodynamics of lateral association in the lipid membrane, all radial distribution functions used in this work were projected onto the two dimensions parallel to the plane of the lipid membrane on a per leaflet basis. These RDFs are hereby referred to as xy-RDFs. By projecting the RDFs in this way, the anisotropic nature of the bilayer is properly handled. If full 3-dimensional RDFs were calculated, the results would not normalize to 1 as a proper RDF does, but to 0, as there is no sampling of particles above or below the membrane. This also guarantees that the PMFs calculated via the reversible work theorem correspond to the PMF of lateral association, which is of unique interest for membrane systems. Doing so also necessitates that each RDF is calculated with respect to only the CG beads in the same leaflet. If this were not done, one would find non-zero values for the xy-RDF near or at $r = 0$, as the distance between beads on different monolayers projected onto the xy plane is not hindered by steric constraints. The interactions of lipid species in the opposite leaflet are not significant, especially for interactions between head groups, but this inclusion would significantly impact the overall RDF calculated. In order to calculate these xy-RDFs, the following method was used. First the solvent was stripped out of the system, as it is not of interest for the calculation of lipid-lipid RDFs or fluctuation spectra. Next, for AA systems, the lipid atoms are mapped onto a Martini-equivalent resolution, specific details of which can be found in the supplemental information. This way, the

AA reference simulations are as close to the Martini system as possible. Next, each trajectory was split into separate trajectories containing a single lipid leaflet. Finally, xy-RDFs were calculated for each leaflet and averaged to obtain a converged RDF for the entire simulation.

For the calculation of lateral association RDFs and entropy-enthalpy decompositions, each bead in each tail was treated distinctly. All results presented utilize the correlations between the beads labeled as part of the "A" tail in the Martini mapping. Because the A and B tails are identical, this result is consistent across each tail, and figures comparing the B tails can be found in the supplementary information along with additional figures detailing RDFs and entropy-enthalpy decompositions not shown in the next section.

2.4 Results and Discussion

Lateral association

System 1: DOPC/DOPS

Figure 2.1 shows comparisons between Martini and CHARMM xy-RDFs for lateral association between DOPC molecules. As expected, Martini matches the distributions of the tail beads well, although it produces slightly over-structured results compared to the AA system. The biggest difference in the RDFs of the tail groups comes in the C1A group, the first tail bead which is bonded to the glycerol group. In this case, Martini fails to capture the location of the first peak properly. For the head groups, Martini 3.0 largely fails to reproduce the RDFs of CHARMM lipids. Figure 2.2, which shows the same RDFs between DOPS molecules, shows similar results. However, there is further deviation between CHARMM and Martini in associations between DOPS serine groups (named CNO in Martini). In this case the AA bilayer exhibits a much larger

amount of structuring around 0.5 nm. Figure 2.3 shows the lateral association RDFS between the two lipid species. These RDFs do not exhibit deviations beyond the single component RDFs.

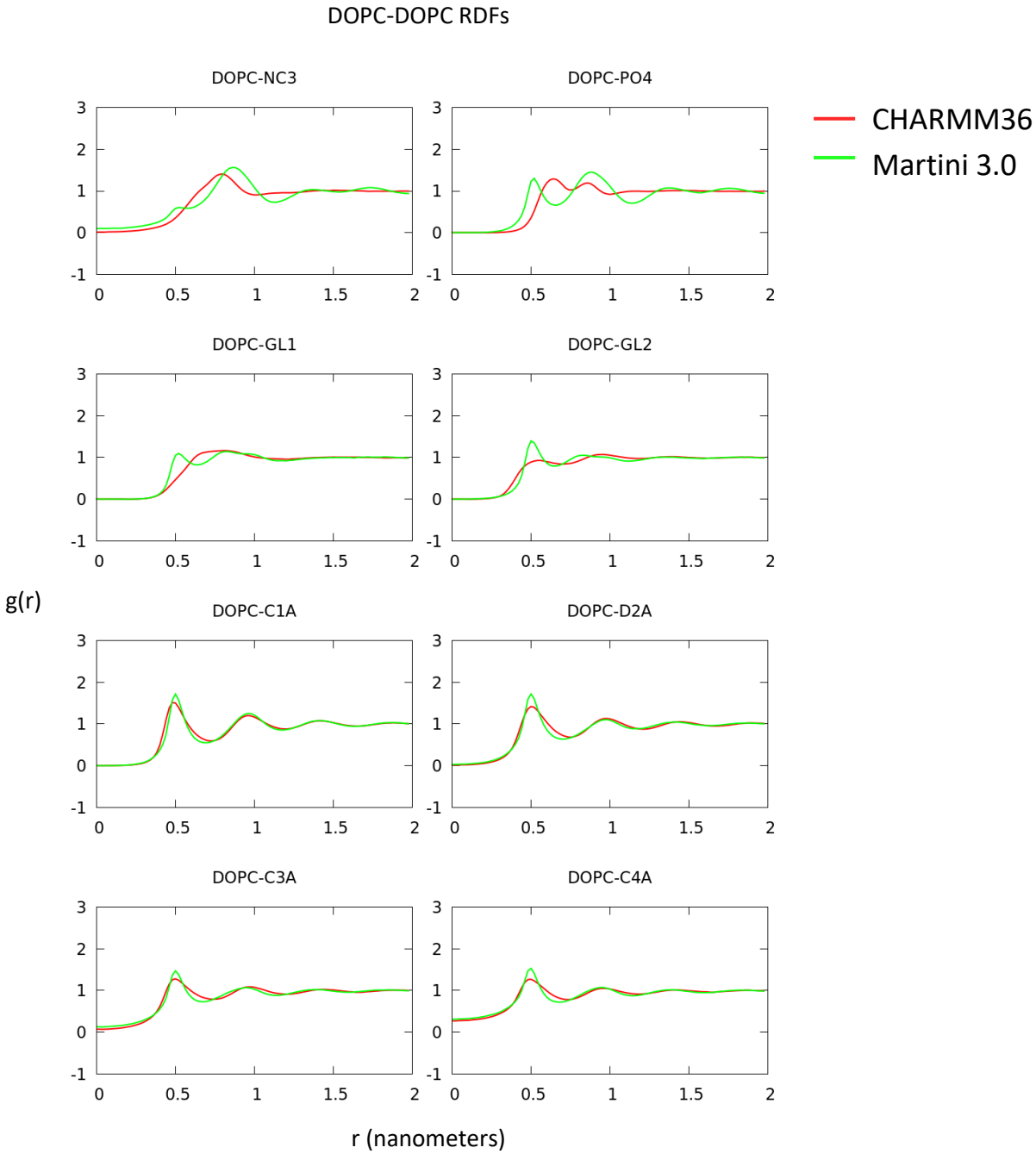

Figure 2.1: Lateral RDFs for DOPC-DOPC interactions in DOPC/DOPS system. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C3A and C4A, Unsaturated tail beads: D2A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

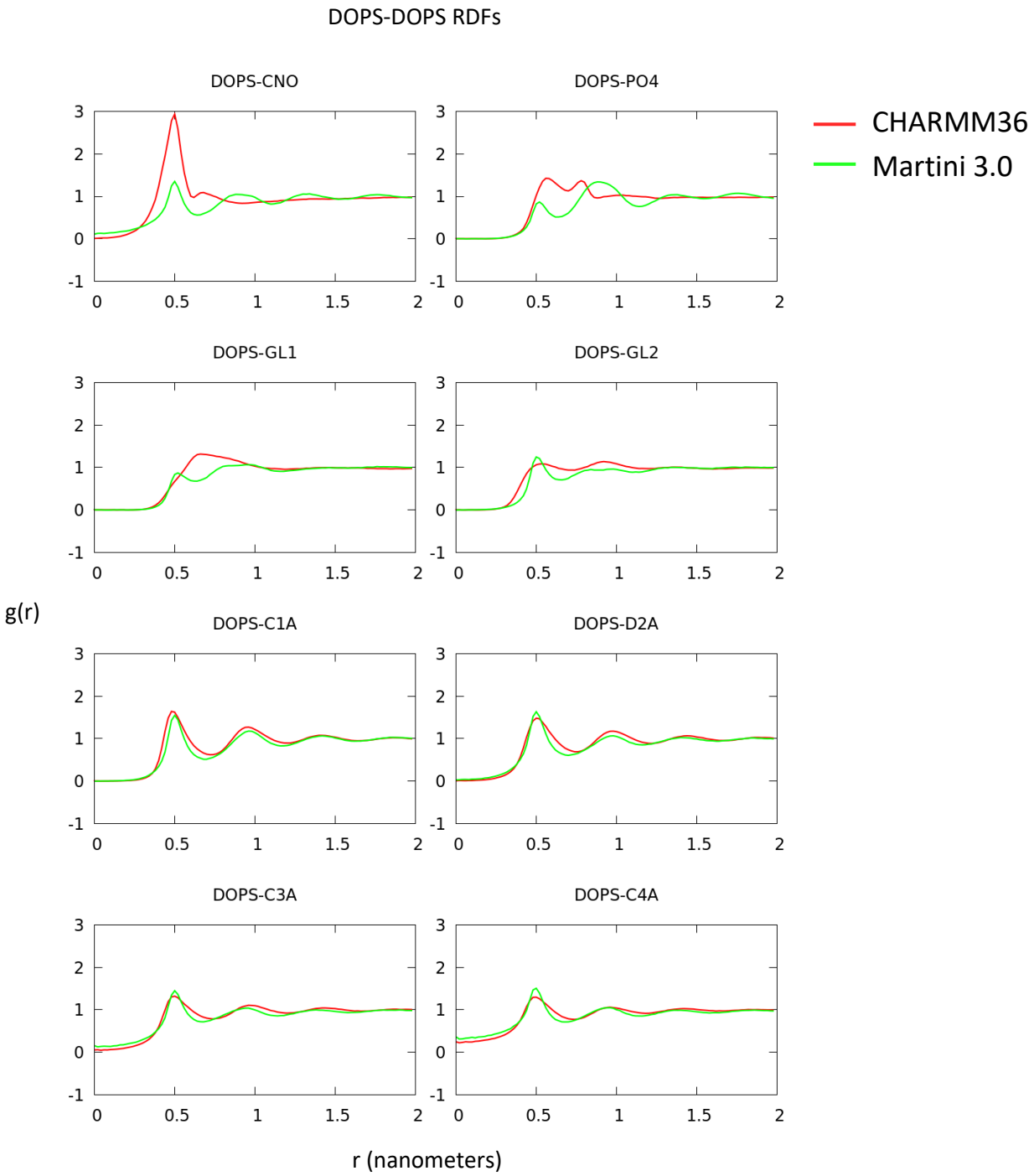

Figure 2.2: Lateral RDFs for DOPS-DOPS interactions in DOPC/DOPS system. Beads are labelled according to the following scheme. Serine: CNO, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C3A and C4A, Unsaturated tail beads: D2A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

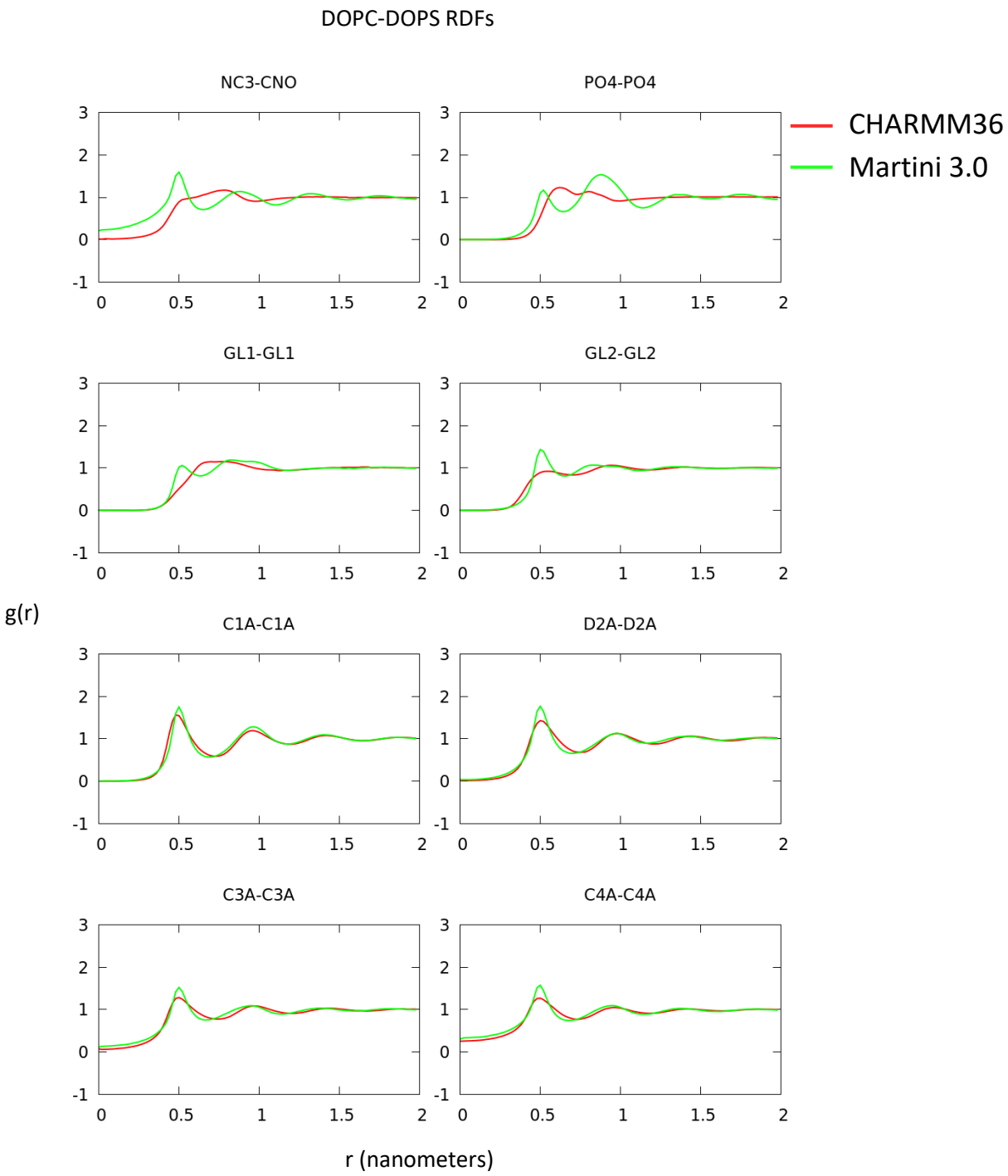

Figure 2.3: Lateral RDFs for DOPC-DOPS interactions in DOPC/DOPS system. Beads are labelled according to the following scheme. Choline: NC3, Serine: CNO, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C3A and C4A, Unsaturated tail beads: D2A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

System 2: DOPC/DPPC

The DOPC/DPPC system for the most part shows similar results for lateral association. Figures 2.4 and 2.5 show results for DOPC-DOPC and DOPC-DPPC association. For the most part Martini can capture the structuring of lipid tails with peaks at the correct distances and most peaks being the correct magnitude. The head groups, as in the DOPC/DOPS system, are largely inaccurate, however there are no large deviations seen in the DOPS-DOPS serine-serine RDF in the previous system. Larger inaccuracies are seen in the DPPC-DPPC tail interactions, shown in figure 2.6. As opposed to RDFs from the previous system and in DOPC interactions in this system, DPPC-DPPC tail interactions fail to capture peak locations for each bead. This is not limited to the first peak either, each subsequent peak is also located further away than in the CHARMM reference data, suggesting that Martini DPPC tails are effectively larger than they should be.

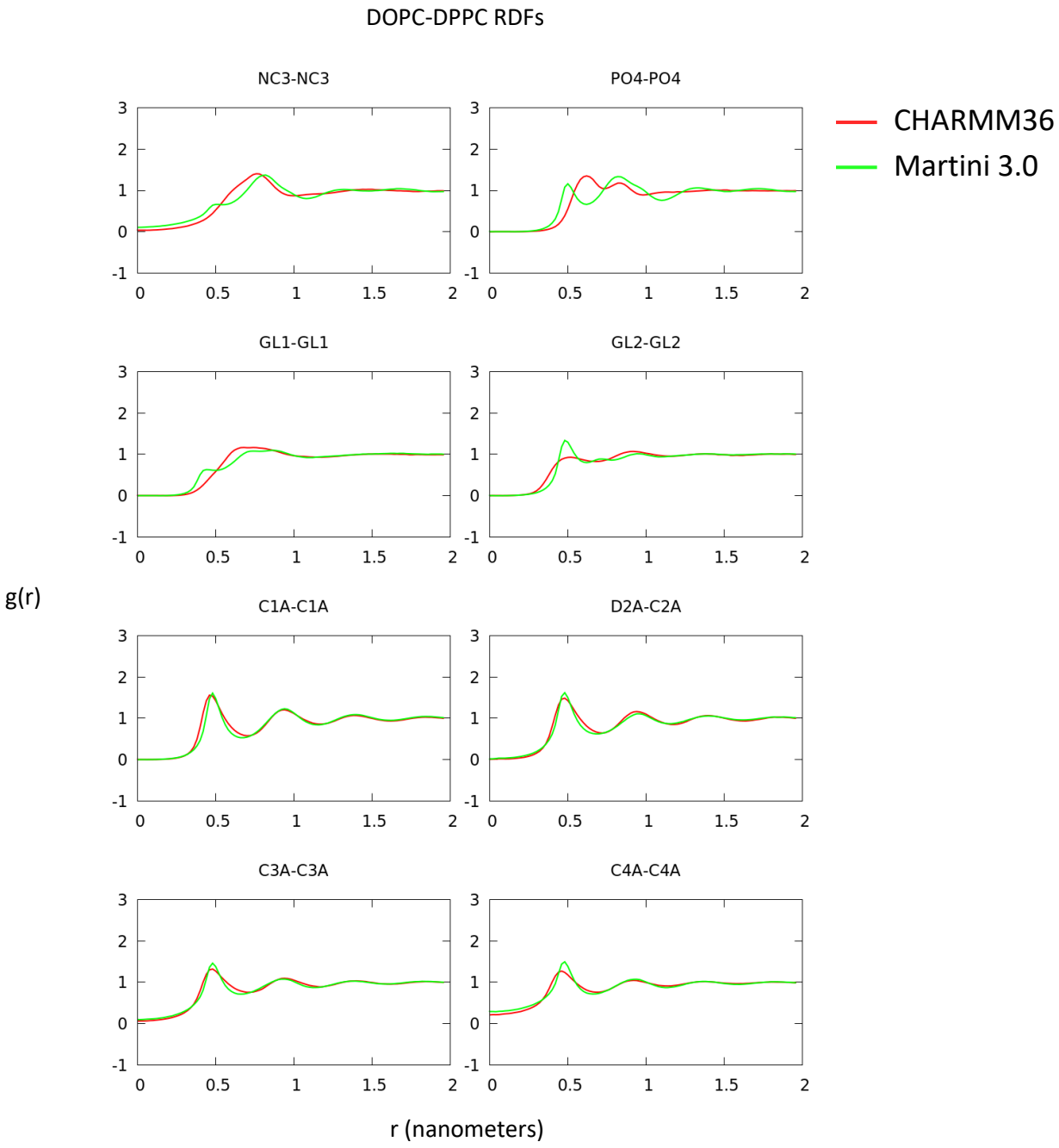

Figure 2.4: Lateral RDFs for DOPC-DPPC interactions in DOPC/DPPC system. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C2A, C3A and C4A, Unsaturated tail beads: D2A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

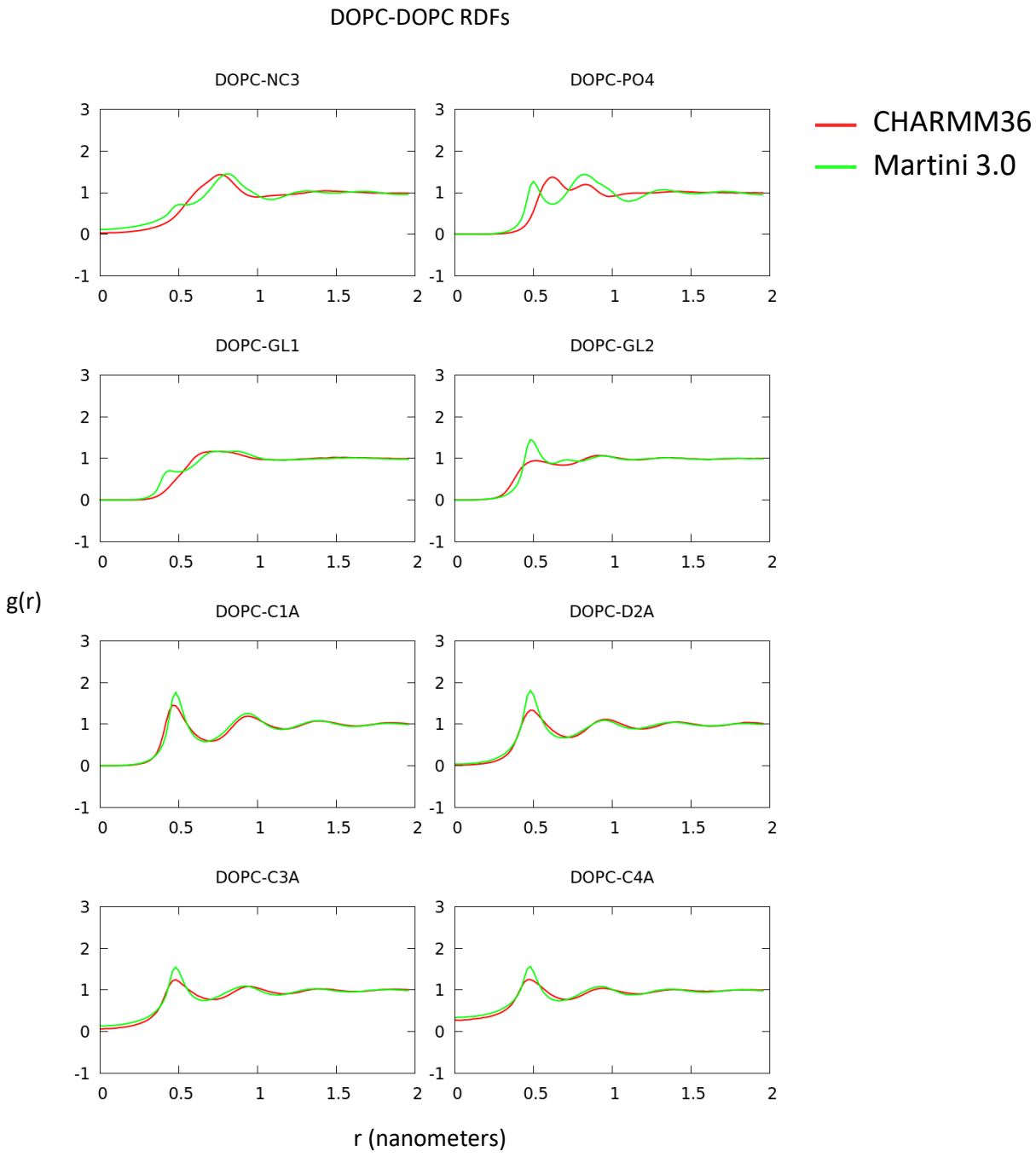

Figure 2.5: Lateral RDFs for DOPC-DOPC interactions in DOPC/DPPC system. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C3A and C4A, Unsaturated tail beads: D2A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

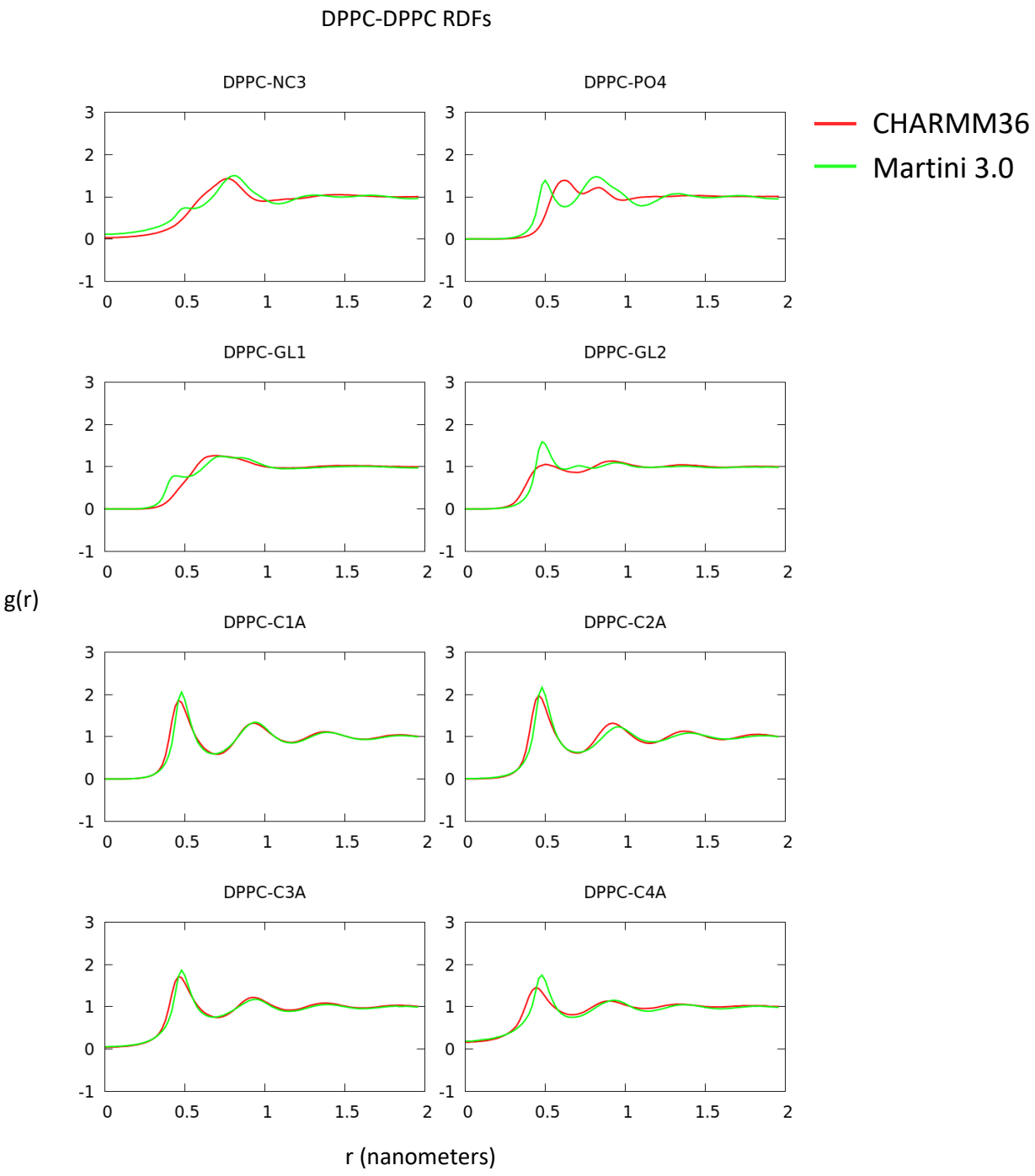

Figure 2.6: Lateral RDFs for DPPC-DOPC interactions in DOPC/DPPC system. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2, saturated tail beads: C1A, C2A, C3A and C4A. Martini 3.0 RDFs are compared to mapped CHARMM36 data.

Entropy-Enthalpy Decompositions:

System 1: DOPC/DOPS

Figure 2.7 shows entropy-enthalpy decompositions for head group DOPC-DOPS interactions. While the PMFs for the glycerol groups are decomposed accurately at least at long ranges, the phosphate-phosphate and choline-serine decompositions exhibit large deviations from the atomistic data. Figure 2.8 shows the decompositions for the tail group DOPC-DOPS. While there are deviations in the peak locations and heights, overall, the tail groups are captured well, with the proper overall trend. Of note, the errors for each value of r are shown as well. At lower r values, these errors raise in magnitude due to the lack of sampling in these regimes. In certain cases, no sampling for certain bins was obtained, and data is not plotted for these regimes. Figure 2.9 shows decompositions for the head groups of the DOPC-DOPC interactions in the DOPC/DOPS system. Again, phosphate beads are poorly captured by Martini while the glycerol beads are qualitatively similar. Interestingly, the choline groups show an additional feature that is not present in choline-choline decompositions for the second system: a strong peak after the initial well which is not observed at all in the atomistic data.

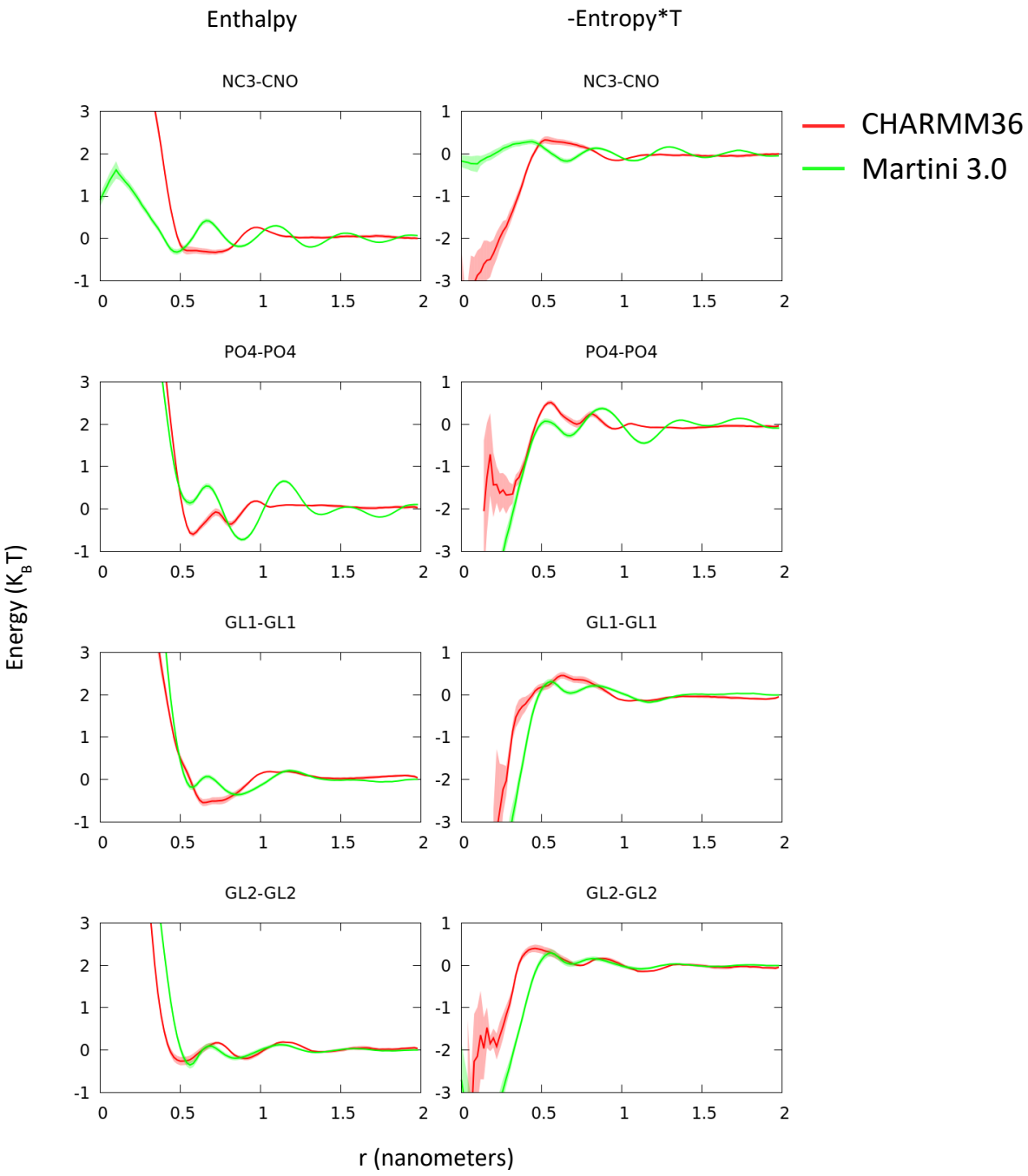

Figure 2.7: Entropy-enthalpy decompositions of DOPC-DOPS head group interactions for DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

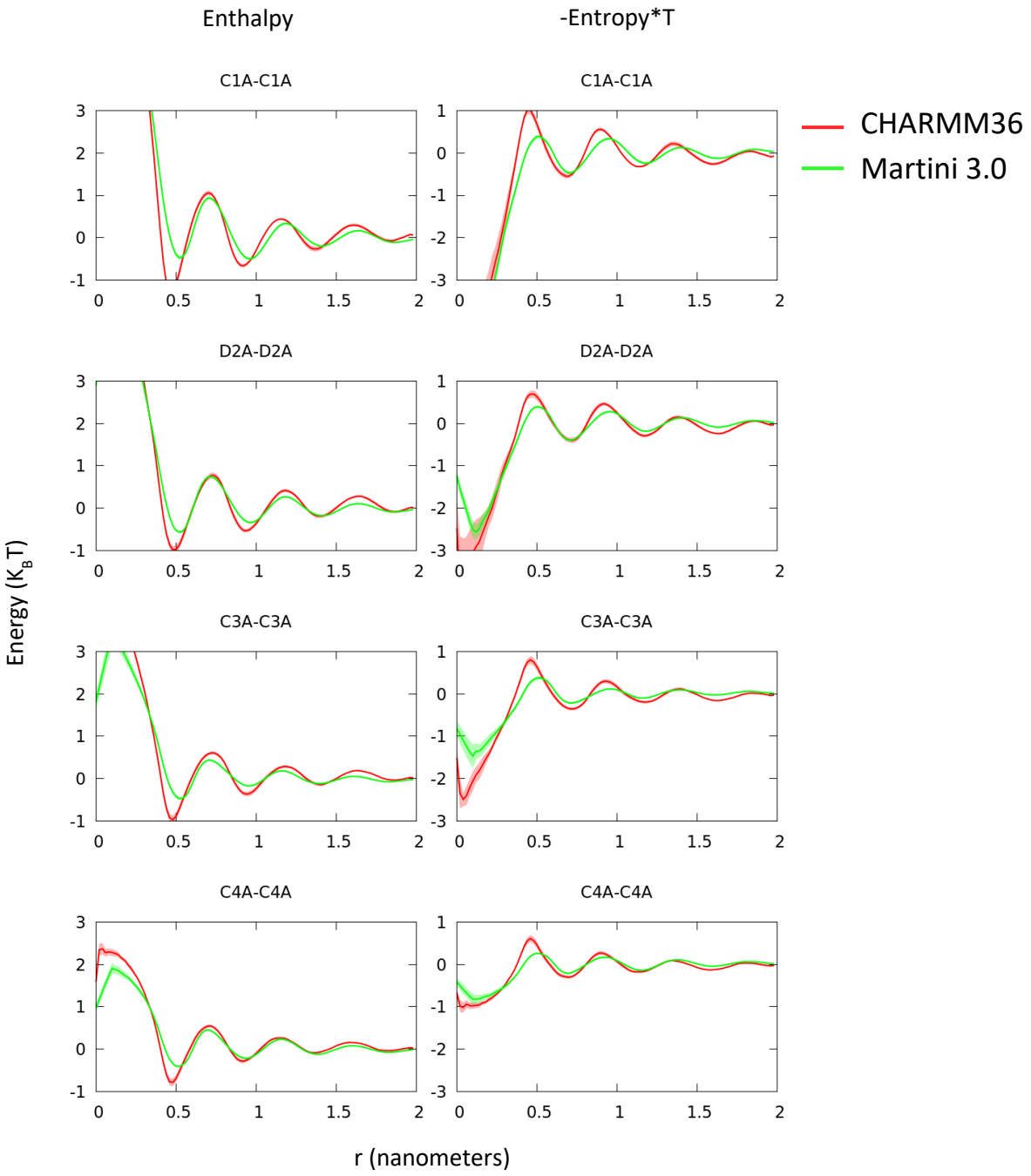

Figure 2.8: Entropy-enthalpy decompositions of DOPC-DOPS tail group interactions for DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Lipid tail beads in descending order: C1A, C2A, C3A and C4A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

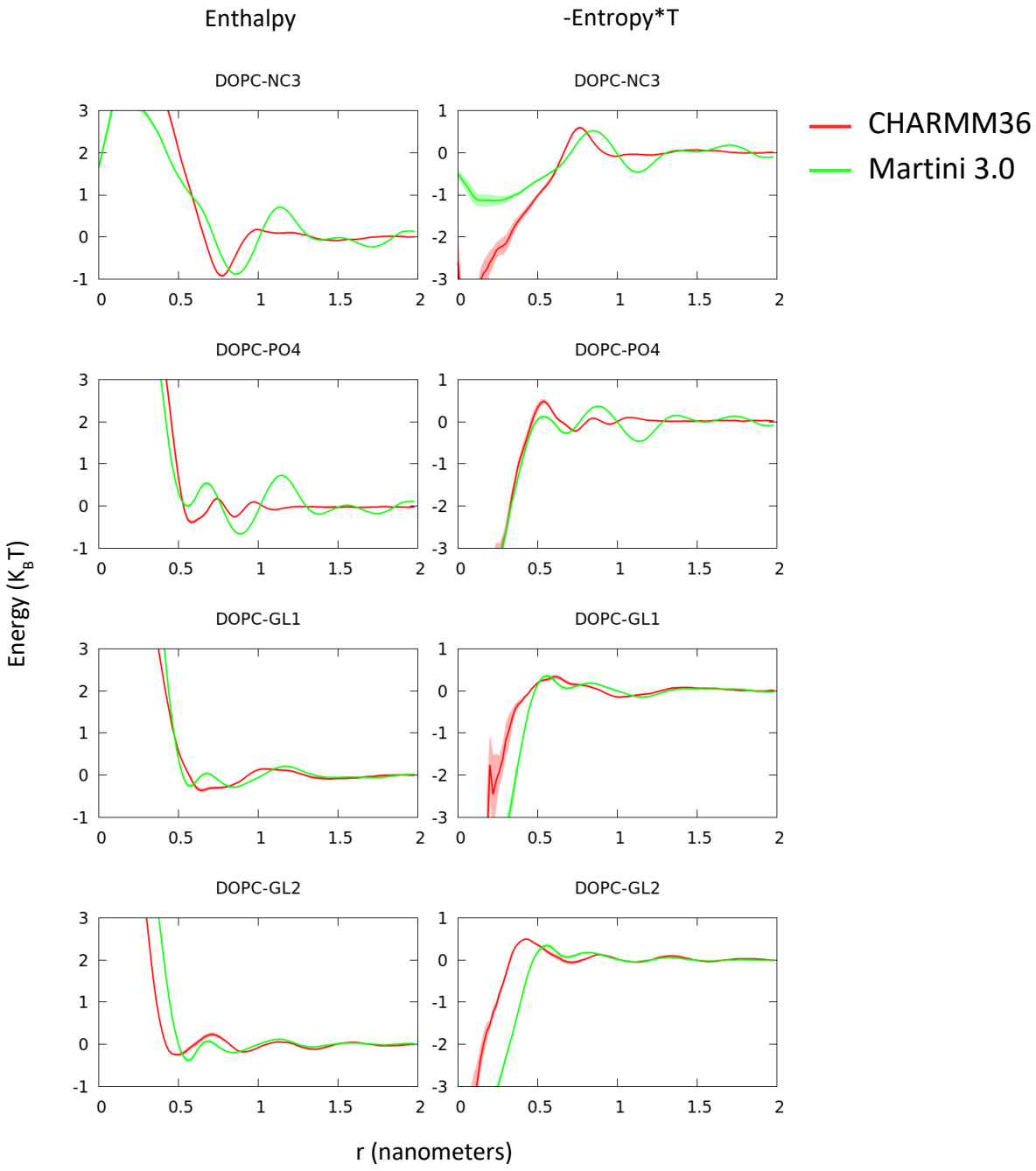

Figure 2.9: Entropy-enthalpy decompositions of DOPC-DOPC head group interactions for the DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

System 2: DOPC/DPPC

The mixed DOPC-DPPC interactions show similar trends for the second system simulated for both phosphate and glycerol groups, shown in figure 2.10. In the case of phosphate, the Martini lipids produce significant structuring which is not present in the mapped atomistic system. The glycerol groups on the other hand show a rough qualitative agreement with the CHARMM lipids, especially at long ranges. The choline-choline decompositions on the other hand are much closer to the reference simulation than the serine-choline decompositions in the DOPC/DOPS system. Aside from some slight structuring at long ranges which is not present in the atomistic data, the decompositions have roughly the same features, though the principal well at around 0.75 nm and the subsequent peak have been pushed further away.

The tail-tail decompositions for the mixed DOPC-DPPC interactions, detailed in figure 2.11, follow the trends of their corresponding RDFs. The under-structuring of tails by Martini is worse in this system than in the DOPC/DOPS system, and the locations of the peaks are all pushed out from the AA reference data. This error is the most prominent in the C2A-D2A interactions in contrast with the DOPC/DOPS system, while the C4A-C4A decomposition almost matches the properties of CHARMM36. Interactions between molecules of the same species follow the same trends with one major exception. The interactions between CHARMM36 DPPC tails exhibit much higher contributions both in the enthalpic and temperature dependent entropic terms than those between DOPC molecules and the DOPC-DOPS tail interactions, as shown in figure 2.12. Martini 3.0 fails to capture this entirely, and the DPPC tails seem to behave much closer to those of DOPC. Effectively, the Martini DPPC tails are much less temperature dependent.

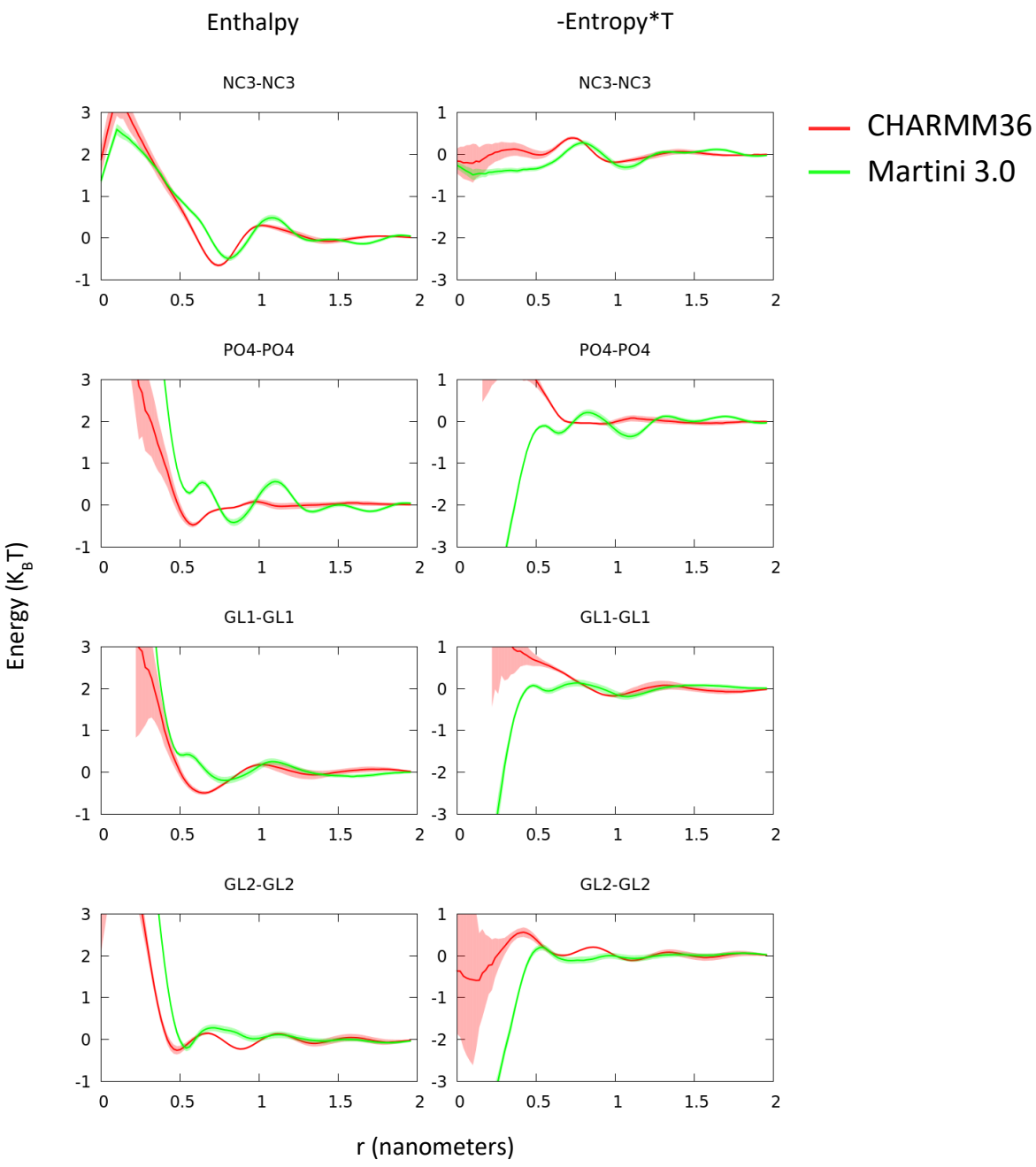

Figure 2.10: Entropy-enthalpy decompositions of DOPC-DPPC head group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Choline: NC3, Phosphate: PO4, Glycerols: GL1 and GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

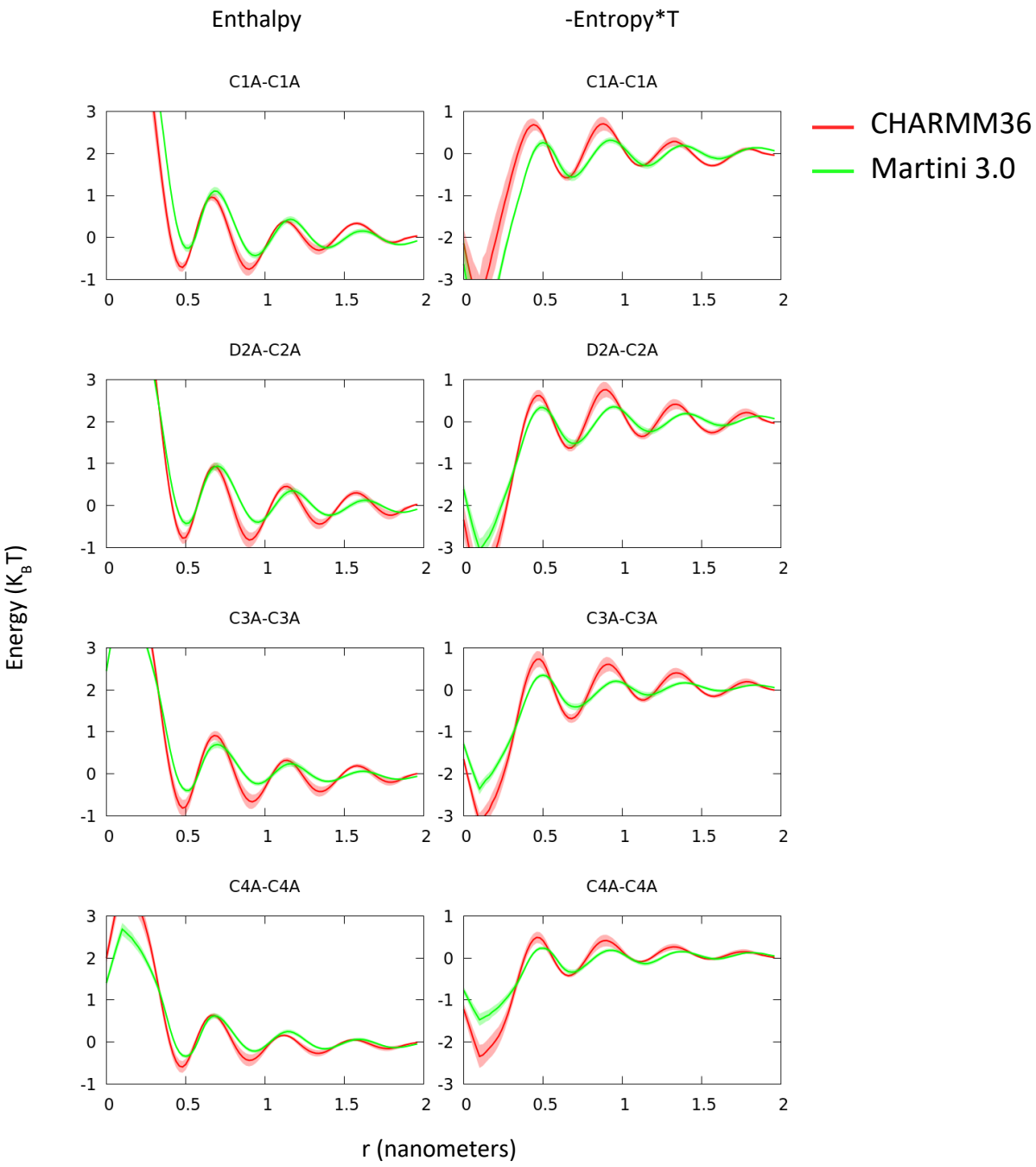

Figure 2.11: Entropy-enthalpy decompositions of DOPC-DPPC tail group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Unsaturated tail bead: D2A, saturated tail beads in descending order: C1A, C2A, C3A and C4A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

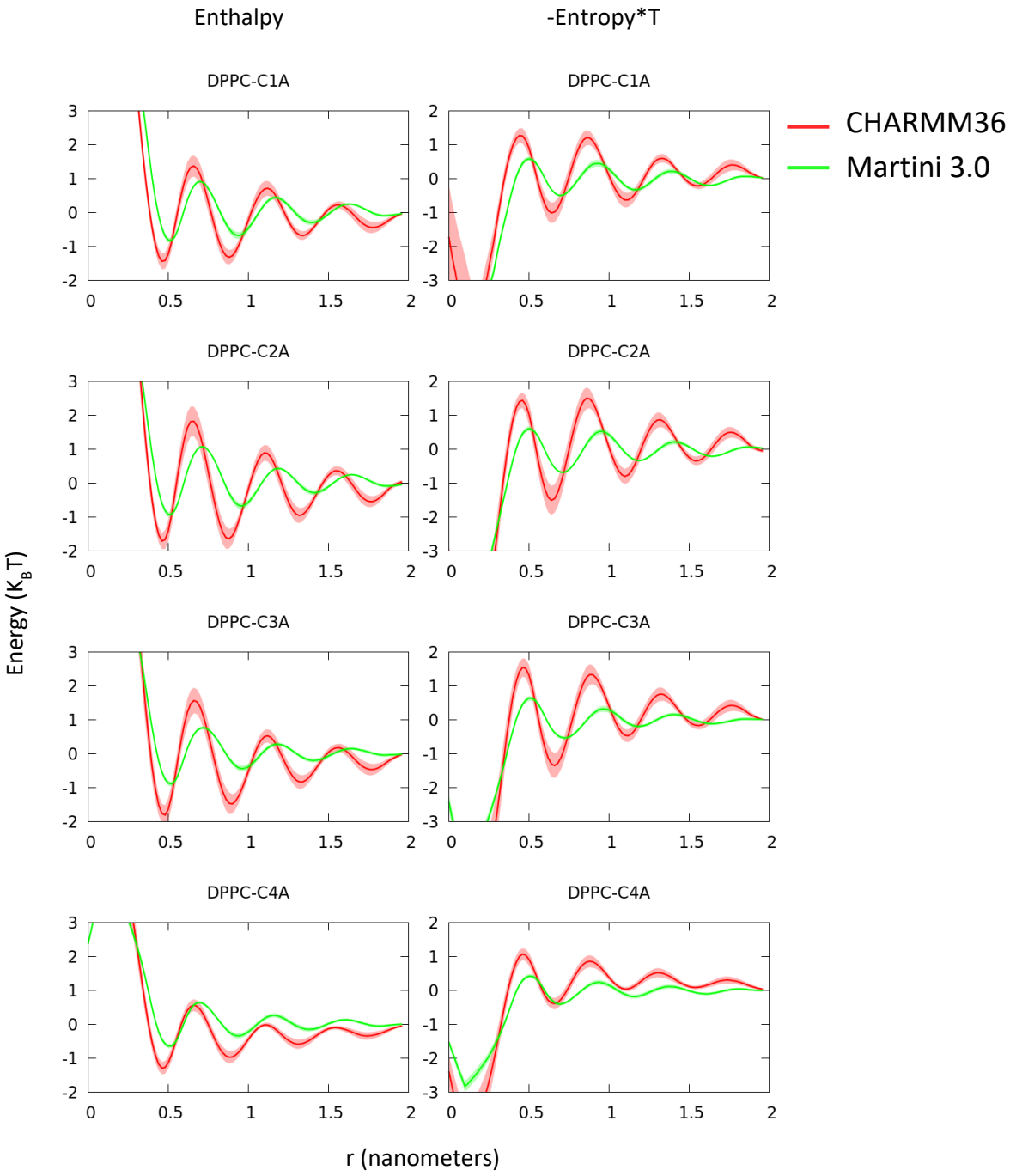

Figure 2.12: Entropy-enthalpy decompositions of DPPC-DPPC tail group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme. Tail beads in descending order: C1A, C2A, C3A and C4A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

Bending Modulus and Height Fluctuation Spectrum:

Table 2.1 details results of fluctuation spectrum based bending modulus calculation. Martini 3.0 consistently produces a bilayer which is less stiff than those seen in the atomistic simulations. Furthermore, the composition of the bilayer seems to have little effect on the Martini bending modulus, while in the atomistic case the DOPC/DOPS bilayer is almost 50 percent stiffer.

Table 2.1: Bending moduli k_c for CHARMM36 and Martini 3.0 lipid bilayers. Each bending modulus was calculated via the Fourier Transform of the membrane height fluctuation spectrum's low wavelength limit.

System	Martini 3.0 k_c (k_BT)	CHARMM36 3.0 k_c (k_BT)
70:30 DOPC/DOPS	28.9 ± 3.7	14.6 ± 1.6
50:50 DOPC/DPPC	26.8 ± 0.5	8.9 ± 1.6

In addition to the bending modulus, Martini fails to capture the height fluctuation spectrum of mapped atomistic bilayers at higher wavelengths. Figure 2.13 shows the full height fluctuation spectrum for the DOPC/DOPS system for both Martini and CHARMM36. Between $q = 1$ and $q = 5$ nm^{-1}, the Martini bilayer shows fluctuations which are up to twice as large as the reference simulation. These fluctuations mediate interactions between proteins on the surface of membranes[70] and thus affect the aggregation of membrane bound proteins. This has consequences for processes such as endocytosis, in which large quantities of proteins such as N-BARs which contain membrane targeting amphipathic helices.[71]

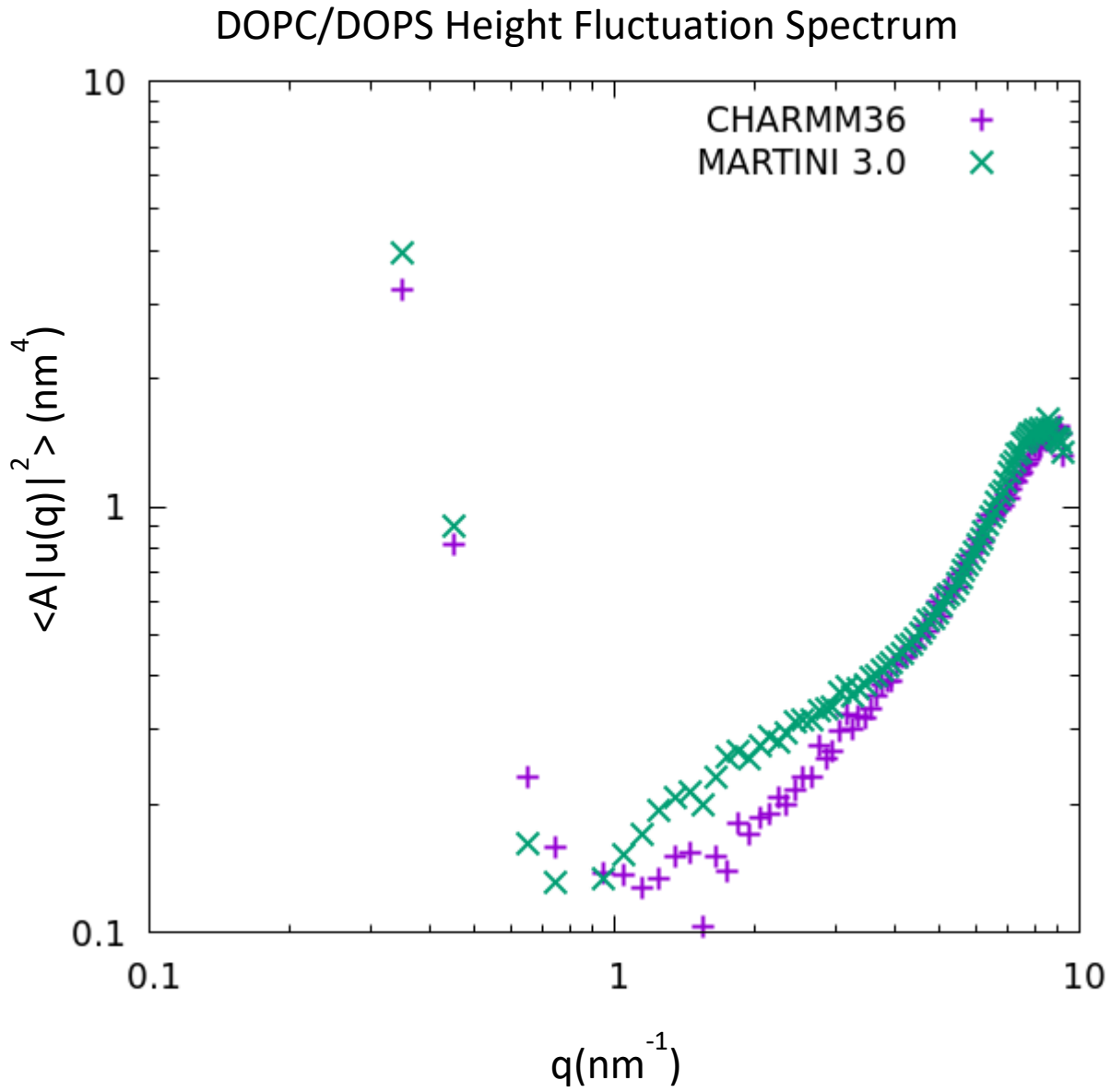

Figure 2.13. Height fluctuation spectrum for the DOPC/DOPS system. Bending modulus calculations are derived from the low q regime, in which fluctuations are directly related to bending rigidity for discrete systems such as lipid bilayers.

Figure 2.14 shows the same spectrum for the DOPC/DPPC system. In this case, the discrepancy between the atomistic and CG bilayers are even greater and persist to higher frequency modes.

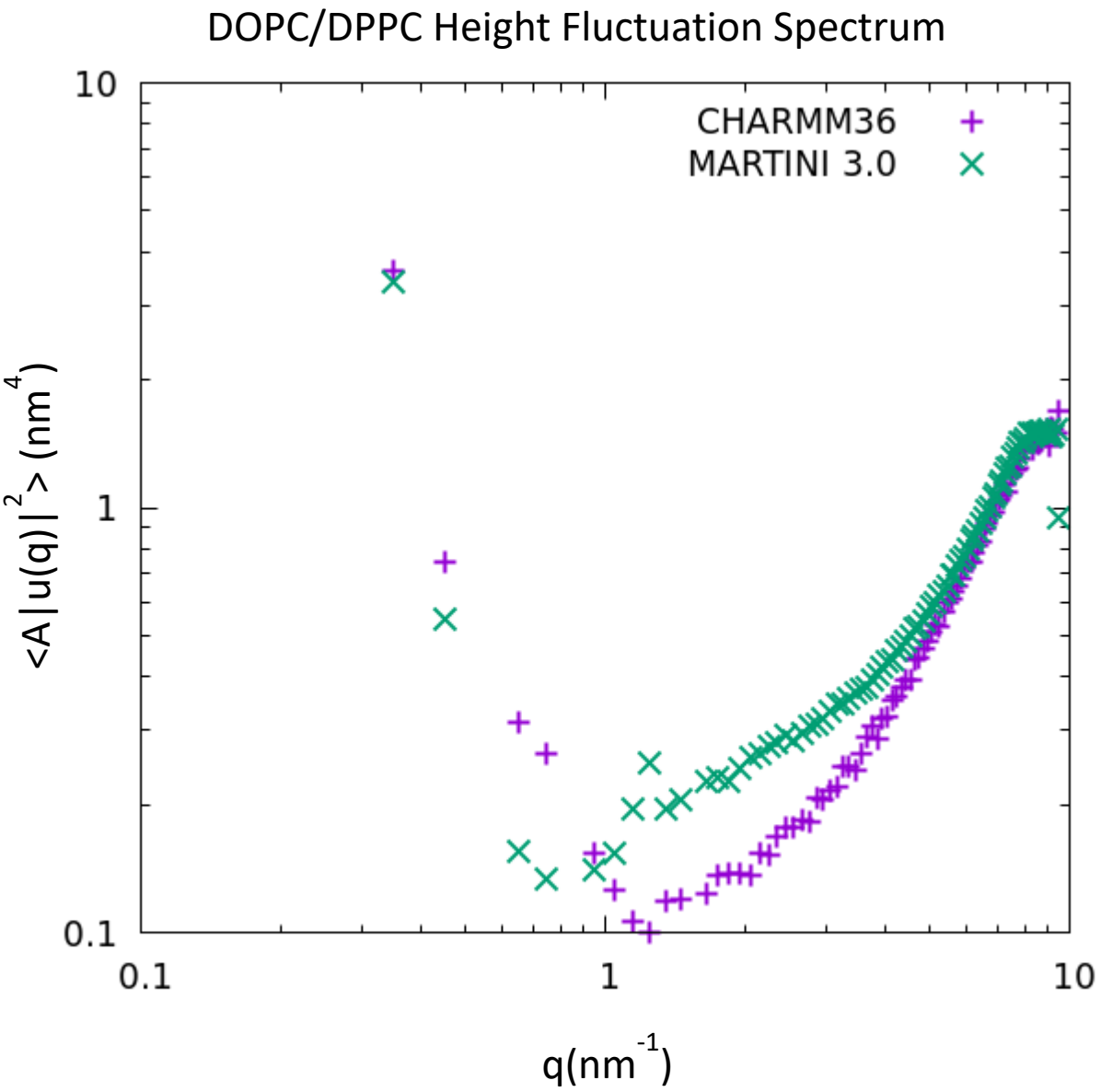

Figure 2.14. Height fluctuation spectrum for the DOPC/DPPC system. Bending modulus calculations are derived from the low q regime, in which fluctuations are directly related to bending rigidity for discrete systems such as lipid bilayers.

2.5 Conclusions

Several conclusions can be drawn about Martini 3.0 membranes in general based on the lateral associations of its tail beads when compared to AA data. For all systems, Martini consistently produced over-structured RDFs in comparison to CHARMM36. However when

looking at the entropy-enthalpy decompositions, Martini seems to consistently underestimate peak and well heights. What this means is that Martini is less sensitive to changes in temperature due to the smaller magnitude of its entropic term but compensates for this via a lower enthalpy term. This over-structuring at room temperature also likely contributes to its stiffer bending modulus.

The results presented in the previous section show that in the case of temperature transferability, Martini 3.0 continues to fail in several critical areas which are sensitive to bilayer composition. The two systems studied provide insight into how lipid bilayers partition entropy and enthalpy in different regions of the membrane. For the first system containing DOPC and DOPS, the heterogeneity from the different lipid species localizes in the head groups. The difference between the two molecules is the presence of a choline group in DOPC versus a serine group in DOPS. Structurally, Martini fails to capture peaks in the serine-serine RDF while consistently failing to reproduce the structural correlations of the phosphate head group. While the choline-choline decompositions are qualitatively similar for CHARMM36 and Martini in the DOPC/DPPC system, the presence of DOPS in this membrane causes Martini to produce peaks which are not seen at all in the CHARMM data. It can be concluded that the heterogeneity of the membrane itself, or Martini's inability to properly represent serine disrupts the otherwise reasonable choline-choline association.

On the other hand, the DOPC/DPPC system exhibits heterogeneity localized within the hydrophobic core of the bilayer. In this case, the lack of double bonds in DPPC creates a bilayer with more flexible tails which are capable of unkinked conformations. Despite this, Martini 3.0 DPPC appears to have a much larger effective size based on the locations of peaks in the DPPC-DPPC tail RDFs and entropy-enthalpy decompositions. In addition to this, AA DPPC is much more sensitive to temperature, noted by the increased magnitude of its entropy-enthalpy

decompositions in the tail groups. The entropic aspect of this is likely the root cause, as Martini tails cannot account for the conformational entropy of AA lipid tails. This lack of entropy then requires a lowered enthalpy to compensate and ensure that the PMF is similar enough to better capture the RDFs. While the previous systems issues seemed in part to stem from the heterogeneity of the lipid heads, in the case of the DOPC/DPPC membrane these issues stem mostly from the DPPC lipids themselves, and do not seem to affect the mixed lipid interactions as much.

Lastly, the bending modulus calculations show that Martini 3.0 consistently produces overly stiff membranes. These membranes are roughly as stiff as Martini 2.0 DOPC + cholesterol membranes which suggests that Martini bending stiffness is insensitive to membrane composition. While this produces reasonable results for membranes which happen to closely match this "standard" Martini stiffness, DOPC/DOPS and DOPC/DPPC membranes should be much more flexible due to their lack of cholesterol. Other analyses of Martini bilayers show similar results for bending moduli calculated via spectral analysis.[72] However the same systems bending moduli were more accurate when calculated via lipid splay angles, and future analyses may be improved if multiple methods are compared.

Overall, the additions to Martini 3.0 in the context of lipid bilayers do not seem to improve on many of the failures that Martini 2.0 lipids suffered from. Attempts to incorporate temperature transferability into the model have not improved Martini's partitioning of entropy and enthalpy into lateral association PMFs, and the new systems analyzed bring to light more issues related to Martini membranes material properties. It is unlikely that Martini can solve these problems without incorporating explicit temperature dependence into the CG force field.[73] These issues are inherently connected to the statistical mechanical nature of CG models, which cannot reproduce entropic contributions from sub-bead level fluctuations and thus do not behave as expected outside

of the specific thermodynamic state point they were parameterized for. The connection between the PMFs of lateral association and the CG potential, which is itself a conditioned PMF, suggest that a linear temperature dependence may be sufficient to recapture the lost entropy in the CG resolution and would likely lead to much more transferrable CG models.

2.6 Additional Figures

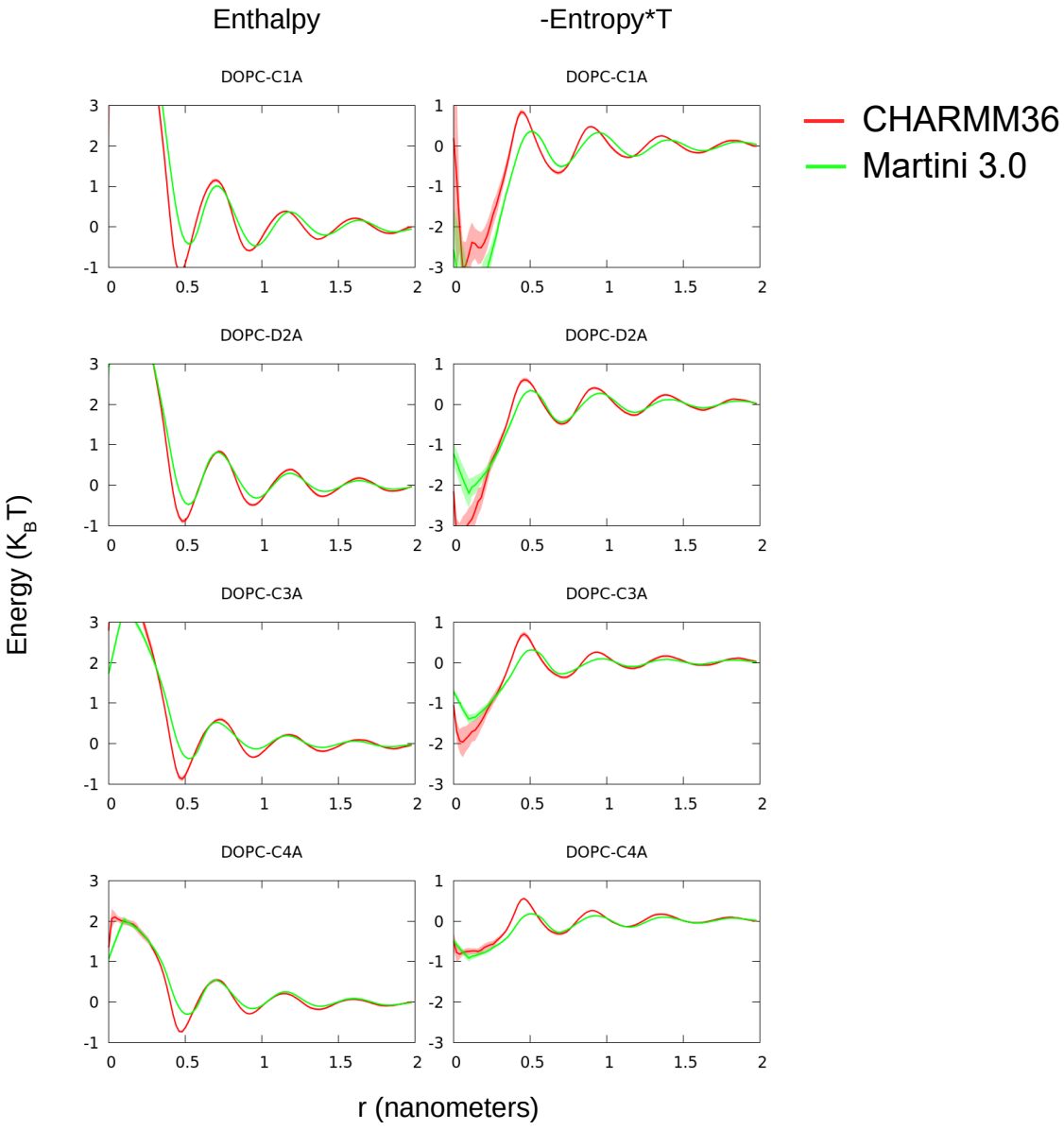

Figure 2.S1: Entropy-enthalpy decompositions of DOPC-DOPC tail group interactions for DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Saturated tail beads in descending order: C1A, C3A, and C4A. Unsaturated tail bead: D2A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

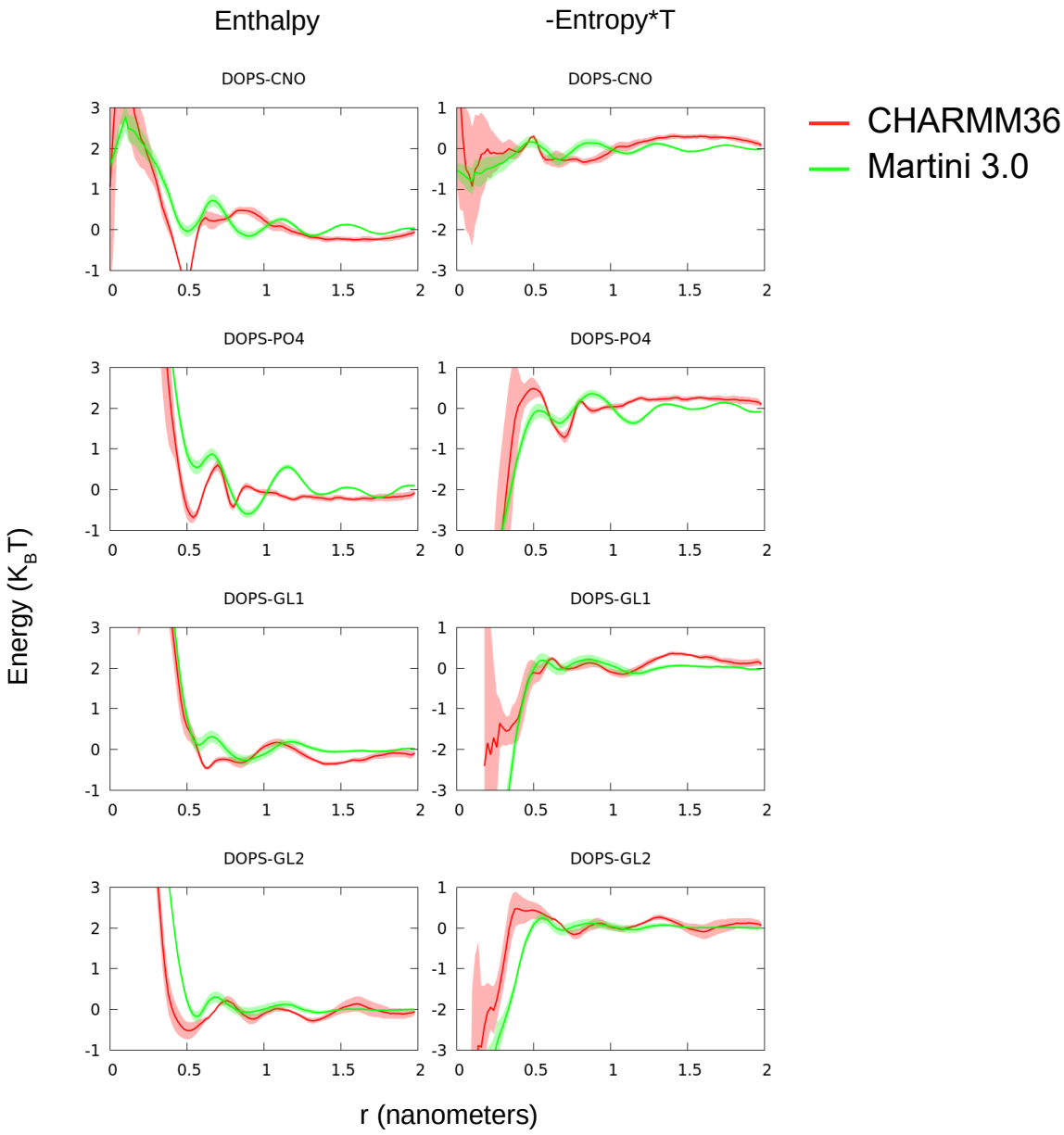

Figure 2.S2: Entropy-enthalpy decompositions of DOPS-DOPS head group interactions for DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Serine: CNO, Phosphate: PO4, Glycerols: GL1, GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

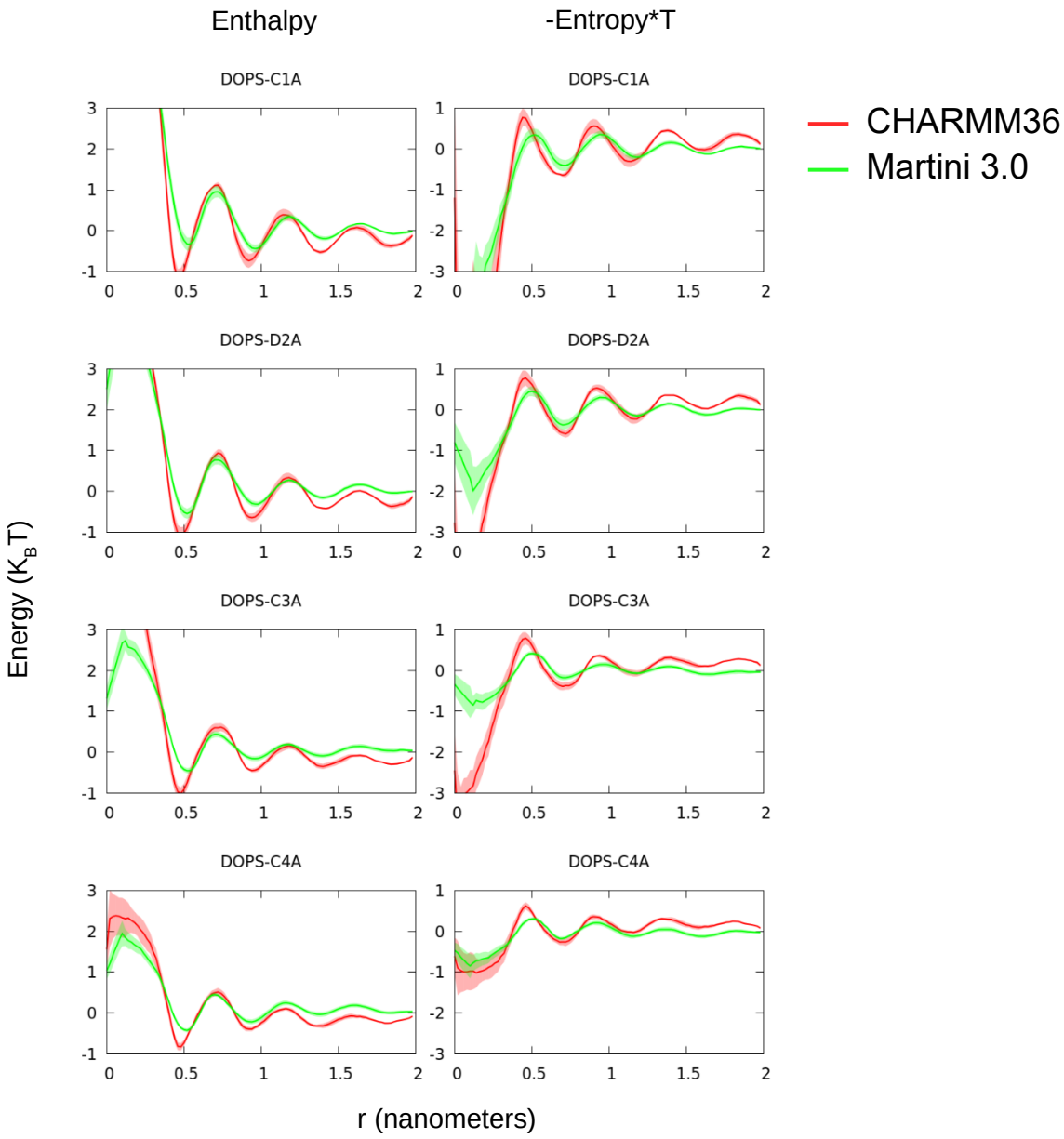

Figure 2.S3: Entropy-enthalpy decompositions of DOPS-DOPS tail group interactions for DOPC/DOPS system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Saturated tail beads in descending order: C1A, C3A, and C4A. Unsaturated tail bead: D2A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

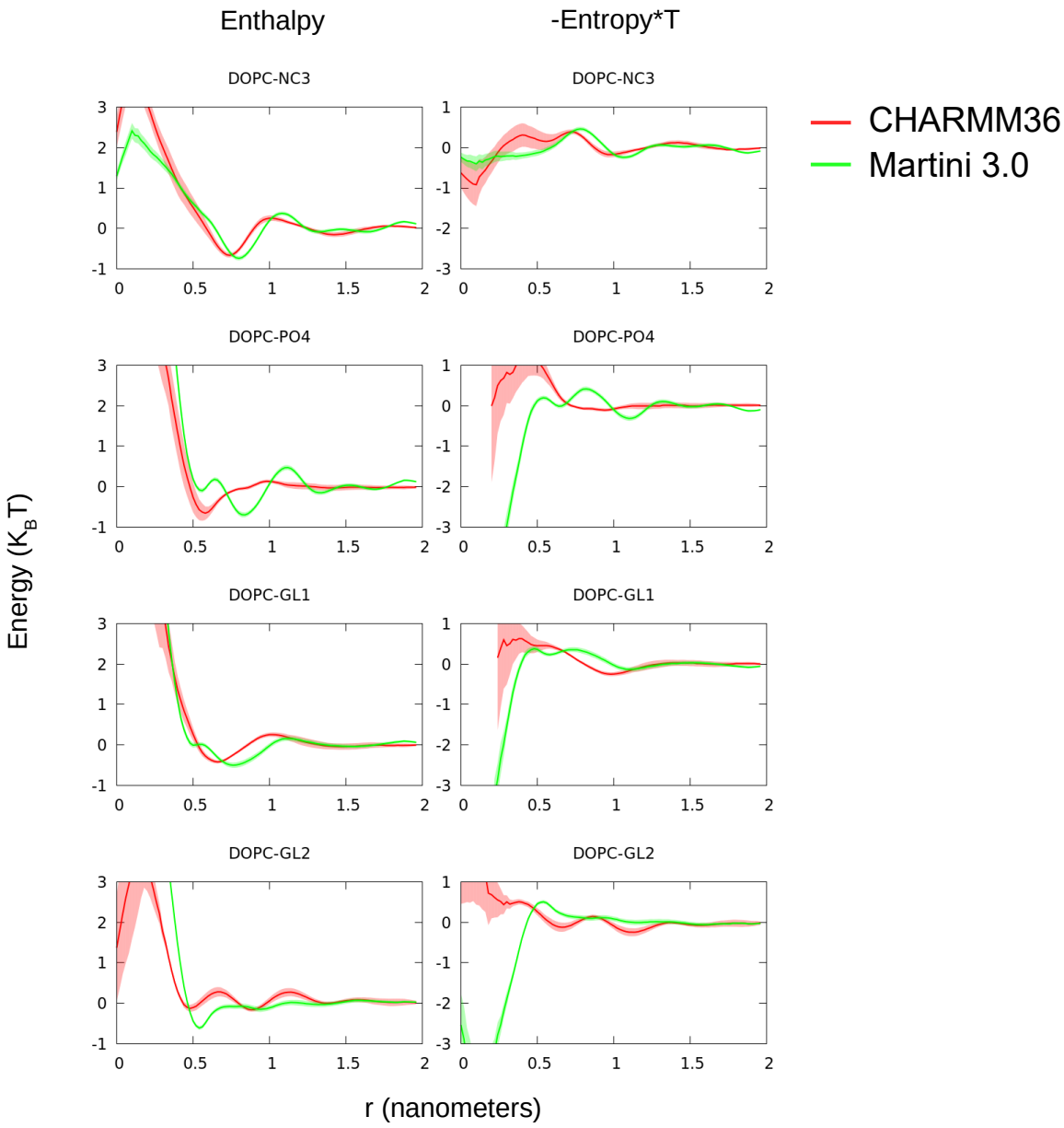

Figure 2.S4: Entropy-enthalpy decompositions of DOPC-DOPC head group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Choline: NC3, Phosphate: PO4, Glycerols: GL1, GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

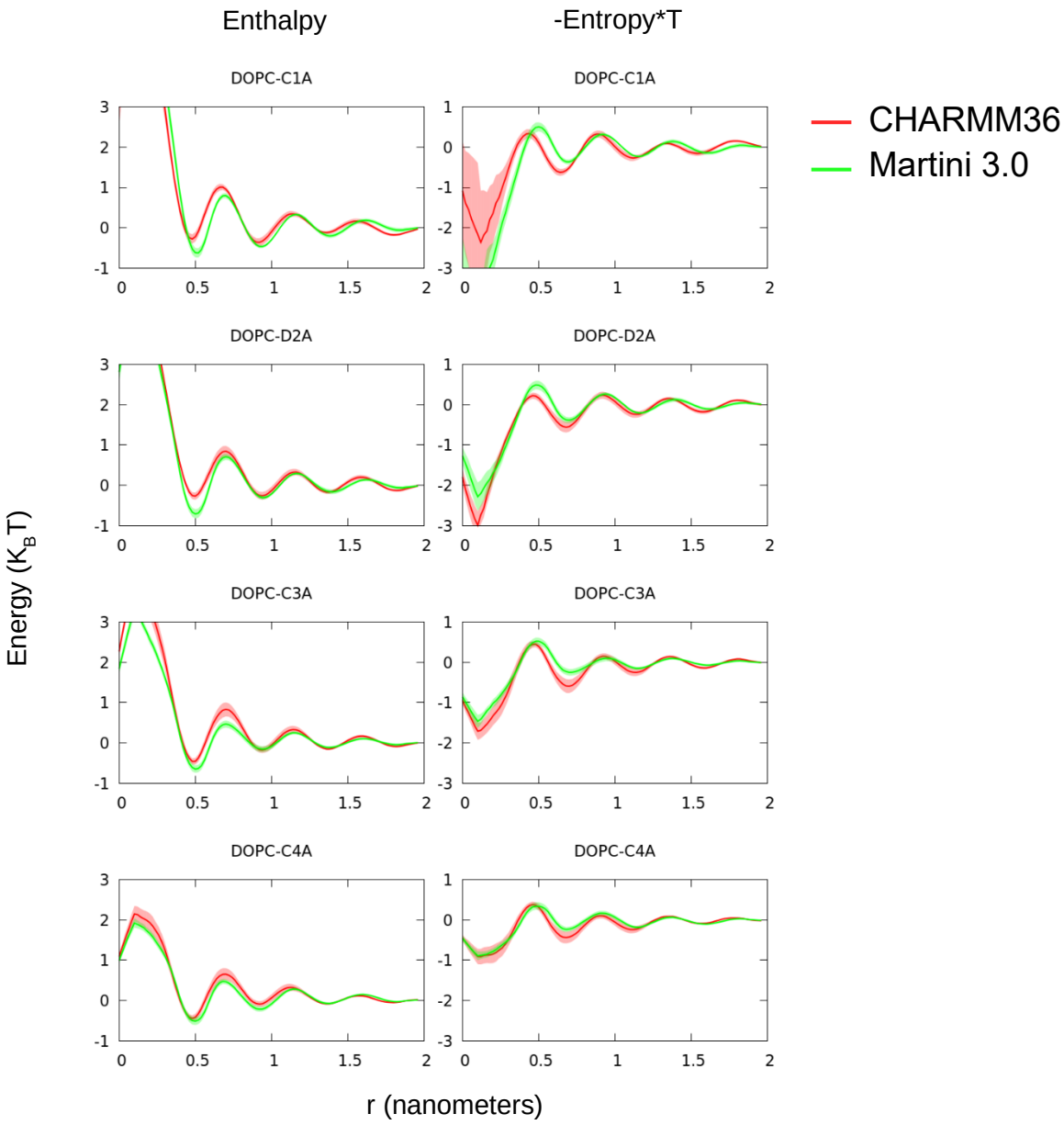

Figure 2.S5: Entropy-enthalpy decompositions of DOPC-DOPC tail group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Saturated tail beads in descending order: C1A, C3A, and C4A. Unsaturated tail bead: D2A. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

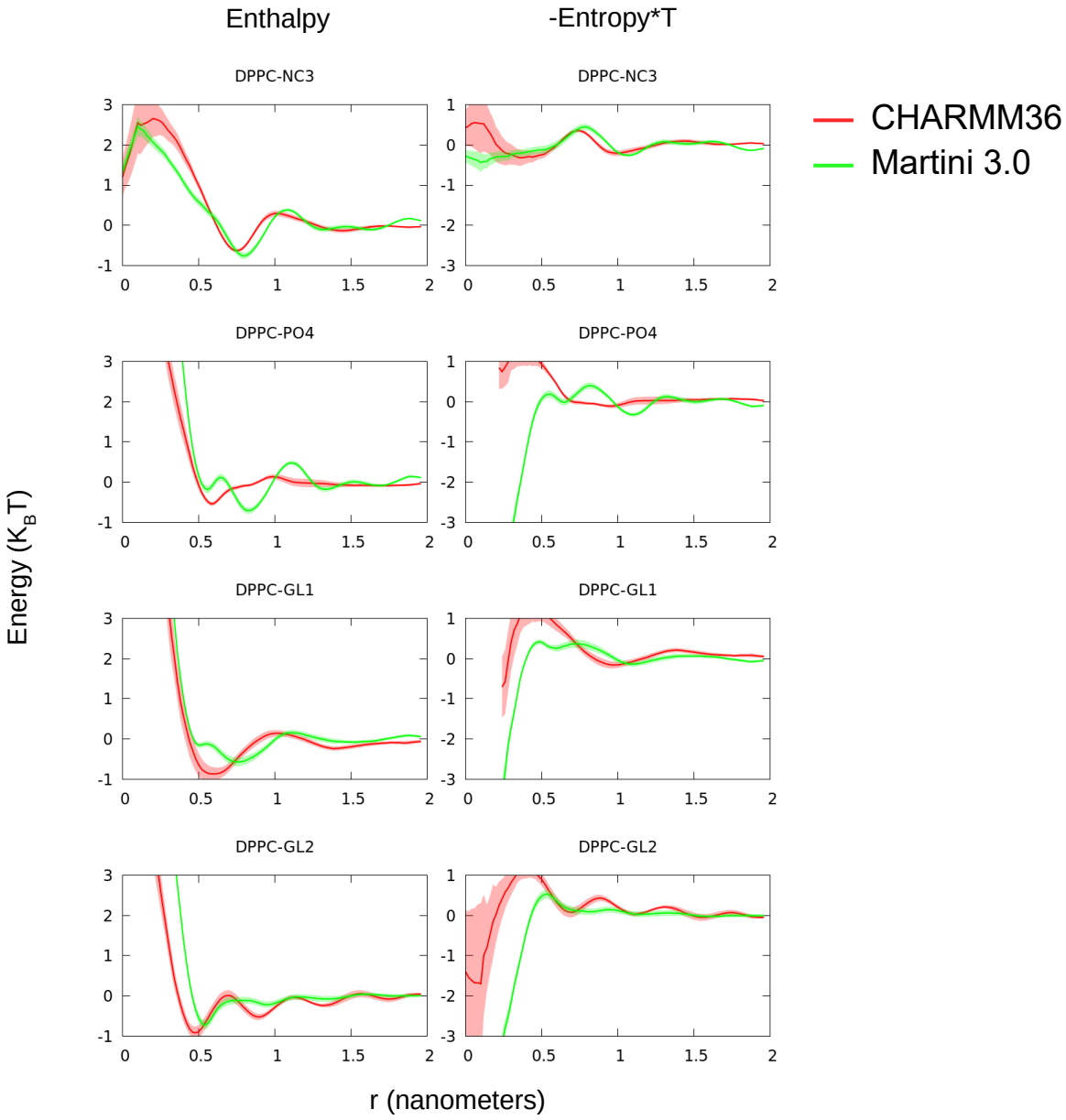

Figure 2.S6: Entropy-enthalpy decompositions of DPPC-DPPC head group interactions for DOPC/DPPC system. Standard errors at each r value are shown. Beads are labelled according to the following scheme: Choline: NC3, Phosphate: PO4, Glycerols: GL1, GL2. Martini 3.0 decompositions are compared to mapped CHARMM36 data.

Chapter 3: High Resolution Coarse-Grained Lipid Models can be Systematically Improved via the Introduction of Solvent Virtual Particles as well as Explicit Temperature Dependence

Abstract

The development of bottom-up coarse-grained (CG) lipid models at a resolution of 4 heavy atoms to 1 CG site has unique challenges which make them often paradoxically less accurate than CG models at lower resolutions. These problems stem from the inherent anisotropy in lipid systems, and the fact that the sampling of lipid head-tail interactions is essentially nonexistent for this reason. At lower resolutions, these interactions tend to be better sampled due to the difference in mapping schemes. When applying force matching to such a system, the resulting bilayer is unstable, either breaking apart, or interdigitating until each leaflet experiences significant overlap. Fixing these issues requires the *ad-hoc* alteration of CG interactions, effectively making it a top-down approach. However, these issues can be partially overcome via advanced coarse-graining techniques. The first approach to solving these issues is to introduce virtual solvent particles, which repartition the forces to better recapture the hydrophobic effect. This, combined with Relative Entropy Minimization produces a much more stable model, which when combined with temperature dependent potential parameterization, reproduces entropy enthalpy decompositions of the reference data much better. However even with all these techniques, such models are incapable of self-assembly, a feature which is possible at lower resolutions. I present a CG model of 1,2-

dimyristoyl-sn-glycero-3-phosphocholine (DMPC) using a high-resolution mapping built with the multiscale coarse-graining method and describe how at this resolution, this methodology fails to produce an accurate model. I then describe steps which can systematically improve the model, including the application of relative entropy minimization, virtual solvent particles, and explicit temperature dependence and demonstrate that even with all of these techniques combined, the model still fails to meet even the quality of lower resolution lipid models, which suffer less from sampling issues.

3.1 Introduction

Molecular dynamics (MD) simulations provide a means to probe chemical systems at atomistic resolutions and timesteps on the order of femtoseconds.[74] This ability makes MD valuable for the study of biomolecular systems at the microscopic scale. While X-ray crystallography can provide experimental insight into the structures of such systems at similar length scales,[75] it can only show scientists a static snapshot of what is happening within the cell. Worse still, crystallographic methods suffer from the act of crystallizing the sample, a process which can push molecules, especially proteins, outside of their native conformations.[76] Cryo-electron microscopy, which avoids crystallization of the sample, can avoid some of these drawbacks,[77] it still shows a static picture, and thus cannot offer any dynamical information about biochemical processes on its own.

Despite this, all-atom (AA) MD suffers from the limitations of current computer hardware. Many biophysical processes take place on the order of seconds or longer or involve a quantity of atoms which is simply inaccessible to AA-MD, which is practically limited to microseconds, and billions of atoms. While these boundaries are pushed and even broken frequently,[78] access to rare,

specialized hardware limits widespread achievement of such simulations. Bottom-up Coarse-graining (CG) methods allow access to much larger time and length scales via a reduction in the number of particles used to represent the system of interest. These particles, or CG beads, interact via a CG Hamiltonian which is parameterized using a smaller amount of AA MD data, and can be used to extend the size of MD simulations for a fraction of the computational cost. Multiscale coarse-graining (MSCG), also known as force matching, is a method to match bottom-up CG models by minimizing the residual of the CG forces between the model force field and the forces obtained via mapping reference AA data down to the desired CG resolution.[66,79,80] Equation 3.1 describes this force minimization scheme.

$$\chi^2[F(\phi)] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |f_I(M(r) - F_I(M(r)|\phi)|^2 \right\rangle \tag{3.1}$$

In this case I is one of N total CG sites and r is the set of atomistic coordinates which map to a CG site via the mapping operator M. f_I and F_I are forces obtained from summing the atomistic forces on that CG site and from the CG force field with parameters ϕ respectively. For most biophysical applications, the model parameters being optimized are b-spline knots, which results in a very expressive, albeit pairwise basis set.[79] Three-body forces can be learned via MSCG by expressing these three body interactions in terms of Stillinger-Weber potentials,[81,82] although this comes at the expense of integration speed. It is often difficult to justify including third order correlations as speed is the key feature of a CG model when compared to corresponding AA simulation.

Another method to produce bottom-up pairwise CG models is relative entropy minimization (REM). Rather than optimize the forces of the CG model to match that of the reference data, REM seeks to optimize the overall static structural properties of the model. In this case, each model is represented by a probability distribution in terms of the spatial coordinates of

each atom. These distributions can be compared directly with each other via the Kullbeck-Leibler divergence (KL divergence) also known as the relative entropy:[83]

$$S_{rel} = \sum_{r} P_{ref}(r) \ln\left(\frac{P_{ref}(r)}{P_{mod}\big(M(r)\big)}\right) \tag{3.2}$$

Where $P_{ref}(r)$ is the probability of observing a certain configuration *r* in a reference distribution, and $P_{mod}\big(M(r)\big)$ is the probability of observing the same configuration (mapped to the CG resolution) in a distribution from a model force field. The relative entropy grows larger as the reference and model distributions become more dissimilar. The gradient of this quantity can also be calculated in terms of model parameters via

$$\nabla_{\phi} S_{rel} = \beta \left\langle \frac{\partial U_{CG}}{\partial \phi} \right\rangle_{FP} - \beta \left\langle \frac{\partial U_{CG}}{\partial \phi} \right\rangle_{CG} \tag{3.3}$$

Where U_{CG} is the CG potential, FP and CG refer the reference atomistic and CG ensembles respectively and ϕ is a model parameter. Then, via gradient descent, the model parameters can be updated to produce a model which shows better agreement with the reference data than before.[84] Such a method is thus iterative, and requires an initial guess for the force field of the system. Given that most common CG models are built on pairwise basis sets, REM has the property of matching radial distribution functions (RDF), also referred to as g(r) extremely well. This is because in the case of a pairwise basis set, the configurational probability distribution is cast in terms of radial distribution functions. Interestingly, under the conditions of limitless sampling and a basis set which fully represents the many-body CG potential of mean force (PMF), both REM and MSCG will produce identical results.[84] However, this is never a realistic case, as sampling will always be inherently limited in the case of CG model production, and it is rare that one would ever seek to produce a CG model that extends beyond three-body correlations.

Naturally, such methods have been applied to a variety of systems, from simple liquids all the way to complex multicomponent biological systems. One class of system that is of great interest to CG modelers is lipid bilayers. Lipid membranes are good candidates for coarse graining for several reasons. First, bulk water, which can be integrated out in a CG model, makes up the majority of atoms in a lipid simulation. Second, many interesting processes in lipid membranes, such as the formation of lipid rafts, occur on long timescales, gated by the diffusion speed of lipids within the larger bilayer.[85] Thus coarse-graining accelerates the sampling of these processes on multiple levels.

The history of MSCG lipid membranes at high resolutions (4 heavy atom to 1 CG site or greater) is fraught with complications. Past implementations of such models are characterized by a large number of CG site types and thus a large and complicated Hamiltonian, and overall noisy potentials.[86,87] Worse still, when removing solvent from the system, a process which is commonly desired for CG lipids as it accounts for the majority of atoms in a lipid simulation while remaining the least interesting part of the system, the resulting force field is unusable without *ad-hoc* alterations.[88] The most critical issue is the fact that MSCG produces interactions between tail brands which lead to highly nonphysical behavior. In most cases, the raw output of MSCG produces a bilayer that at best interdigitates to the point that the membrane is no longer a bilayer, and at worst fails to stay together at all when starting from an assembled bilayer. This series of issues stems from a variety of places, some of which are inherent to all CG models, some to only lipid systems, some to implicit solvent lipids, and some specifically to high resolution implicit solvent lipids.

First, there are two major issues which are inherent to all classes of CG models: transferability and representability. As discussed previously, transferability issues arise when a CG

model is incapable of correctly matching the behavior of the system outside the thermodynamic state point for which it was parameterized. Even advanced, well parameterized models fail to properly represent the entropy of the AA system and thus cannot be transferred to other temperatures. Representability issues are also inherent, as a variety of observables are dependent on the full atomistic representation of the system, and calculation of pressure etc. becomes impossible.

When considering specifically CG lipid bilayers, one of the biggest issues stems from the fact that membranes have an inherently large amount of anisotropy. Due to the structure of the bilayer, each interaction is only sampled within a relatively flat slice of the simulation box, with interactions between the bulk solvent (if present) take up most of the observed sampling. There is also relatively little sampling between lipid heads and tails in a stable bilayer because of this structuring, which in turn leads to noisy potentials that can produce bad forces which can ultimately destabilize the resulting model. While lipid flip flopping can allow for lipid heads to interact with lipid tails as they cross from one leaflet to another, the process is rare and in pure simulated bilayers is only observed via enhanced sampling methods such as transition path sampling.[89,90] This is a difficult problem to address for CG lipids. Including sampling from non-bilayer configurations or enhanced sampling simulations could indeed improve the amount of sampling, however this would involve sampling the system out of equilibrium, which would lead to non-Boltzmann statistics and removes CG consistency between the reference and model data.[66] Sampling can also be improved by simply including more reference data, but this still fails to produce a good model with an amount of sampling which would be considered reasonable. Even 250 ns of reference data on a 1200 lipid bilayer is insufficient.

Second are issues inherent to all solvent free lipid models. While solvent is generally less interesting to study for lipid systems, it is nevertheless critical to membrane formation and stability. It is the hydrophobic effect after all which drives lipid assembly, not attractive interactions between lipid tails. Practically, this means that for force matching models of lipid bilayers without explicit solvent representation, the hydrophobic forces from bulk water are naturally pushed onto the lipid tails. The resulting model will typically have a highly attractive potential between lipid tail beads, especially those beads at the ends of the lipid tails.

Last are the issues that specifically affect these types of models at higher resolutions. High resolution CG models create an additional issue through the complexity of their mapping. While lower resolution models may incorporate a total of 4 CG types[91], models at the higher resolution necessitate 6 or more, shown in Figure 3.1. While this does not appear at first glance to be a large increase, it ends up creating a far more complex model. While a 4-type model will require 10 individual non bonded interaction types, there is a geometric expansion of interaction types with respect to the number of types. A 6-type model necessitates 21 separate interactions, and each parameter of each interaction (which in the case of spline-based models can be hundreds per

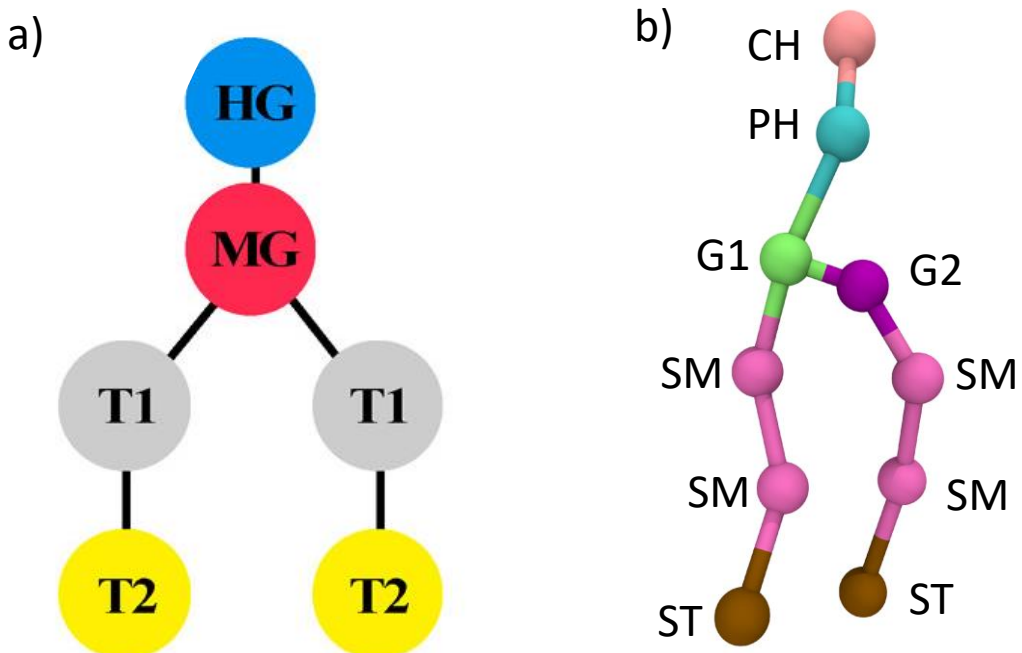

Figure 3.1: a) "Low" resolution mapping of a lipid molecule onto 6 CG sites with 4 types. This leads to a total of 10 nonbonded interaction types. Adapted from [91]. b) "High" resolution CG lipid mapped onto 10 sites with 6 types. This leads to more than double the number of nonbonded interactions.

interaction) is a potential point of failure. This has the effect of exacerbating sampling issues caused by the previous two issues, whereby tail beads which would normally be nominally closer to the head groups because of the nature of the mapping scheme are pushed farther away, and thus are even more poorly sampled.

While many of these problems are difficult to solve without resorting to applying unreasonable amounts of reference data to the problem, some of them can be approached in more strategic ways which do not obviate the need for a CG model in the first place. The area which has seen a large amount of effort dedicated to it is the representation of solvent. As mentioned, membranes form explicitly because of interactions between hydrophobic lipid tails and water, and

thus it is difficult to build a successful model without at least some form of solvent representation. Plenty of models have been developed for which water is explicitly represented, both top-down and bottom-up[17,86,87] yet these models suffer greatly due to the increased computational cost, or worse yet, inaccurate water models. In the case of bottom-up coarse graining, each water molecule must be represented as an individual bead, which hardly eliminates degrees of freedom from the system when compared to mappings of even 4 heavy atoms per bead. In the case of top-down models such as Martini, each water bead represents multiple water molecules, which solves the problem of efficiency (at least partly, fully implicit solvent models will be even faster) but runs into problems due to the unphysical water beads freezing at or near room temperature.[52,53]

There is a middle ground between full solvent representation and implicit solvent which has been applied to a great degree of success, at least in lower resolution CG lipids: virtual solvent particles.[91] For the purpose of this work a virtual particle is defined as a CG site which has no linear mapping to atoms in the reference trajectory. In the case of virtual solvent particles, they are solvent beads which represent the water molecules that are tightly associated with the lipid head groups, which are then bound to the CG lipid heads in an effort to represent only the solvent which is exerting the most influence on the bilayer. In the reference trajectory, this means that the virtual solvent particles may map to different individual water molecules as they are free to leave the bilayer surface or move to different areas, and adjacent virtual particles may contain the same water molecule within them. Predictably, parameterizing virtual solvent particles via MSCG is difficult, as the mapping of forces must be linear for MSCG models to remain consistent with their underlying AA data.[66] Instead, in order to obtain virtual particle forces, a REM model of the system must first be obtained, which can be used to estimate the forces on the virtual particles in the mapped reference data. After this step, those forces can be used as a stand in for actual reference

forces, and force matching can then be performed. Additionally, the resulting MSCG model can then be used as input for further optimization via REM.[91] This method was used to generate a 6 (plus 1 virtual) site 1,2-dioleoyl-sn-glycero-3-phosphocholine DOPC lipid model which exhibited not only bilayer stability, but also the capacity to self-assemble into a variety of morphologies depending on the initial concentration, which is not seen in implicit solvent bottom-up lipids.[91] The introduction of the virtual particles also improved the bending modulus of the membrane. Without the virtual particles, both MSCG and REM produced bilayers which were much stiffer than corresponding AA membranes, while introduction of the virtual particles in combination with REM and MSCG together resulted in a bilayer which is nearly as flexible as the reference data.

In this work I investigate the steps necessary to build more accurate high resolution (4 heavy atoms per bead) CG models of a 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid bilayer. Five models for CG DMPC are presented. The first is an implicit solvent MSCG model parameterized using a much larger amount of sampling than has been applied to such high-resolution models in the past, which is referred to as IS-MSCG. The second is a model built using the first model as input for a long REM optimization, which shows improved stability without *ad-hoc* force field alteration, referred to as IS-REM. Next, I apply a virtual solvent site to each lipid molecule and demonstrate that this improves overall bilayer stability, albeit at short timescales, referred to as VS-REM. I then apply MSCG using virtual site forces from the previous model and non-virtual forces from the reference data, which behaves largely the same as the initial MSCG model without virtual particles, referred to as VS-MSCG. Finally, I present a REM virtual solvent model with explicit temperature dependence, which improves temperature transferability, but is still unable to avoid certain structural issues, referred to as TT-VS-REM. The work is laid out as follows. First the methodology behind each step in the model creation is discussed, including

atomistic reference simulation details, hyperparameters for MSCG, REM and the inclusion of virtual sites. Next, data showcasing the advantages and disadvantages of each model are presented, with a short section describing the total amount of work required to generate each model. Lastly, the data is discussed in detail, and conclusions about the viability of models at this resolution are made.

3.2 Methods

The initial reference dataset used to generate the DMPC models was generated by simulating a system of 1260 DMPC lipids using the CHARMM36 force field[92] with TIP3P water[93] to generate accurate lipid statistics using GROMACS.[94] Inputs were generated using the CHARMM-GUI membrane builder[68,69,95] with the system containing 22.5 Å of bulk water above and below the membrane to ensure minimal correlation across periodic images. Sodium and chloride ions were added to the bulk solvent to bring it to 0.15 M according to physiological conditions.[96] The system was first equilibrated for 50 nanoseconds at NPT conditions, 1 atmosphere and 300 K, followed by a 250 ns NVT production run at 300 K. Forces were dumped every 10 ps for a total of 25,000 frames of force data. This trajectory was split into 25 equal subtrajectories of 1000 frames each to facilitate parallelized force matching. While previous attempts to force match high resolution DMPC used far less sampling, as little as 64 lipids simulated for 400 ps,[87] the larger dataset was chosen to ensure that any issues related to noisy potentials and instability were not related to insufficient reference data.

The system was mapped to 10 CG beads per lipid. The phosphate and choline were each assigned their own bead types, named PH and CH respectively. Each glycerol was assigned a separate site type as well (G1 and G2) to account for the odd number of carbons incorporated into

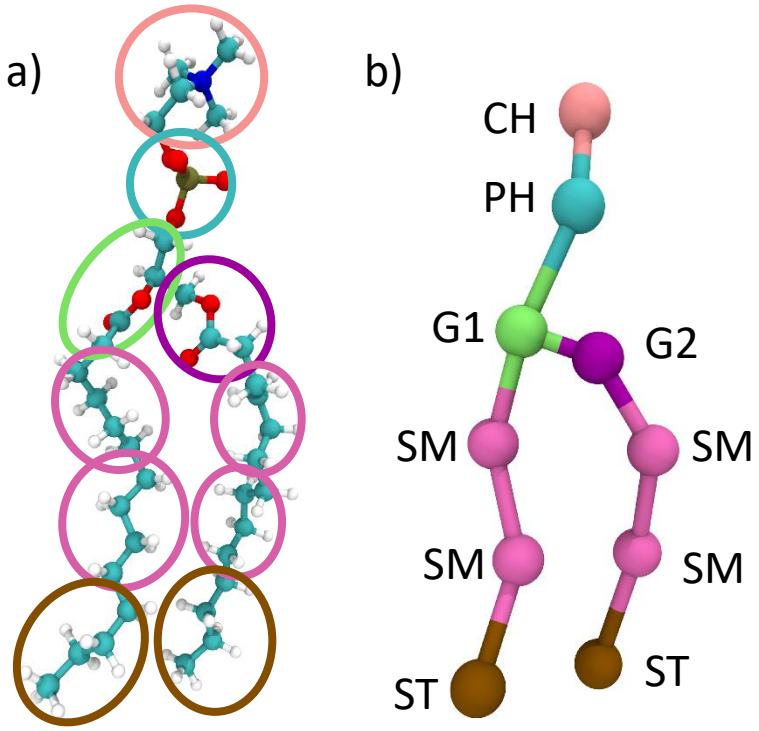

Figure 3.2: a) Atomistic DMPC molecule grouped to approximately 4 heavy atoms per group. Each group is colored so that only groups with the same chemical structure are the same color. b) CG DMPC molecule mapped according to groupings in a). Each bead's position is set to the center of mass of the grouped atoms, and the mass is set to the sum of the group's atoms.

each glycerol bead. The tails were mapped onto three sites each. The first two tails from the glycerol were assigned one site type (SM), while the terminal bead was assigned a separate type (ST). This choice was made to prevent attractive forces holding the leaflets together from affecting the mid-tail region of the system. Figure 2 details which atoms were assigned to which site. Each bead's position was mapped to the center of mass of all atoms in the bead, while forces were mapped additively according to center of mass mapping rules required to generate consistency with the reference atomistic data. Water, as well as all sodium and chloride, was excluded from the mapping to produce a lipid only mapped atomistic trajectory.

For the IS-MSCG model of DMPC, all 21 possible pairwise interactions were optimized using 3rd order B-splines with control points placed every 0.1 Å. All bonds and angles were

optimized as well using 3rd order B-splines, with control points placed every 0.25 Å for bonds and every 5 degrees for angles. Nonbonded interactions were subjected to a long-range cutoff of 25 Å. Inner cutoffs for all interactions, as well as outer cutoffs from bonded and angular interactions were determined by the following procedure: One of the 1000 frame mapped subtrajectories was analyzed to produce histograms of each interaction by distance with a bin width corresponding to the spline resolution of each interaction type. The lowest bin (or highest bin in the case of bonded and angular outer cutoffs) which contained at least 100 examples was selected as the cutoff. As each interaction can contain potentially hundreds of parameters, Bayesian regularization was applied to the outputs of force matching for 250 iterations to ensure smooth potentials.[97,98] Forces from 1-2 and 1-3 bonded interactions were excluded from force matching of non-bonded interactions.

The second DMPC model created, IS-REM, used REM to optimize the force field generated by the IS-MSCG model. As an iterative model, REM benefits from an initial guess force field which reproduces the reference configurational distribution as closely as possible. Since the force field generated by MSCG caused the leaflets of the bilayer to break into micelles, several interactions were altered to provide REM with a better starting point. The bilayer collapse was resolved by replacing the terminal ST beads on each lipid tail with SM beads, leading to a model which at least formed a bilayer with properly separated leaflets. When REM was performed however, the ST beads were reintroduced for the purpose of model optimization. These beads were considered two different types which started with the same potentials. Failure to do so resulted in a model which fell apart over the equilibration period after a few iterations. Inner cutoffs for each non-bonded interaction were also reselected to prevent the relative entropy between model and reference data from diverging near the cutoff. If the inner cutoff is too small, the model can sample

configurations which are never seen by the reference model, and thus the ln (P_{ref}/P_{mod}) term in the relative entropy becomes undefined. This numerically causes the potentials generated by REM near the inner cutoff to become increasingly negative, which stabilizes such configurations even further, forming a feedback loop. To avoid this, REM was performed for a short number of iterations until such negative spikes appeared in the potentials. Then the inner cutoffs were then moved out until there was only one spline knot left in the unsampled region (essentially the beginning of the hard wall region). This knot was held fixed, while only knots from the sampled region were allowed to update. By selecting the cutoffs this way, this problem can be avoided for the actual REM iterations. The spline order for the interactions was increased to 4th order splines to account for the fact that REM natively outputs potential splines as opposed to MSCG which outputs splines of forces, the negative derivatives of the potentials. The last change made to the MSCG model was to change the spline knot resolution for the non-bonded interactions from 0.1 to 0.5 Å. This was done to avoid oscillations in the potential which arise due to the fact that each iteration of REM uses less sampling than MSCG.

For each iteration of REM, reference data collected from 50 frames of reference data was compared to 50 frames of data from the previous iteration's CG model. Each non-bonded interaction was allowed to update at each iteration, with a maximum step size for each parameter of 0.002 k_BT. Low step sizes were chosen to ensure stability across iterations, as larger ones tended to oscillate between stable and unstable bilayers without improving the model. This step was run for 1000 iterations to produce the final model.

The VS-REM model was generated using the IS-REM model as a starting point. The virtual solvent particles were mapped by a different strategy than the non-virtual sites. In this case, the mapped reference trajectories (with water included) were mapped to the CG resolution described

above. Then the RDF of water about the phosphate bead was calculated in order to estimate the size of the first solvation shell. Figure 3.3 shows this RDF, where the first solvation shell was found to extend to approximately 7 Å away from the phosphates.

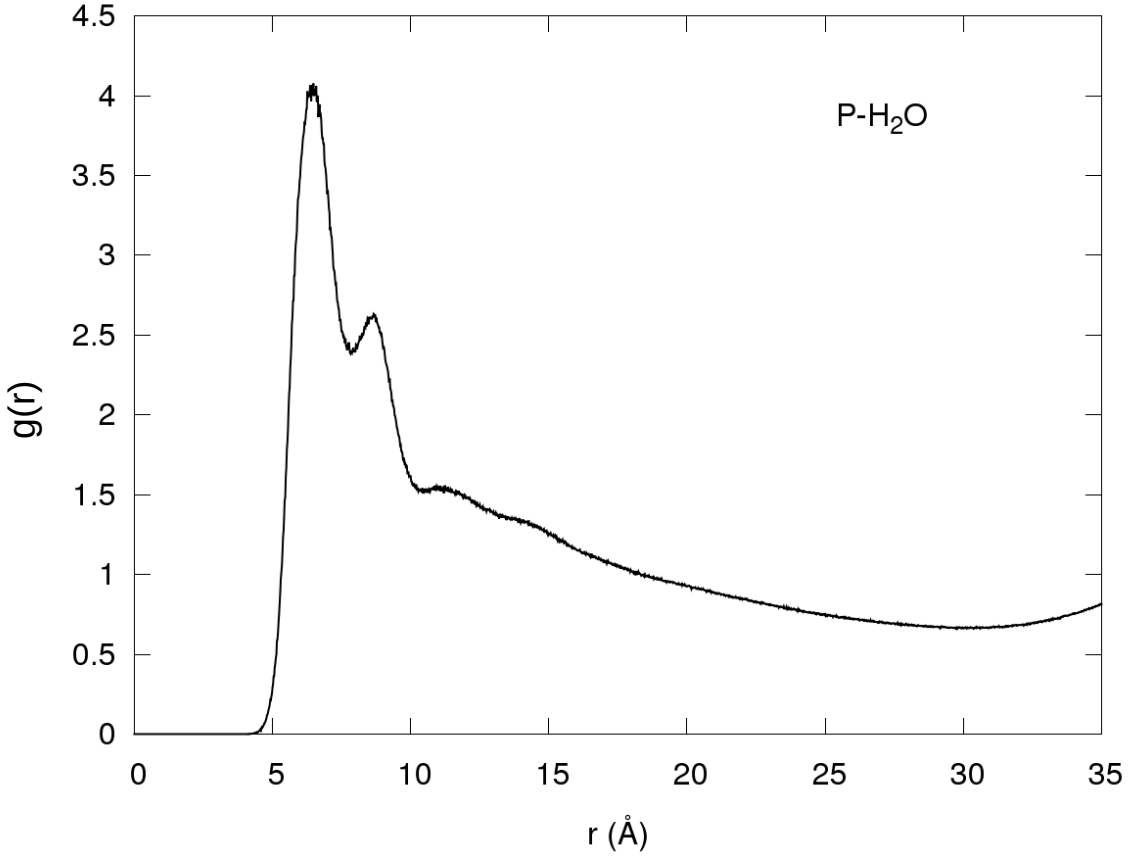

Figure 3.3: Phosphate water RDF for mapped atomistic trajectory. The first solvation shell appears to end at approximately 7 Å away from the phosphate bead, while the second shell extends from 7 to 10 Å. The RDF calculated does not converge to 1 before 35 Angstroms due to the absence of water within the bilayer.

Following the selection of the solvation shell radius, a virtual site (named VS) was mapped as the center of mass of all the water molecules in this radius from each individual phosphate group on a frame-wise basis. This virtual bead was then covalently bound to the phosphate to create an alternative lipid topology with a semi-explicit solvent representation shown in Figure 3.4.[99] As the forces obtained by summing the forces on each atom corresponding to each virtual site do not

guarantee CG consistency with the mapped reference data, they were ignored, and positional data alone was included.

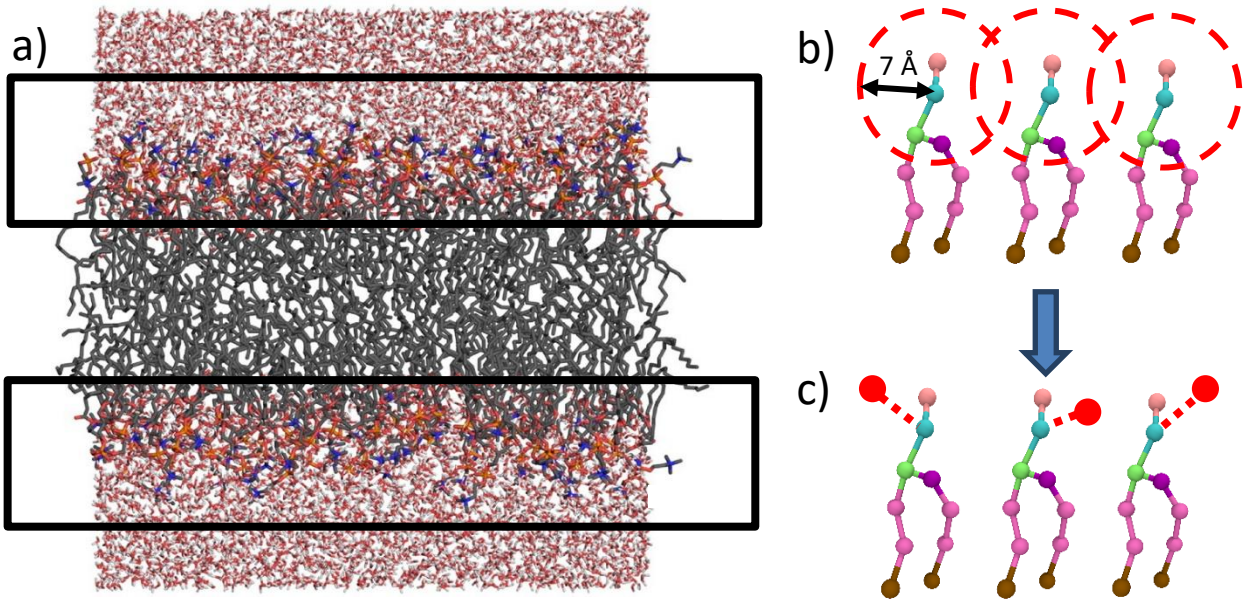

Figure 3.4: a) An atomistic lipid bilayer. The water within the black rectangles differs from bulk solvent in that it diffuses from the bilayer much more slowly. This network of water is the primary solvent that the lipid heads interact with. Adapted from [99]. b) CG lipids shown with the first phosphate solvation shell. Notably, these shells overlap, and individual waters may be within multiple lipids solvation shells. c) CG DMPC with solvent virtual sites bound to the phosphate beads.

In order to begin REM for the VS-REM, guess potentials for interactions involving the VS beads were assumed to be soft cosine potentials which decayed from 1 to 0 kcal/mol over the course of 1 Å. Bonded interactions between the VP and PH were assumed to be harmonic with a force constant of 2 kcal/angstrom^2 and an equilibrium distance of 1 angstrom, corresponding to the average position of the VP beads relative to the phosphate beads in the reference data. Selection of inner cutoffs for non-bonded VP interactions followed the scheme used for non-virtual particles, and the soft cosines were centered around this cutoff. The topology was altered at this stage to switch the ST beads to SM beads. This resulted in a less complex force field with 6 types including

the VS beads. Following the same iteration scheme as the IS-REM model, 250 iterations were required to converge VS-REM.

To create the VS-MSCG model, the forces for non-virtual particles were obtained from the reference data, while the VS-REM model was used to estimate the virtual site forces in the reference data mapped to the virtual site resolution. This hybrid reference data was then used to force match the model, using the same parameters used to generate the IS-MSCG model. Two versions of this model were generated with separate topologies, one using the original IS-MSCG lipid mapping (with separate types for middle and terminal tail beads), and the other using the IS-REM topology (with only SM bead types for the tails).

Finally, the TT-VS-REM model was generated using the VS-REM as input for REM. Final coordinates for the initial reference data were used to spawn three new trajectories, each annealed to 290, 310 and 320 K respectively. Once equilibrated, each system was integrated for 20 ns. Each reference trajectory was used to parameterize a separate VS-REM model at a different temperature using the same parameters. For each temperature, the model at the 500th iteration was selected for analysis.

For all CG models, simulations were carried out in LAMMPS at NVT conditions.[100] Unless otherwise specified, each simulation was carried out at 1 atm and 300 K. Additionally, due to the smoother free energy surface upon which CG models evolve, a larger timestep of 10 fs was used for production simulation, while 5 fs timesteps were used for equilibration, unless otherwise specified. Each system was simulated using special bonds between 1-2 (bonded) and 1-3 (angular) interactions.

For the calculation of membrane bending moduli, simulations were carried out on larger lipid bilayer patches. For lipid simulations, the bending modulus is calculated via the low

frequency modes of the height fluctuation spectrum, and thus more accurate calculations of the bending modulus can be obtained from larger systems. These larger lipid patches were generated by replicating the original 1260 lipid system used to parameterize the models 3 times in both the x and y direction, resulting in a total 11340 lipids, or approximately a 60 nm by 60 nm lipid patch. Each system was equilibrated to de-correlate the replicated bilayer and then simulated for 50 ns, with frames captured every 0.01 ns, for 5000 total frames.

To calculate the entropy-enthalpy decompositions of the PMFs of association for each model, 100 ns of simulation were carried out for each system using the smaller 1260 lipid 20x20 nm lipid patch. Each system was simulated at 290, 300, 310, and 320 K, with a short equilibration performed before each production simulation. Positions were saved every 0.01 ns for a total of 10000 frames, which were used to calculate RDFs for each interaction at each temperature. The PMF of each interaction was then calculated using these RDFs via the reversible work theorem. As this PMF is a conditional free energy, it can be decomposed into an entropic and enthalpic component. For each value of r, the value of the PMF is a linear function of temperature. The y-intercept of this function is the enthalpy of the PMF, while the slope corresponds to the negative of the entropy of the PMF. Linear regression was used to estimate the values and errors of entropy and enthalpy at each distance, which can then be plotted to see how each model's entropy and enthalpy are partitioned.

3.3 Results and Discussion

Results of the five DMPC models are discussed as follows: First each model is discussed from a qualitative standpoint, covering overall bilayer stability and structure. Then, each model is compared quantitatively, with bending modulus, radial distribution functions and entropy-enthalpy

decompositions used to assess the overall quality of the models from the standpoint of statistical thermodynamics. Lastly, the amount of effort required to generate each model is reviewed as speed is a critical feature of CG models.

IS-MSCG

The IS-MSCG model fails to produce a proper bilayer. While it does not completely fall apart despite the total lack of solvent representation, the bilayer splits into multiple cylindrical micelles in a short amount of time. Figure 3.5 shows a snapshot of this behavior. While this mapping scheme produced stable bilayers for explicit solvent top-down models when using a Lennard Jones style potential to replicate the RDFs of the system, force matching fails due to large forces mapped to the interactions involving ST beads, despite the fact that similar trends are observed in bottom-up models.[101]

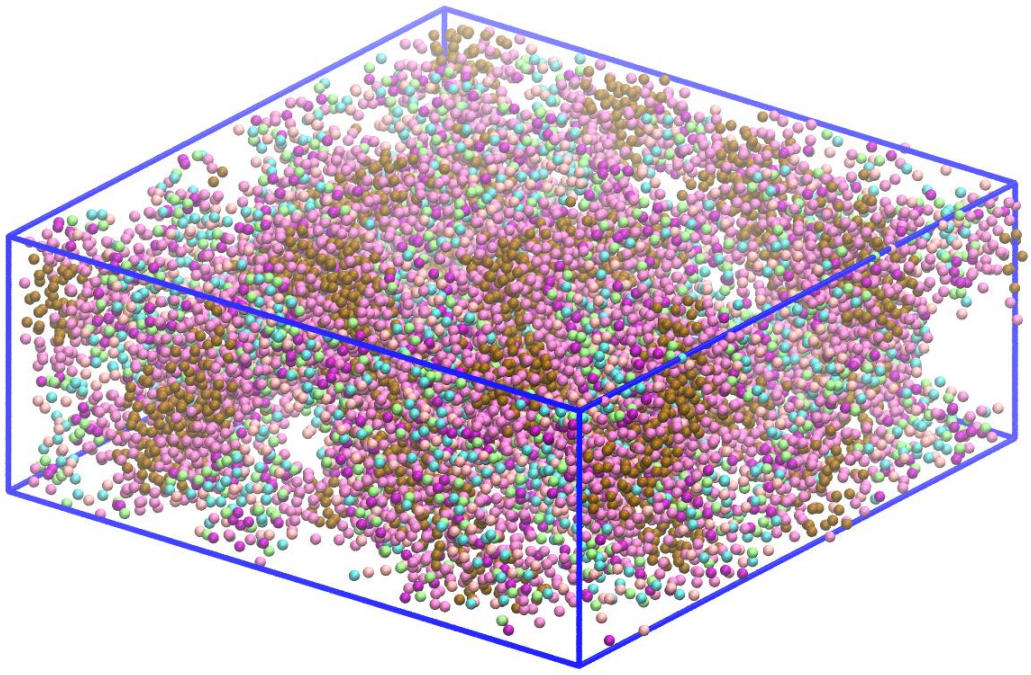

Figure 3.5: Snapshot of IS-MSCG DMPC model after 10 ns of simulation. The system started in a fully formed bilayer which slowly became a series of separate cylindrical micelles. Upon replacing the ST beads with SM beads, the initial membrane form is retained.

Figure 3.6 shows a comparison of SM and ST potentials obtained via force matching. Previous studies of DMPC using a similar resolution resolved similar issues by replacing the ST beads with additional SM beads.[88]In the case of the IS-MSCG model, this method also works. However, by changing the topology before force matching is performed, the model fails in a different way. The forces mapped to the tail beads in this case are altogether too low to keep the

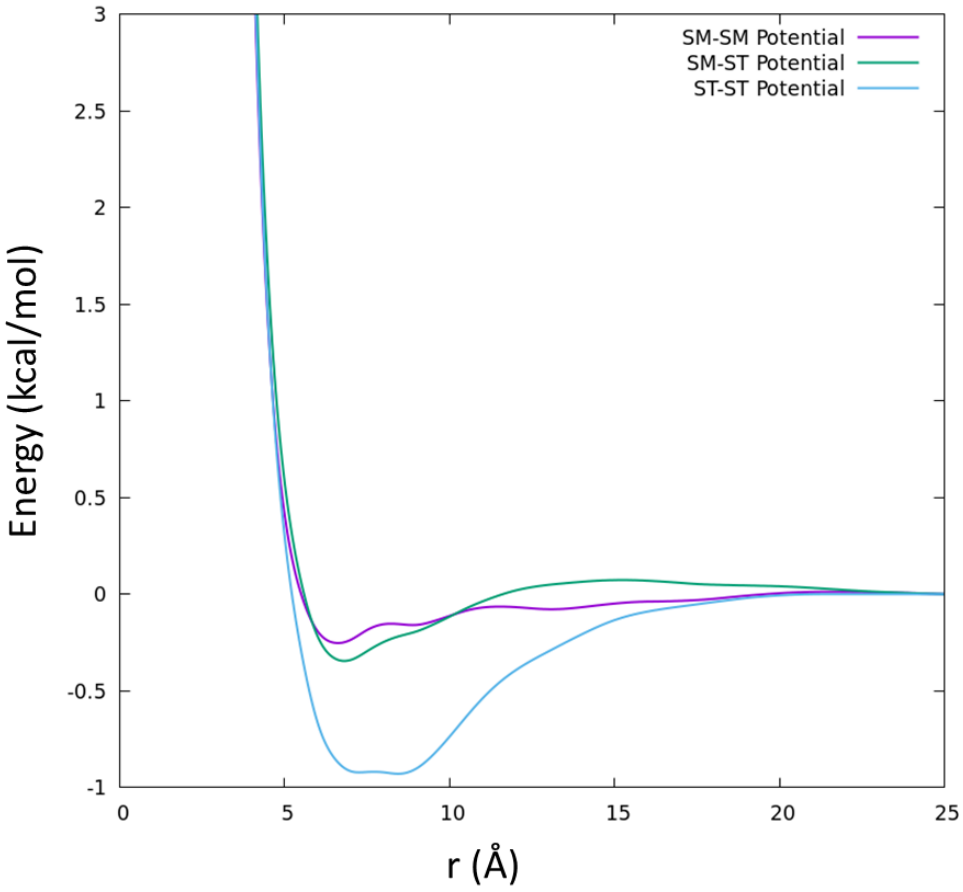

Figure 3.6: Tail-tail potentials for unaltered IS-MSCG model. The ST-ST interaction is especially attractive which leads to collapse of the bilayer as the ST beads aggregate to a degree which destabilizes the entire system. Replacing these interactions, along with all other ST containing interactions, restabilizes the bilayer.

leaflets together, and the bilayer disassembles rapidly. Thus, for all analysis performed in this work, the IS-MSCG model refers to the model parameterized with ST beads but simulated as though each ST bead was an SM bead.

IS-REM

The IS-REM model does not suffer from the micelle formation or disassembly issues which the IS-MSCG model demonstrates. However, the bilayer that it does form has a tendency to porate. These issues fail to resolve even after 1000 iterations of REM were run, despite the much smaller step size used than in previous REM models of lipids at lower resolutions. Figure 3.7 shows a snapshot of the IS-REM model which has porated. This bilayer still manages to match the RDFs of the reference data which is possible due to the anisotropy of the bilayer. Because REM works by matching radial densities and the bilayer is flat, the densities can match up radially around the edges of the pores. In addition the lips of the pores are far outweighed by the sampling of the overall bilayer, which may wash out any inaccuracies they are causing.[102]

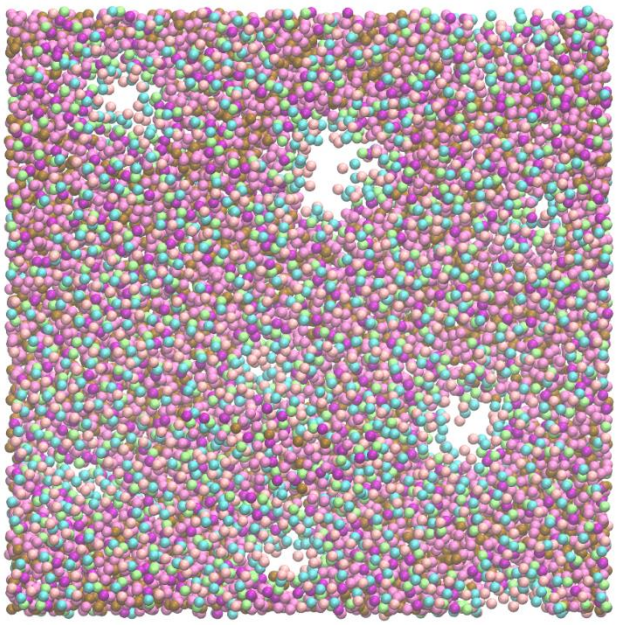

Figure 3.7: A snapshot of the IS-REM model taken after 50 ns. The bilayer has developed several pores. These pores are transient, and can merge with one another, but the bilayer always contains several once they develop.

VS-REM

The VS-REM model produces a bilayer which is qualitatively much better than the previous 2 models. Despite the fact that the model used the IS-REM model as a guess to the REM solution

(with additional virtual site parameters), the virtual sites manage to produce a more stable bilayer with no poration. While certain iterations of REM exhibit a slight amount of poration, the final model selected produces a hole-free bilayer.

VS-MSCG

The VS-MSCG model suffers from similar problems as the IS-MSCG model. Three separate versions of the model were tested: one with all tail beads parameterized as SM beads, one with separate ST beads, and one with separate ST beads which are replaced with SM beads after force matching. The SM only model exhibits significant interdigitation. Both leaflets collapse into one another forming a nearly unordered layer of lipids. When the ST beads are included, the model is far too flexible and eventually falls apart. When these beads are replaced with SM beads, the model still falls apart, forming small membrane like patches which spread throughout the simulation box. Because a functional bilayer could not be formed, the VS-MSCG model is excluded from quantitative analysis such as, the bending modulus and entropy-enthalpy decompositions. Figure 3.8 shows snapshots from simulations from all three versions of the model.

TT-VS-REM

The TT-VS-REM model is composed of four separate models, each tuned to a different temperature using the same REM parameterization scheme. For this reason, all analysis of the model which were performed at 300 K, namely RDFs and bending moduli, are identical to those of the previous VS-REM which was used as the starting point for TT-VS-REM. Qualitatively, the model is roughly the same as VS-REM, though integrating the 290, 310, and 320 K models for longer time periods demonstrate some issues. Outside of 300 K, the TT-VS-REM model begins to porate over time when started from a stable bilayer.

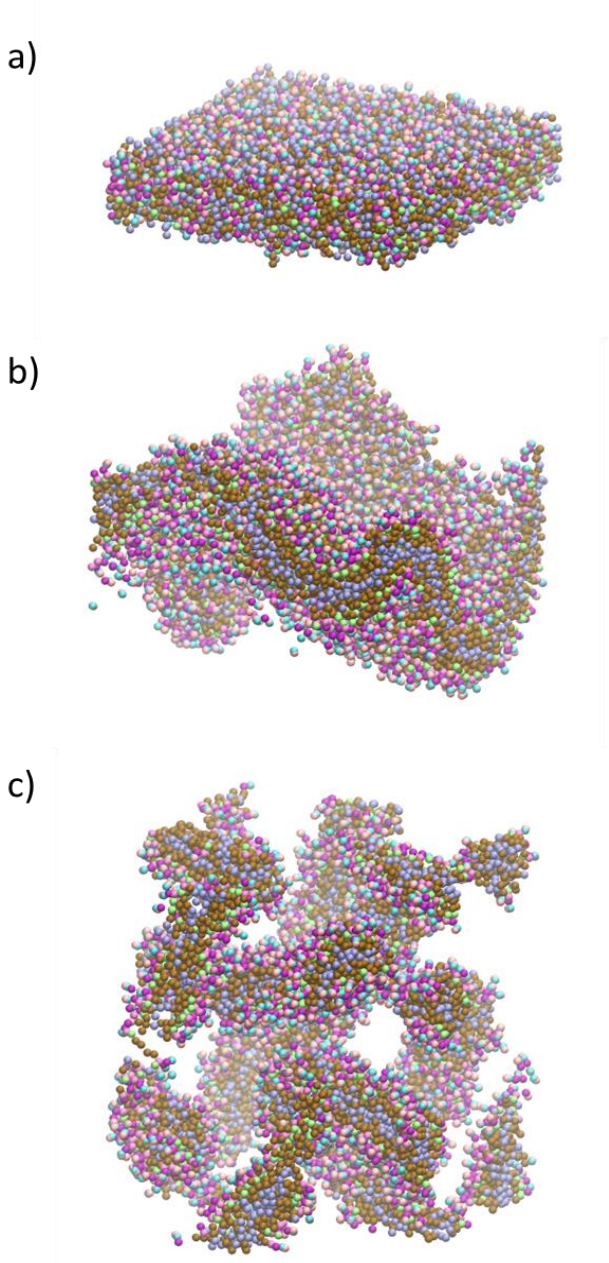

Figure 3.8: a) Snapshot from VS-MSCG model parameterized with purely SM tail beads. The bilayer maintains its overall shape, but the leaflets have merged. b) Snapshot from VS-MSCG model with ST beads. The bilayer deforms to an extreme degree, eventually disassembling. c) Snapshot from VS-MSCG model with ST beads swapped to SM beads after parameterization. Bilayer quickly forms small patches.

Figure 3.9 shows selected RDFs comparing the IS-MSCG, IS-REM, VS-REM, VS-MSCG and TT-VS-REM models to mapped AA reference data. ST beads are treated separately even though they are treated as SM beads for the purpose of MD integration. This was done to

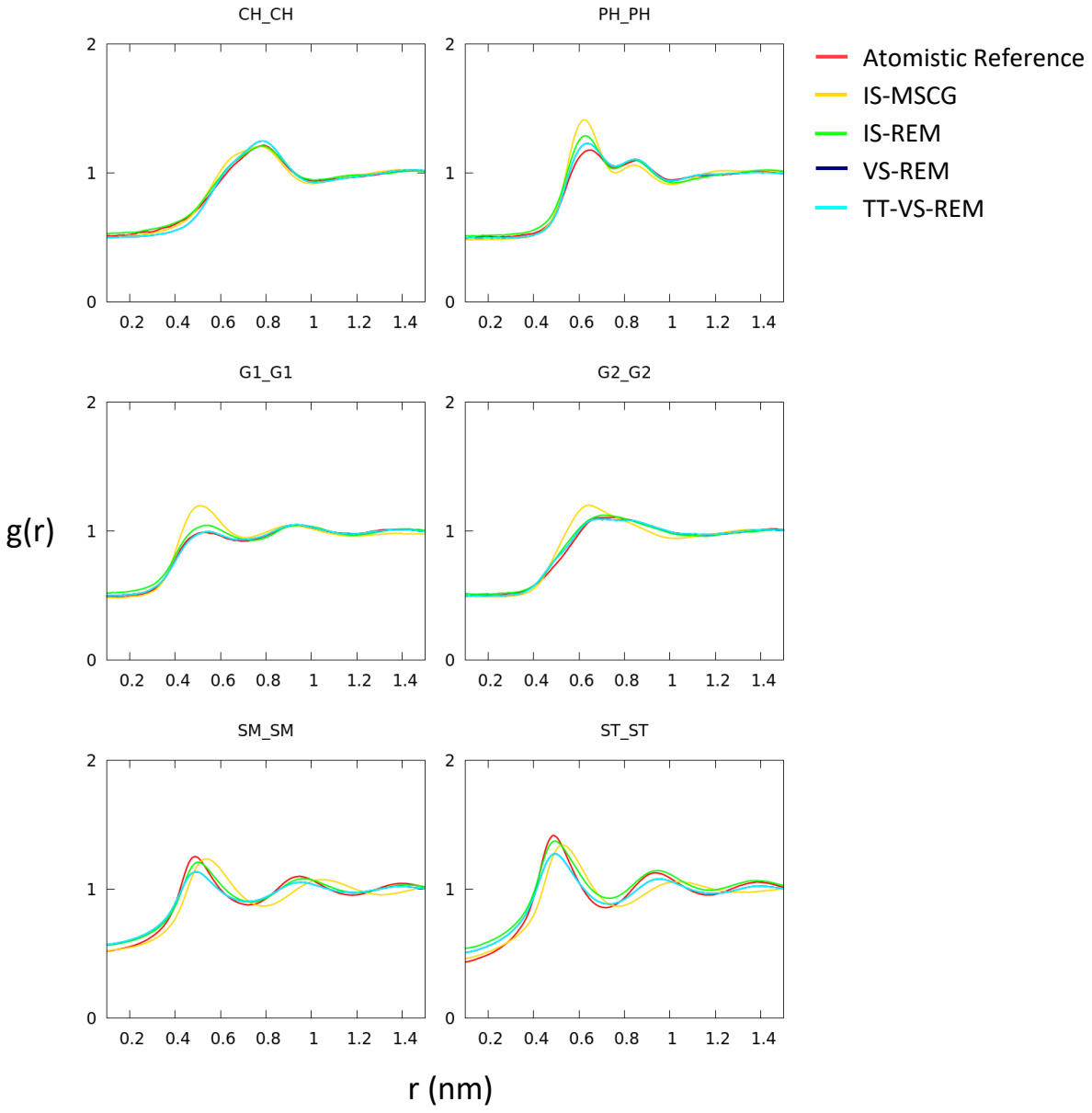

Figure 3.9: a) Selected RDFs comparing atomistic, IS-MSCG, IS-REM, VS-REM and TT-VS-REM DMPC models. RDFs decay to ½ at distance 0 due to effects from the opposite leaflet. VS-MSCG model omitted as it cannot form a stable bilayer.

interrogate the structural and thermodynamic properties of the ST beads. These beads seem to play a unique role when compared to the SM beads. First, they are primarily responsible for membrane integrity in implicit and semi-explicit solvent models regardless of whether they are treated separately. Secondly, they play a unique role in the generation of all force matching based models, where they must be treated as distinct sites for the purpose of parameterization but must be treated as SM sites for the purpose of integration. In general, the REM based models perform better than the IS-MSCG. As REM targets structural correlations this is not surprising. What is surprising is the difference between the IS-REM and VS-REM models. Both models perform best for certain interactions, with IS-REM outperforming the virtual site model within the hydrophobic core and within the choline-choline interactions. On the other hand, VS-Rem (and TT-VS-REM by construction) performs better for glycerol and phosphate RDFs.

By simulating all models across a range of temperatures (290 K to 320 K) and calculating RDFs at each temperature, the enthalpic and entropic components of the PMFs of lateral association in each model were analyzed. Figure 3.10 shows selected enthalpy decompositions comparing each of the five models with the reference data, while figure 3.11 shows the entropic component of the same interactions. Again, the terminal tail (ST) bead was treated separately from the middle (SM) beads. As expected, TT-VS-REM outperforms all other models due to its explicit temperature dependence, as well as its use of REM which targets structural correlations which are directly connected to the RDF, which is in turn directly connected to the PMF of lateral association. One exception to this is the well at 5 Angstroms in the G1-G1 interaction, which TT-VS-REM misses entirely despite VS-REM capturing it very well. In fact, all models aside from TT-VS-REM show a well at this location. Importantly, the virtual sites alone do not seem to improve the model to any real degree, as the IS-REM and VS-REM produce similar results. In fact for certain

interactions, such as CH-CH, the IS-REM model better captures the locations of the peaks, although the VS-REM model captures the heights of those peaks much better. In most cases, the force matching models perform worse than the REM based ones, which is again not particularly surprising given that these decompositions rely on accurate RDFs, which is not directly related to the forces being matched. Here the tail bead interactions provide interesting results. For the SM-

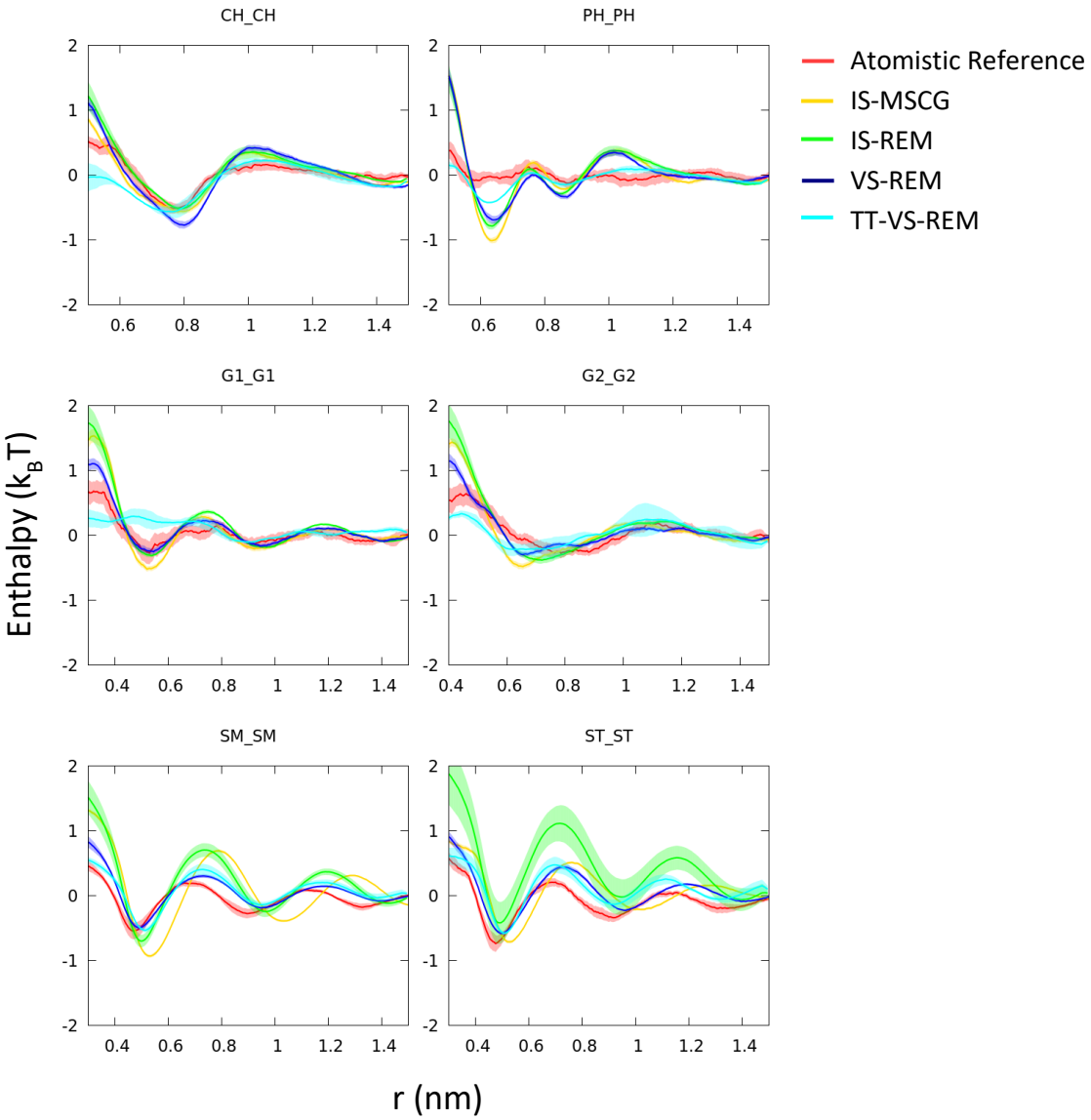

Figure 3.10: a) Enthalpic components of selected inter-bead PMFs comparing Atomistic, IS-MSCG, IS-REM, VS-REM, and TT-VS-REM DMPC models. VS-MSCG omitted due to its inability to form a stable bilayer.

SM interaction, the IS-MSCG model tends to over-structure the interaction as much as the IS-REM model, while the VS-REM model and TT-VS-REM model perform much better. However, for the ST-ST interaction, the IS-MSCG model performs much better than IS-REM, matching the atomistic peak height and locations better than VS-REM.

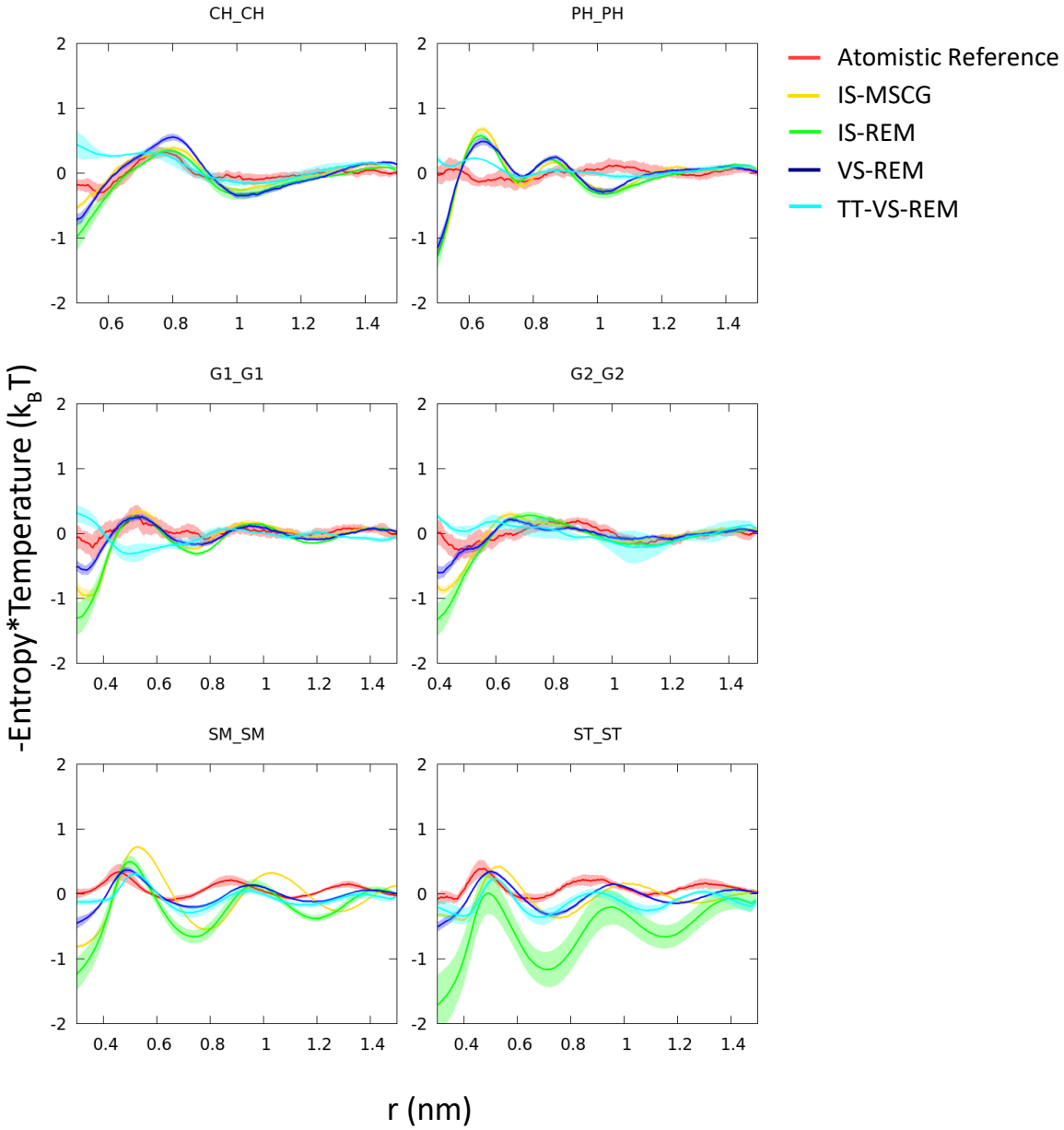

Figure 3.11: a) Entropic components of selected inter-bead PMFs comparing Atomistic, IS-MSCG, IS-REM, VS-REM, and TT-VS-REM DMPC models. VS-MSCG omitted due to its inability to form a stable bilayer. Temperature value chosen to be 300 K.

Table 3.1 compares the bending modulus of each model to that of the mapped atomistic reference data. The bending moduli were calculated using the low frequency limit of the Fourier transformed height fluctuation spectrum. The bilayer midpoint and height were calculated with respect to the PH bead, which provides a consistent estimate across both AA and CG resolutions. Previous studies comparing CG and AA bending moduli have demonstrated that when a phosphate bead contains just the central phosphorus atom and the four phosphorus bound oxygens, the AA and mapped AA fluctuation spectra match perfectly.[56] The addition of virtual sites has a clear positive effect on the bending modulus. The IS-MSCG and IS-REM models produces a bilayer which is far too flexible. The flexibility of the IS-REM model is strange, as typically REM models without virtual sites are far stiffer than reference data.[91] This can partially be explained by REM's sensitivity across iterations. There may be much stiffer bilayers in nearby iterations. Unlike previous applications of virtual sites to CG lipids, incorporating virtual site forces into the MSCG model fails to produce a stable bilayer, and thus the bending modulus cannot be calculated.

Table 3.1: Bending moduli for Atomistic, IS-MSCG, IS-REM, VS-REM and TT-VS-REM DMPC models calculated using height fluctuation spectrum. VS-MSCG model does not produce a bilayer.

DMPC Model	**Bending Modulus k_c (k_BT)**
Mapped Atomistic	30.7 ± 5.7
IS-MSCG	14.1 ± 0.3
IS-REM	5.0 ± 2.5
VS-REM, TT-VS-REM	30.5 ± 1.2
VS-MSCG	N/A

One critical aspect of a bottom-up CG model is the amount of time and effort required to create it. Thus, it is important to discuss the amount of effort required to create each of the models discussed here alongside their structural and thermodynamic properties. Each step in the model generation process eats away at time which could be spent simply simulating the system the CG

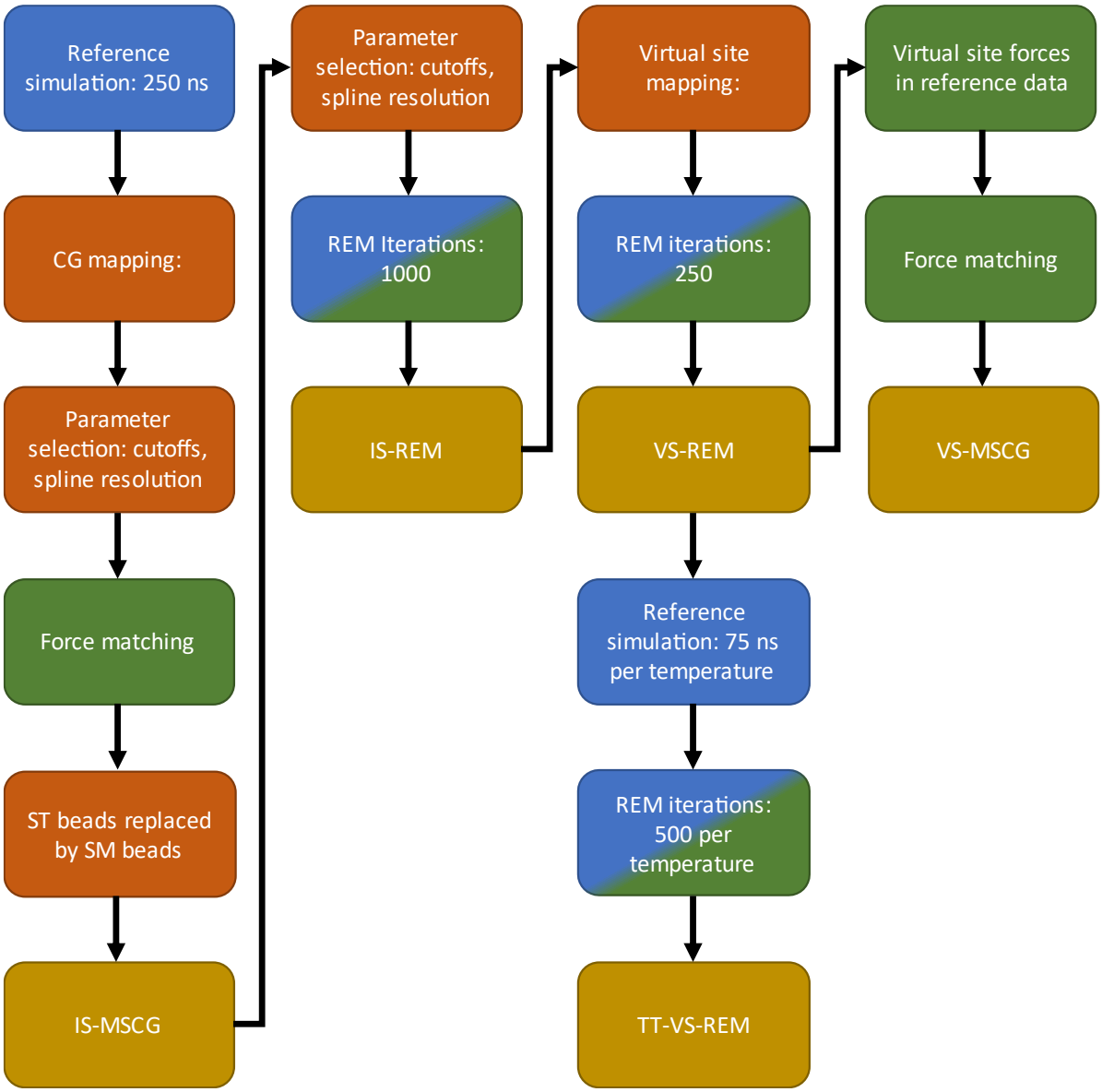

Figure 3.12: Workflow for generating each of the five DMPC models. Blue boxes involve atomistic or CG MD simulation which can be time consuming. Green boxes are CG algorithms and are generally not time consuming unless iterated. REM iteration involves both simulation and algorithmic optimization. Orange boxes involve parameter selections which may not be obvious for a given system and can take experimentation. Yellow boxes are finalized models.

model is built to speed up, and if enough time is spent on model development, there may be no point in generating the model in the first place. Figure 3.12 describes the workflow required to generate each model as well as the dependencies of each model.

For the IS-MSCG model, 250 ns of production simulation on a 1260 lipid membrane were used alongside force matching. The collection of reference data in this case was by far the most time-consuming step, as force matching a system of this size takes only a few hours. Still, this amount of reference data is far larger than previous studies have used (400 ps and 64 lipids total). In addition, there was effort involved in identifying the poorly sampled ST interactions and knowing to replace them with SM interactions. This step may or may not be the same for all systems.

The IS-REM model required the production of the IS-MSCG model, as well as 1000 iterations of REM using a small step size across two stages. Each iteration utilized reference data from the initial 250 ns simulation, but also involved short CG simulations (1,000,000 timesteps) to be carried out at each iteration. Performing this step using a single compute node took a considerable amount of time, roughly 10 days as the step size required to stabilize the bilayer enough to even create a non-porating model was much lower than previous efforts at lower resolutions. Lastly, the REM scheme used to create this model was not straightforward. Previous attempts to optimize every potential at every iteration resulted in rapid bilayer disassembly, and only when the tail interactions were optimized by themselves first was a bilayer produced, albeit a poor one.

VS-REM required the IS-REM model, 250 additional iterations of REM to optimize the force field including virtual sites, and access to a mapped atomistic trajectory containing solvent in order to map the positions of the virtual sites for REM reference data. This poses an issue as

water is commonly stripped from lipid trajectories at runtime in order to save space (the total reference set took up approximately 2.5 TB as forces for each atom were saved for use in force matching).

Going from the VS-REM model to the VS-MSCG model took the shortest amount of time, as force matching is not an iterative process and large trajectories can be force matched in parallel, cutting the model optimization time down from weeks to hours. The only other step required to generate the VS-MSCG model was estimating VS forces on the mapped reference data via a rerun, which is a fast step even compared to force matching.

By contrast, generating the TT-VS-REM model from the VS-REM model took the largest amount of computation aside from the initial reference simulation. Each temperature's model required 500 REM iterations and a shorter AA simulation at that thermodynamic state point, for a total of 75 ns of simulation and 1500 REM iterations. This is also to produce a model which can transfer over a range of 30 K, when real use cases may require a much larger temperature range. It is also not obvious how one would simulate the system in between the selected state points, without calculating the functional form of the temperature dependence of each interaction at each distance. This idea is elaborated on further in the following section of this chapter.

Integration speed is the last part of the equation determining how useful a CG model is. If a model takes a considerable amount of time to generate but integrates extremely quickly compared to the corresponding AA model, it may still be worth using. In the case of all 5 CG models, a 10 fs timestep was utilized, which is 5x the timestep of the reference CHARMM system. Additionally, the reference system contained 298 K atoms including waters, while the IS and VS models contained only 12600 and 13860 CG sites respectively, for an overall reduction in the system size

of roughly 94 %. Holding all other aspects constant, including the increase in sampling due to the smoother CG PMF, this equates to a speedup of approximately 125x.

3.4 Conclusions

Bottom-up CG models of implicit solvent lipid bilayers suffer from a variety of issues stemming from the inherent anisotropy of membranes. Lipid dynamics being limited to lateral diffusion (except for a small amount of flip-flop of lipids between leaflets which is not observed often enough to matter) lead to far less sampling between lipid heads and tails than is needed to produce effective potentials using force matching or REM. In addition, the lack of explicit solvent representation leads to unrealistic forces between lipid tails as the model attempts to keep the individual leaflets together without any water to impose amphiphilic ordering of lipid heads and tails. Based on previous studies into CG lipids at a variety of resolutions, these problems seem to be exacerbated in higher resolution CG models, approximately 4 heavy atoms to one CG site. The increased complexity of the mapped topology for these models introduces a geometric expansion in the number of model parameters. This in turn creates a geometric expansion in the number of failure points, given that individual spline parameters can destabilize the membrane if they are sufficiently far from the ground truth result. It also has the effect of pushing the centers of mass between the head groups and terminal tail groups further apart than in mappings of 5 or more heavy atoms to one CG site.

In this work, 5 CG models of DMPC lipids are presented. The first, IS-MSCG, uses a 10-site mapping for DMPC, no solvent representation, and force matching to optimize model parameters. The second, IS-REM, uses IS-MSCG as a starting point for REM. VS-REM builds upon this model and incorporates virtual solvent particles to represent the water within the first

solvation shell of the DMPC phosphate groups. VS-MSCG uses VS forces from VS-REM in order to apply force matching to a model with these virtual sites. Finally, TT-VS-REM incorporates explicit temperature dependence to the model by reparametrizing VS-REM across a range of temperatures.

Qualitatively, each model aside from the VS-REM model has serious issues which preclude their usefulness as CG models. The IS-MSCG model suffers from the aforementioned issue in which several potentials in the force field must be swapped in an *ad-hoc* manner after parameterization. While this results in a stable bilayer, it undermines one of bottom-up CG's main advantage, that being systematic parameterization. The capacity of force matching, and other CG methods, to produce a usable model given nothing but a reference trajectory and a set of hyperparameters (e.g., cutoffs, spline resolution) makes them attractive and easy to use for those without extensive experience in generating CG models. However, the high-resolution lipid models consistently fail via interdigitation or complete bilayer collapse when force matching is applied and require special knowledge of how such models have been created in the past in order to resolve.

The issues with the IS-MSCG model are limited to the interactions containing an ST bead. As previously discussed, this is not surprising in some ways, as the implicit solvent representation of the IS-MSCG model tends to push attractive forces into the terminal tail beads. Despite the hypothesis that virtual sites would improve this specific area of the force field, the VS-MSCG model suffers from the same qualitative issues, suggesting one of two things. Either the virtual sites are not representing solvent forces in the way they were hypothesized to, or the resolution at which the model was mapped is simply a poor choice for lipid models. Successful applications of virtual sites to lipid models[91,103] suggest that the latter is the problem, as even implicit solvent models at this resolution do not exhibit these failures.

While the IS-REM model forms a stable bilayer, it forms multiple pores after a small length of time making it useless as a CG model. TT-VS-REM models outside of 300 K also porate. REM's tendency to generate models with this sort of behavior likely stems from the RDFs being suboptimal targets for CG optimization when the system is largely two dimensional.

The VS-REM model is qualitatively the best of the models presented, with no interdigitation, collapse, or poration. However, when compared to other virtual site models, it is found lacking. VS-REM DMPC cannot self-assemble, while 6 site models of DOPC with an included virtual site not only form bilayers from random starting configurations, but also form other membrane topologies such as vesicles, tubules, and branched tube networks.[91]

Quantitatively, the TT-VS-REM model performs best when targeting entropy-enthalpy decompositions. This is, as mentioned previously, unsurprising due to the explicit temperature dependence of the model, and the fact that CG models in general cannot transfer outside of the thermodynamic state point they were parameterized at.[30] In addition, the introduction of the virtual sites also leads to a much more accurate bending modulus. Its biggest drawback is its tendency to porate slowly at temperatures outside of 300 K, an issue which could likely be solved by additional REM iterations or a different set of REM parameters. Both implicit models suffer from poor bending moduli, with IS-REM being far too stiff and IS-MSCG being too flexible. Additionally, IS-MSCG produced the worst entropy-enthalpy decompositions overall. IS-REM on the other hand outperformed even the virtual site models in several cases, though it was overall similar in quality to the IS-MSCG model. Most surprisingly, the VS-MSCG model completely fails to produce a bilayer regardless of which choice is made for the terminal tail beads. For lower resolution bilayers, the act of incorporating virtual site forces into the mapped atomistic data produces the best results. In this case, it seems that virtual sites not only fail to overcome the issues

that the IS-MSCG model had but exacerbate them instead. This issue may be resolved via application of different force matching methods, such as iterative force matching which has been shown to improve MS-CG force fields in situations where structural distributions are not matched well or when a poor basis set is used.[104]

While VS-REM and TT-VS-REM are clearly the most accurate models overall of the ones generated, they are also among the most difficult to create. To include the virtual particles, a combination of force matching and REM was utilized in addition to the reference data collection. While previous virtual solvent lipid models at lower resolutions utilized a similar amount of sampling (hundreds of nanoseconds of simulation on roughly the same sized lipid patch), the resultant models are much better at capturing the dynamics of the membrane and are even capable of self-assembly. TT-VS-REM incurs an additional three sets of REM optimization to build in temperature transferability across a range of 30 K. Interestingly, VS-MSCG is the most similar model to the previously developed 6 sites plus 1 virtual site DOPC model in methodology but fails for the exact same reason that the IS-MSCG model does, suggesting that all REM iterations performed add little to no benefit to the model.

REM constitutes one of the major limiting steps in generating these models. The low step size required to stabilize the bilayer across iterations means that many more iterations are required to build the model, which adds a considerable amount of time to the process before data collection on the CG model can even begin. The model was parameterized on 5 cascade lake nodes, which could perform roughly 100 iterations of REM per day. With that in mind, REM alone took over a week to complete for the IS-REM model, with no real way to accelerate it via parallelization.

Overall, the amount of effort and time required to parameterize these models is high compared to their accuracy, especially in the light of more accurate models which work at an even

lower resolution. While one might expect to gain accuracy as a tradeoff for a higher resolution and thus more computationally expensive CG model, this does not appear to be the case for lipid bilayers. This is troublesome from one viewpoint: a model which removes less information from the atomistic reference data should be able to capture the properties of the reference data more accurately. However, from another perspective this is a good thing: There is no downside to creating a lower resolution, and thus faster lipid model, and in fact one avoids a series of pitfalls which slow down and complicate the model generation process. I conclude that in the future study of CG lipid models, the high-resolution regime should be avoided.

The collection of models analyzed here can also provide insight into the usefulness of virtual particles, REM, and temperature dependent CG force fields. The introduction of virtual sites in the VS-REM model finally allowed for a stable bilayer to be modeled with no poration, a problem which plagued the IS-REM model. They also notably improved the structural properties of the bilayer, cutting the bending modulus by over an order of magnitude down to values which closely align with the reference data. In addition, the usage of REM itself solved the issues caused by MSCG, either interdigitation or collapse of the bilayer, a problem which even the virtual sites couldn't address. Lastly, the explicit temperature dependence of TT-VS-REM produced a model which correctly decomposes the PMFs of association of the model into their enthalpic and entropic components.

These results provide a signpost towards better understanding how to generate accurate and efficient CG lipid models. Virtual particles are an excellent way to improve a model with only a marginal increase in computational cost. While the VS models presented here relied on a non-linear mapping between the CG and AA regimes, this may not be necessary for future models. Variational Derivative REM (VD-REM) is a machine learning based framework which uses

forests of decision trees to learn the relative entropy gradients of arbitrary virtual sites.[103] This means that virtual particles can be parameterized using REM which have no mapping whatsoever in the AA regime. This method was used to create a 6 CG site + 1 virtual site model of DOPC without using the positions of water molecules, which exhibits similar self-assembly behavior to similar models using solvent-based mapping. One could also apply this method to virtual particles which reside in between leaflets, which might be able to further address issues caused by CG forces on lipid tails. In this case, the virtual particles are not a representation of any real particles but can be thought of as an efficient way to expand the basis set of the CG force field in a way without resorting to expensive multi-body correlations.

The accuracy of the TT-VS-REM model's entropy-enthalpy decompositions suggests that more research into incorporating explicit temperature dependence is needed. In principle, each parameter should have a linear dependence in temperature, which is supported by prior explorations of temperature dependence in liquid systems.[58] This linearity suggests that temperature dependent terms could be fit to each value of r for a potential energy function, which could be then altered on the fly to produce a potential which can transfer across thermodynamic state points in the middle of a simulation.

Chapter 4: Coarse-Graining with Equivariant Neural Networks: A Path towards Accurate and Data-Efficient Models

Abstract

Recently machine learning has entered its way into the mainstream of coarse-grained molecular simulation. While a variety of methods for incorporating deep learning into these models exist, many them involve training neural networks to act directly as the CG force field. This has several benefits, the most significant of which is accuracy. Neural networks can inherently incorporate multi-body effects during the calculation of CG forces, and a well-trained neural network force field outperforms pairwise basis sets generated from essentially any methodology. However, this comes at a significant cost. First, these models are typically slower than pairwise force fields even when accounting for specialized hardware which accelerates the training and integration of such networks. The second, and the focus of this paper, is the need for considerable amounts of data needed to train such force fields. It is common to use 10s of microseconds of data to train a single model, which approaches the point of obviating the CG model's usefulness in the first place. As it

turns out, this “data-hunger” trap that neural networks predicting molecular energies and forces experience is caused in large part by the difficulty in learning force equivariance, that is to say the fact that force vectors should rotate while maintaining their magnitude in response to an equivalent rotation of the system. We demonstrate that for coarse-grained water, networks that inherently incorporate this equivariance into their embedding can produce functional models using datasets as small as a single frame of reference data, while networks without inherent symmetry equivariance cannot.

4.1 Introduction

Molecular dynamics (MD) has proven to be a powerful tool to study the molecular underpinnings behind a variety of interesting phenomena in the biological and material sciences.[105,106] MD’s efficient integration of Newton’s equations of motion provides a fast method by which statistical and information of a molecular system can be averaged over to arrive at conclusions about the thermodynamics and dynamics of the system. A variety of MD techniques have been developed to interrogate systems at a range of different accuracies. *Ab-Initio* molecular dynamics (AIMD) and path integral molecular dynamics (PIMD) based method incorporate quantum level treatment of electronic and nuclear degrees of freedom respectively allowing for the accurate simulation of light nuclei and even chemical reactivity.[107,108] On the opposite end of the spectrum is coarse-graining (CG), which seeks to accurately simulate a molecular system at a resolution below that of atomistic MD.[19] In recent years there has been a large amount of interest in applications of machine learning (ML) to molecular simulations both at the quantum[109–116] and CG[103,117–120] level. While many of these methodologies apply machine learning to generate a pairwise force field, most work by treating the machine learning model, typically a deep neural

network (DNN), itself as the force field. On the quantum end, DNN force fields tend to integrate faster than full quantum treatments of atoms as they can typically be evaluated as a series of matrix multiplications and are thus useful as a method to speed up integration while maintaining an acceptable level of accuracy. On the CG end however, ML based methods tend to be slower than a simple pairwise Hamiltonian but are much more accurate as they can naturally incorporate many-body correlations and are much better at fitting data than the linear regression or relative entropy minimization (REM) based methods typically employed.

ML has a long history of drawing inspiration from nature to create powerful models which can tackle problems which were previously considered intractable. The architecture of the first neural networks were, as their name suggests, inspired by neural function in the brain.[121] Similarly, convolutional neural networks take advantage of processes seen in animal eyes in order to identify and classify features above the single pixel level in image processing.[122,123] Both the brain and eye are remarkably powerful tools for learning and image processing respectively, so it is no surprise that these models can succeed at such tasks. In the application of DNNs to molecular systems, it should come as no surprise that taking inspiration from physics can lead to excellent results. Successful DNN based methods incorporate physical constraints into their architectures and training schemes to improve the often-nebulous connection between DNN regression and physical reality and to speed up training. Typical DNN based force fields, such as CGnet and CGSchnet, utilize the same objective function as the multiscale coarse-graining (MSCG) method,[66,67] the sum of squares of the residual forces between the mapped reference data and the CG model. The difference lies in the usage of a neural network to learn these forces versus force matching's least squares regression over each model parameter, typically either Lennard Jones parameters or b-spline knots.

Further physical intuition is applied to the structure of the networks themselves. These DNNs work in two stages. The first is a featurization stage in which the raw Cartesian coordinates for each particle in a configuration are converted into more natural internal coordinates while the second is a neural network which learns particle wise energies and forces from these featurized configurations. The simplest featurization scheme corresponds to converting the coordinates into interparticle distances or their inverses as well as particle types, which can be done directly as is in the case of Cgnets.[117] These features are then subjected to physically inspired energy priors, harmonic for bonded particles and repulsive for non-bonded ones. This frees the energy predicting network from needing to learn those features of the CG Hamiltonian and allows it to learn corrections to the priors instead.

Another approach for featurizing molecular configurations which generates even better results is to embed these features into a graph neural network as in Schnets and CGSchnets[109,118] which are naturally suited to representing molecular systems. Each node of the graph represents a CG site, and each edge a distance between the two CG sites representing each node. Convolutions over these graph elements can be performed analogously to convolutions over pixels in 2-D images, giving graph neural networks a powerful tool to pool information across a variety of spatial scales.[124] These networks can then be trained to learn an effective embedding of the CG configuration which optimally predict CG forces and energies, improving over the set of hand selected internal coordinates used by CGnets. This embedding network fits into the previously discussed architecture in between the original featurization into internal coordinates and before the energy prediction network. This method also produces networks which are inherently transferable, as the embedding network can develop an effective embedding for any configurations so long as

there are no CG types which haven't been seen by the network. For systems such as proteins, this is easy to accomplish, so long as the training dataset contains all 20 amino acids.

Both the regular handpicked featurization and the graph neural network implementation ensure that the resulting CG features are invariant to rotation and translation. This means that energies predicted by the neural network are insensitive to translations and rotations of the configuration, consistent with how these interactions work in the real world and in classical MD simulations. However, as these neural networks operate in the space of scalar distances and output energies, they rely on the traditional gradient of the potential to calculate forces. The directional elements of the distance vector between particles encode important information which directly relates to the directionality of the force between particles. While the energy of a CG configuration is indeed rotationally invariant, the forces are rotationally equivariant instead. Rotationally equivariant symmetries give the same results as invariant symmetries (e.g. the magnitudes of forces do not change with respect to rotation of the system, as well as additional properties related to vector encoded information. For example, when a system is rotated, the direction of the force vectors on each particle (an equivariant property) rotate the same amount and in the same direction. This relationship can be described more precisely by describing a symmetry as a group G which operates on vector spaces X and Y. A function $f(x)$ which maps from X to Y is equivariant with respect to G if

$$D_Y[g]f(x) = f(D_X[g]x) \tag{4.1}$$

Where $D_X[g]$ and $D_Y[g]$ are representations of element g in the vector spaces of X and Y respectively.

Recently, a class of equivariant neural networks has been developed which incorporate force equivariance for atomistic molecular systems.[111,112,116,125] These methods incorporate full vector information of the relative positions of atoms in addition to higher order tensor information to guarantee that the magnitude of the forces produced by these networks are invariant to rotation, translation and reflection (also known as the Euclidian or E(3) symmetry group) while the unit vectors describing the directions of these forces are equivariant under these operations. These properties impose a restraint on the networks based on real physics which in theory should make the networks far more capable of representing and predicting molecular forces. Specific implementations of equivariance differ from architecture to architecture, but this work focuses on the NequIP and Allegro models. These architectures take advantage of convolutional operations natural translational and permutational equivariances, and furthermore enforce that the convolutional filters are products of radial functions and spherical harmonics, which are rotationally invariant in order to achieve full E(3) equivariance:[126]

$$S_m^{(l)}\left(\overrightarrow{r_{ij}}\right) = R(r_{ij})Y_m^{(l)}\left(\widehat{r_{ij}}\right) \tag{4.2}$$

where $S_m^{(l)}\left(\overrightarrow{r_{ij}}\right)$ is a convolutional filter over full distance vectors between atoms, r_{ij} is the scalar distance associated with $\overrightarrow{r_{ij}}$ and $\widehat{r_{ij}}$ is the corresponding unit vector. Allegro and NequIP differ in that NequIP is globally equivariant. NequIP achieves global equivariance via a message passing layer which passes messages from adjacent graph nodes. These layers can learn a variety of functions, from graph convolutions to graph-wide targets which encode information about the entire system.[127] Allegro removes this message passing layer and only achieves local equivariance as a cost to allow for parallelization of network evaluation, allowing it to scale to much larger systems.

The results of these equivariant networks address a key weakness that ML suffers from: the data requirements of training neural networks. To generate an effective DNN, one must supply a vast amount of training data, which can in turn make the resulting model less useful as it inflates the time required to calculate results. For example, prior training of a CGSchnet architecture to produce a force field for chignolin, a mini-protein containing 10 amino acids, required 180 microseconds of reference simulation.[118] On the other hand, a hetero-elastic network model for full length integrin, containing 1780 residues, was generated using 0.1 microseconds of MD simulation.[128,129] This disparity calls into question the usefulness of DNNs as force fields when applied to super-atomic resolution CG models even considering the highly accurate results they generate. As it turns out, a primary reason for the large number of examples required to train DNN force fields is equivariance with respect to molecular forces. Rotation invariant DNN based methods must learn the equivariance of forces via training reinforcement, which adds considerably to the data cost of creating these models. Equivariant neural networks build this information into the model inherently and have been shown to predict interatomic energies and forces for small molecules when shown three orders of magnitude less training data than symmetry invariant architectures with even greater accuracy.[112]

While training neural networks to predict energies and forces of atomistic systems from *ab-initio* quantum data is not the same as training a CG model from atomistic data, there is a natural analogy of learning to predict forces at a lower resolution from higher resolution data. In the former case, the high resolution is the quantum description of the system, while the low resolution is the atomistic description. In the case of CG, the high resolution is the atomistic description, while the low resolution is some chosen super-atomic resolution. A key difference remains in that most methods which learn atomistic descriptions from quantum data treat bonded and non-bonded

interactions as the same, and rely purely on internal coordinates to decide this, while CG DNN methods use labels and alternative energy priors to do this. This is necessary for complex CG systems as bond breaking and forming is typically ignored for these models, and because the length scales of bonded interactions can easily match and overlap with that of the non-bonded ones. For this reason, the currently available equivariant neural network-based methods must be used carefully when applying them to CG systems.

There are certain cases for which the methods are fundamentally identical. The simplest case is that in which there are no bonds whatsoever, and each CG bead corresponds to an entire molecule. In this case, the act of making a CG model is equivalent to the act of reducing the quantum description of a nonreactive single particle, such as helium, to its atomistic representation. For this reason, this work is limited in scope to the coarse graining of single site liquid, namely single site water. Water is also an ideal test case for a DNN CG method due to the high levels of correlation caused by hydrogen bonding. For this reason, single site water models tend to fail to predict proper center of mass radial distribution functions (RDFs) for water unless they incorporate many-body correlations.[130–132]

In this work we present an analysis of DNN based CG models of single site water, using a rotationally invariant CG model and a rotationally equivariant model. For the invariant model, the Deep Potential Molecular Dynamics method with smoothed embedding (DeePMD)[133,134] is utilized. For the equivariant model, the Allegro model[125] is utilized. Each method is applied to water in the limit of low sampling: a maximum of 100 consecutive MD frames are used to train each model. This work is laid out as follows: First a discussion of the methods used is given, with hyperparameters for all ML methods as well as all MD simulation parameters. A discussion of DeePMD and Allegro models is also presented. Following this, results for each model are

presented. Pairwise RDFs are analyzed and compared to mapped atomistic reference data. Three body angular correlations are also analyzed. Finally, the stability of each force field in the low sampling limit is discussed as well. These results are then discussed and conclusions on the usefulness of equivariant particle embedding in the field of CG modeling are drawn.

4.2 Methods

In order to generate the dataset used to train the models, LAMMPS[100] was used to simulate 512 TIP3P[93] water molecules for a total of 10 nanoseconds in the NVT ensemble. A Nosé-Hoover thermostat[135,136] was used to keep the simulation at 300 K, and frames were captured every 2 ps for a total of 5000 frames, though far fewer were used in the training of the Allegro and DeePMD models. The resulting trajectory was mapped to a 1 CG site per water resolution using a center of mass COM mapping scheme. This was then passed as a training dataset to DeePMD and Allegro.

The DeePMD method consists of both an embedding network and a fitting network. The embedding takes pairwise distances as input and output a set of symmetry invariant features which include three-body information such as angular and radial features from nearby atoms, denoted by the authors as the se_e2_a embedding. Notably, this embedding network is not a graph neural network. However, before the interatomic distances are fed into this matrix, they are converted into a set of coordinates based on inverse distances:

$$\{x_{ij}, y_{ij}, z_{ij}\} \rightarrow \{s(r_{ij}), \hat{x}_{ij}, \hat{y}_{ij}, \hat{z}_{ij},\} \tag{4.3}$$

$$s(r_{ij}) = \begin{cases} \dfrac{1}{r_{ij}}, & r_{ij} < r_{c1} \\ \dfrac{1}{r_{ij}}\left\{\dfrac{1}{2}\cos\left[\dfrac{\pi(r_{ij} - r_{c1})}{(r_{c2} - r_{c1})}\right] + \dfrac{1}{2}\right\}, & r_{c1} < r_{ij} < r_{c2} \\ 0, & r_{ij} > r_c \end{cases} \quad (4.4)$$

where x_{ij}, y_{ij} and z_{ij} refer to the x, y and z projections of r_{ij} the distance between two particles i and j, and $\hat{x}_{ij} = \frac{s(r_{ij})x_{ij}}{r_{ij}}$, $\hat{y}_{ij} = \frac{s(r_{ij})y_{ij}}{r_{ij}}$, and $\hat{z}_{ij} = \frac{s(r_{ij})z_{ij}}{r_{ij}}$. This set of features is then converted via the embedding network into a matrix of features which preserves the rotational, translational and permutational symmetry of the system. These features are passed through per-atom subnetworks which compute the energy contribution from each atom to the total system energy. The gradients of these per-atom energies can then be used to calculate interatomic forces during the course of an MD simulation. During training, the DeePMD model sees individual atoms as training samples which can be batched as usual.[134]

Allegro is an extension of the NequIP model which trades global equivariance gained via a message passing graph neural network for local equivariance in order to provide much greater scaling capabilities.[125] In allegro, two sets of features are generated by the initial featurization for each pair of particles. The first is a scalar set of features which consists of interatomic distances and labels for each chemical species in the interaction. This feature set is symmetry invariant, as in the case of DeePMD. The second feature set contains unit vector information which correspond to these interatomic distances which are projected onto spherical harmonic functions. These features are then embedded through a series of layers onto a new equivariant feature set which is then fed into a multilayer perceptron (MLP) which predicts the energy of the interaction. The total

energy of the system can be calculated as the sum of these energies, and the forces can be calculated again via the gradients of these energies.

For both Allegro and DeePMD models, a common network size was selected to ensure that differences in the performance of the models was most strongly correlated with the amount of training samples. The embedding networks were composed of three layers with widths [8, 16, 32]. In the case of DeePMD models, this format is converted into a ResNet[137], for which no timestep was selected. The energy fitting networks were also composed of three layers each with widths [32, 32, 32]. Each network was given a maximum cutoff for the environment of each atom of 7 Angstroms. Training parameters such as numbers of epochs, learning rates, and early stopping were left up to the defaults of each model archetype to ensure that each model was trained according to its normal usage. Specific parameters for each model may be found in the Supplementary Information in the form of actual DeePMD and Allegro input files.

A total of 5 models were trained according to the preceding description. For DeePMD, two models were trained, one using 100 frames (or 200 ps) of training data, and one using 10 frames (or 20 ps) of data. Three Allegro models were trained, using 100, 10 and 1 frame of training data respectively. Each model was frozen and simulated for 2,500,000 timesteps with frames selected every 1000 timesteps in order to calculate CG RDFs and 3-body angular correlation functions. Each model was tested on a simulation of 3916 water molecules using a Nose-Hoover thermostat in the NVT ensemble at 300 K just as the reference data was. All models, except for the 10 frame DeePMD model, were simulated using a timestep of 2 fs, while the 10 frame DeePMD model used a 0.5 fs timestep, which is explained in the results section.

To calculate the RDFs, an outer cutoff for each model was selected to be 10 Angstroms. 3-body angular distributions, $P(\theta)$, for water were calculated using the following equation:

$$P(\theta) = \frac{1}{N}\left\langle \sum_{I} \sum_{J \neq I} \sum_{K > J} \delta\left(\theta - \theta_{JIK}\right) \right\rangle_{R<R_C} \tag{4.5}$$

Where N is a normalization constant equal to the largest value in the calculated sum and R_c is the cutoff radius. For these correlation functions, an outer cutoff of 4.5 Angstroms which corresponds with the second solvation shell of water originating from its tetrahedral ordering.[1] 3-body correlations were calculated between 30 and 150 degrees with a bin width of 1 degree, which captures the full extent of the 3 body correlations seen in tetrahedral water.[131]

4.3 Results and Discussion

Simulation stability:

Every model trained except for the 10 frame DeePMD model was stable when simulated using a 2 fs timestep. The 10 frame DeePMD model suffered from severe energy drift at this time-step and the system quickly falls out of a liquid state. Figure 4.1 shows sample coordinates for the

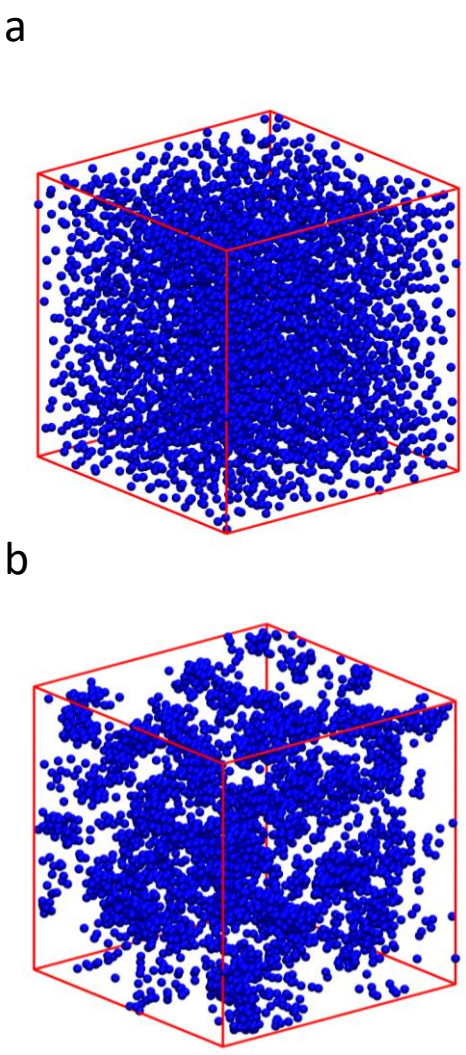

Figure 4.1: a) Snapshot of Allegro water model trained on 10 frames of AA reference data. The configuration was captured after 2,500,000 timesteps at 2 fs per timestep, or 5 nanoseconds of total simulation. b) Snapshot of DeePMD water model trained on 10 frames of AA reference data. In this case, the configuration was captured after 2,500 timesteps at 0.5 fs per timestep, or 1.25 picoseconds of total simulation.

10 frame DeePMD and Allegro models, detailing the extent of the collapse. Interestingly, even the repulsive priors present in the DeePMD architecture cannot prevent the particles reaching unphysically close distances. Within 2,500 timesteps, the water sites coalesce into small clumps. To simulate the model long enough to collect data, a 0.5 fs time-step was selected, as a 2 fs timestep lead to an almost immediate loss of multiple beads from the simulation box. In both DeePMD models, a 500 fs damping coefficient was not strong enough to stabilize the system at 300 K and both exhibited a lower temperature throughout the simulation. On the other hand, each Allegro model performed stably and did not naturally tend towards a lower temperature or collapsed state.

RDFs

Figure 4.2 shows RDFs for each model compared to that of the mapped reference system. Each stable NN model does a good job of capturing the structural correlations of liquid water, a difficult task for a single site CG water model. While none of the models fully capture the depth of the well directly beyond the first peak, all models aside from the 1 frame Allegro model are close, with the 100 frame DeePMD model performing the best. The 1 frame Allegro model overestimates the well depth and overestimates the height of the first peak more than any other model, though the overall trend is for this peak to be overestimated. None of the models can capture the small peak around 6 Angstroms either, with both 100 frame models performing slightly better than the 10 and 1 frame Allegro models. Surprisingly, while it overestimates the height of the peak the most, it also captures the location of the peak the best, with every other model peaking slightly past where the reference data does. The most noteworthy differences come from the 10 frame DeePMD model, which deviates significantly because of its propensity to collapse individual water

molecules onto one another. While the 1 frame Allegro model is the least accurate of the stable models, it still qualitatively captures the shape of the RDF, and in multiple cases, is slightly more

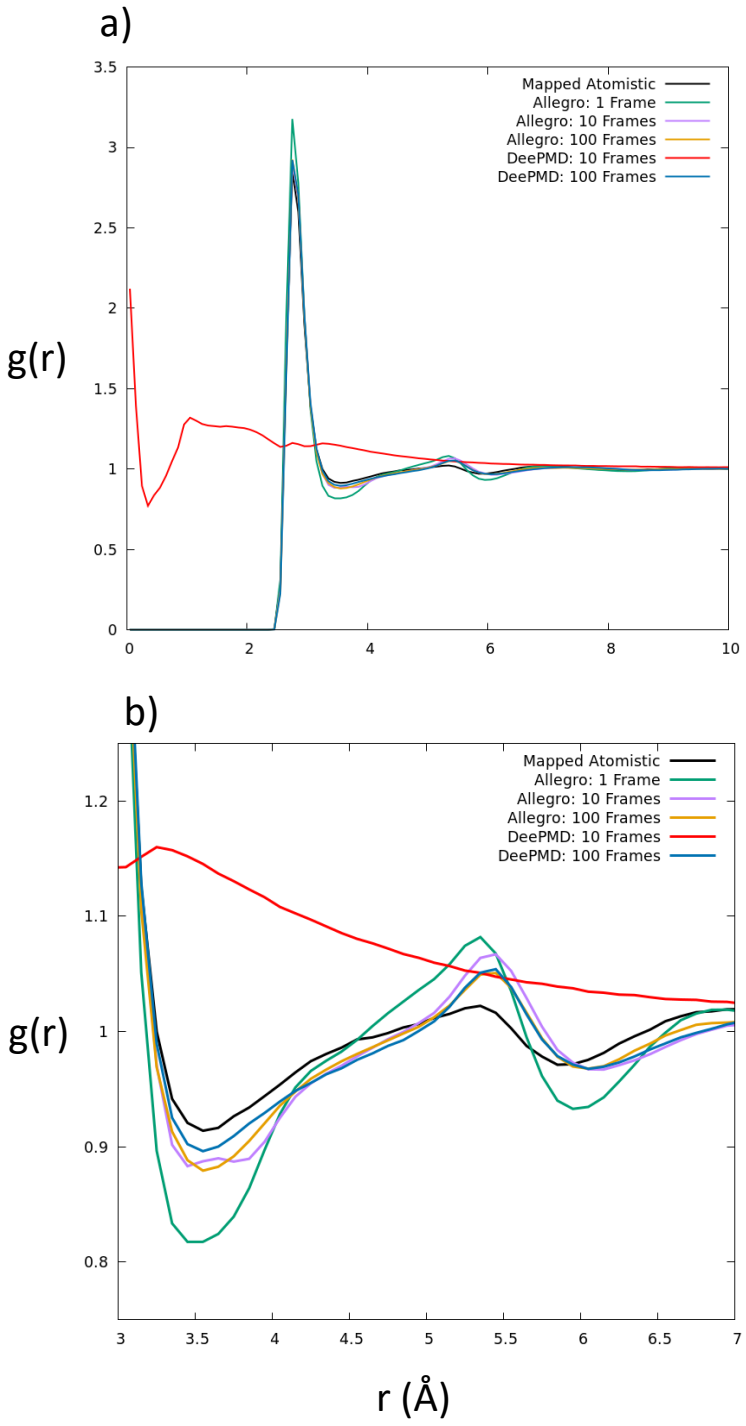

Figure 4.2: a) Radial distribution functions of CG water for Allegro and DeePMD models compared to reference atomistic data. b) Detail of RDFs in top panel between 3 and 7 Angstroms.

accurate with where peaks and wells are located. There is an overall trend of increasing quality as the number of training examples increases, present in both the Allegro and DeePMD models.

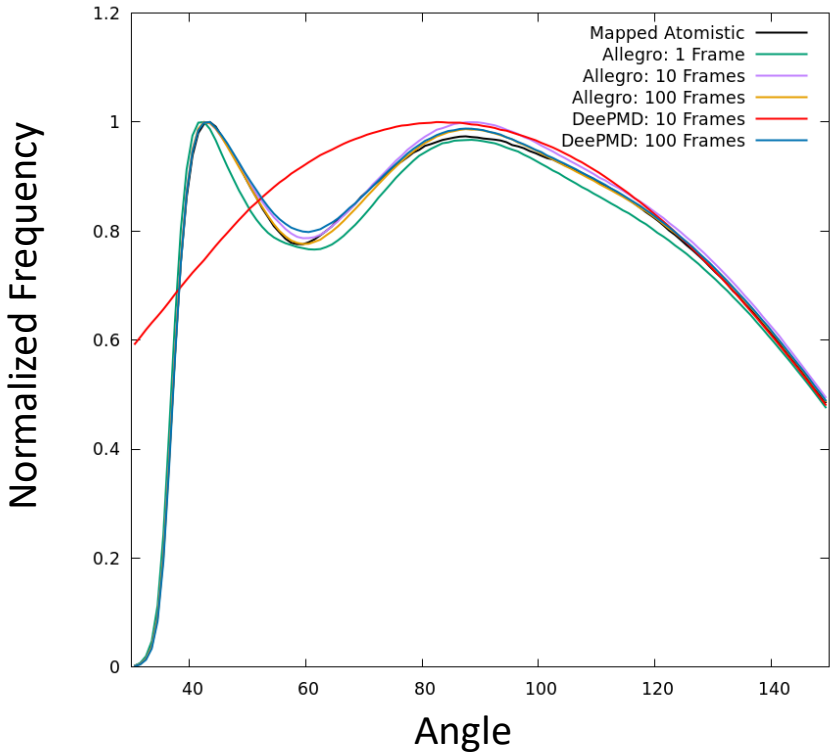

Figure 4.3: Three-body angular distributions between triplets of CG waters for each Allegro and DeePMD model compared to mapped atomistic data. Distributions were calculated using Eq. 4.3.

Three Body Correlations

Each model aside from the 10 frame DeePMD model performs reasonably well by this benchmark, although this is not very surprising given how expressive a DNN based force field can be. The true value of a DNN based force field is its ability to capture many body correlations which are simply out of reach of all but the most advanced pairwise potentials. Figure 4.3 shows water-water-water triplet angular distributions for each model parameterized in comparison to the mapped atomistic TIP3P water. As with the RDFs there is an overall trend of increasing accuracy

with respect to increased amounts of training data. Between the models trained on 100 frames of data, the Allegro model outperforms DeePMD, with much better representation of the well around 60 degrees and comparable accuracy everywhere else. Even the 10 frame Allegro model gives a more accurate result for this well, although it overestimates the second peak around 90 degrees the most of all stable models. Of the stable models, the 1 frame Allegro model is again the least accurate, predicting far lower probability values everywhere except for the first peak at 45 degrees. As before, the 10 frame DeePMD model completely fails to capture the correlations of water.

4.4 Conclusions

In this work we compare symmetry invariant and symmetry equivariant neural networks in their capacity to generate accurate CG force fields in the limit of low training data. Two architectures, DeePMD and Allegro were chosen for symmetry invariance and equivariance respectively. We show that symmetry equivariant models can form stable CG water models with just a single frame of reference data.

It is not surprising that holding model architecture constant the models trained on more reference data outperformed those parameterized on less. In all cases, the 100 frame models were able to accurately capture RDFs and 3-body correlations, though there is certainly room for improvement in both the Allegro and DeePMD models. It is likely that these architectures could produce even better models with the same training datasets if more hyperparameter sweeping was performed. In particular, the fitting and embedding networks were chosen to be far smaller than those used in previous studies to generate atomistic force fields from quantum mechanical simulations. For example, the original Allegro models utilized fitting networks with three hidden layers of 1024 neurons each, instead of the 32-width network used in the current work. This choice

was made to maximize their speed as CG models depend on integration speed to enhance the sampling of the underlying atomistic simulation. Despite this, these models integrate far slower than even a corresponding all-atom system, with the fastest model, Allegro, integrating at a speed of ~25 ns/day on a small 3916 particle system, even when utilizing 4 GPUs.

While there is still much room for improvement in the integration speed of DNN based CG force fields, the addition of equivariant embedding clearly reduces the amount of training data required to generate a stable model by orders of magnitude. Though it was not the most accurate model, Allegro could train a force field which reproduced all qualitative features of the 2- and 3-body correlations of water, with a reasonable amount of quantitative accuracy, even when using a single frame of training data containing 512 total training examples. In comparison, DeePMD could not create a model which stably formed a bulk liquid using 10x the amount of training data. This sidesteps one of the biggest hurdles for generating DNN CG force fields and suggests that incorporating physical intuition and restraints may increase their training efficiency. With additional advances in DNN integration speed, methods such as these could become the state of the art for CG modelling in the future. Explicit inclusion of bonded CG beads could expand the capacity of these models to much more complicated systems for which traditional CG methods fail.

Chapter 5: Centroid Molecular Dynamics Can Be Greatly Accelerated Through Neural Network Learned Centroid Forces Derived from Path Integral Molecular Dynamics

This chapter was originally published under the same title as Timothy D. Loose, Patrick G. Sahrmann, and Gregory A. Voth, *J Chem. Theory Comput.,* **2022** *18*(10), 5856-5863 DOI: 10.1021/acs.jctc.2c00706

Abstract

For nearly the past 30 years, Centroid Molecular Dynamics (CMD) has proven to be a viable classical-like phase space formulation for the calculation of quantum dynamical properties. However, calculation of the centroid effective force remains a significant computational cost and limits the ability of CMD to be an efficient approach to study condensed phase quantum dynamics. In this paper we introduce a neural network-based methodology for first learning the centroid effective force from path integral molecular dynamics data, which is subsequently used as an effective force field to evolve the centroids directly with the CMD algorithm. This method, called Machine-Learned Centroid Molecular Dynamics (ML-CMD) is faster and far less costly than both standard "on the fly" CMD and ring polymer molecular dynamics (RPMD). The training aspect of ML-CMD is also straightforwardly implemented utilizing the DeePMD software kit. ML-CMD is then applied to two model systems to illustrate the approach: liquid para-hydrogen and water. The results show comparable accuracy to both CMD and RPMD in the estimation of quantum

dynamical properties, including the self-diffusion constant and velocity time correlation function, but for significantly reduced overall computational cost.

5.1 Introduction

The accurate simulation of quantum dynamics is limited by the computational complexity of solving the time-dependent Schrödinger equation. While classical molecular dynamics (MD) based on empirical force fields are capable of utilizing certain information from quantum calculations, the treatment of the nuclei as point particles as well as the inherent limitations of a pairwise decomposable description of the intermolecular interactions limits their accuracy.[138,139] The most straightforward approach to address this accuracy problem is to perform full electronic structure calculations for each timestep to calculate the force vectors on each particle, such as in the case of *ab initio* molecular dynamics (AIMD).[140,141] However such methods can be prohibitively expensive for all but the smallest systems. Additionally, these calculations must be performed repeatedly over the course of a simulation which highly limits the timescales accessible to study with these methods. Many alternatives have been employed to limit the cost of representing the effects of quantum electronic structure within MD simulations, such as treating the majority of the system with classical mechanics and saving the quantum mechanical calculations for specific regions as in hybrid quantum mechanics/molecular mechanics.[142]

Irrespective of the underlying accuracy of the representation of the forces on the system nuclei, the challenge of quantum dynamics is compounded by the fact that for many systems of interest exhibit nuclear quantum effects (NQEs), even at thermal equilibrium. Path Integral (PI)

methods can capture NQEs in such equilibrium circumstances. This technique relies on the imaginary time formulation of Feynman path integral quantum mechanics, in which the classical principle of least action is generalized to quantum systems via functional integration over all paths that system can take between an initial and final point.[143] In the case of molecular simulation, PI based methods re-cast a quantum mechanical description of the system of interest into an isomorphic classical one where each quantum particle is represented by a set of P classical quasiparticles or "beads", for which standard MD algorithms can be utilized.[144,145] In particular, it can be shown that, in the P $\rightarrow \infty$ limit, the static equilibrium properties of a quantum particle can be described by the Boltzmann statistics generated by the following Hamiltonian:

$$H = \sum_{n=1}^{P} \left[\frac{p_n^2}{2m_n'} + \frac{1}{2} m\omega_P^2 (x_n - x_{n+1})^2 + \frac{U(x_n)}{P} \right] \qquad (5.1)$$

where m_n' are fictitious mass parameters, $\omega_P^2 = \sqrt{P}/\beta\hbar$, and P is the total number of replicas or beads chosen for the discretized imaginary time path. Each bead represents the particle at a discrete position in imaginary time, so $x_{P+1} = x_1$ in order to guarantee that only paths which begin and end in the same place are considered for each configuration. This Hamiltonian effectively describes a collection of classical-like particles each acting under a potential $U(x)/P$ that are attached by harmonic oscillators to the adjacent particles in a "ring polymer" or "necklace". (Note that the notation here is for a single quantized particle in a one-dimensional potential, but the notation is readily generalized to more than one particle in three dimensions.) Molecular dynamics simulations performed using the Hamiltonian in equation 1 is Path Integral Molecular Dynamics (PIMD). PIMD provides a route to calculating quantum static equilibrium properties in a computationally feasible fashion, generally achieving converged results at P~30 replicas for a system such as liquid water at ambient temperature.[145] However, the PIMD Hamiltonian cannot

be used to estimate quantum dynamics and must be regarded only as a sampling tool for equilibrium statistics.[144,145] The PIMD approach is also still significantly slower than classical MD due to the increased complexity of the system being simulated, but largely scales the same way as does classical MD in terms of computational cost (i.e., there is simply a "cost" prefactor proportional to the value of *P*).

Centroid Molecular Dynamics (CMD), introduced by Cao and Voth nearly thirty years ago, provides a means to estimate certain quantum dynamical information from the discretized imaginary time path integral via the dynamics of the imaginary time path centroid moving in a classical-like fashion under the mean centroid force.[146,147] This method is motivated by Feynman's observation that the imaginary time path centroid is the most classical-like variable of a quantum system.[143,148] The original papers introducing the CMD concept replied largely on *ad hoc* arguments to justify the method, but two subsequent papers[149,150] in 1999 provided an exact formulation of centroid quantum dynamics and also a route to deriving CMD as an approximation to those exact dynamics. These latter two papers are sometimes not cited by authors when discussing CMD so the primary content of that work is briefly reviewed here for completeness.

Formally, a quasi-density operator (QDO) for the centroid density can be defined which can be used to formulate the exact dynamics of the imaginary time path centroids.[149] In one dimension, this QDO is equal to

$$\varphi(x_c, p_c) = \frac{\hbar}{2\pi} \int_{-\infty}^{\infty} d\zeta \int_{-\infty}^{\infty} d\eta e^{i\zeta(\hat{x}-x_c)+i\eta(\hat{p}-p_c)-\beta\hat{H}}, \qquad (5.2)$$

where x_c and p_c are centroid positions and momenta, $\widehat{H}$ is the system's Hamiltonian and $\beta = 1/k_B T$. The centroid distribution function can be obtained by tracing this operator

$$\rho(x_c, p_c) = Tr[\varphi(x_c, p_c)]. \tag{5.3}$$

Evaluation of this trace gives a classical-like form for the centroid distribution function which separates the position and momentum components when a Cartesian coordinate system is used, such that

$$\rho(x_c, p_c) = e^{-\frac{\beta p_c^2}{2m}} \rho(x_c) = e^{-\frac{\beta p_c^2}{2m}} e^{-\beta V_c(x_c)} \tag{5.4}$$

where $V_c(x_c)$ is the effective centroid quantum potential of mean force. Integrating over the centroid position and momentum variables thus yields the standard quantum partition function

$$Z = \int \int \frac{dx_c dp_c}{2\pi\hbar} \rho(x_c, p_c). \tag{5.5}$$

This partition function can be used to calculate the average of a physical observable corresponding to an operator $\hat{A}$ as

$$\langle \hat{A} \rangle = \frac{1}{Z} \int \int \frac{dx_c dp_c}{2\pi\hbar} \rho(x_c, p_c) \hat{A}(x_c, p_c; t), \tag{5.6}$$

where the time dependent centroid variable $A_c(x_c, p_c; t)$ is defined as

$$\hat{A}(x_c, p_c; t) = Tr\left[\hat{\varphi}(x_c, p_c) e^{\frac{i\widehat{H}t}{\hbar}} \hat{A} e^{-\frac{i\widehat{H}t}{\hbar}}\right] / \rho(x_c, p_c). \tag{5.7}$$

For the exact dynamics of the centroid variables, a normalized time dependent QDO can also be defined as

$$\hat{\delta}_c(t; x_c, p_c) = e^{-\frac{i\hat{H}'t}{\hbar}}\hat{\delta}_c(x_c, p_c)e^{\frac{i\hat{H}'t}{\hbar}} \tag{5.8}$$

$$\hat{\delta}_c(x_c, p_c) = \hat{\varphi}(x_c, p_c)/\rho(x_c, p_c). \tag{5.9}$$

where $\hat{H}'$ is a time-independent Hamiltonian upon which the system evolves ($\hat{H}' = \hat{H}$ in the usual equilibrium case). It is important to note that the QDO in this derivation is not a typical density operator: while it is Hermitian with non-negative diagonal elements, it is not positive-definite. This operator can be used to describe the exact dynamics of the path centroids; however this is not a useful approach for non-trivial systems for which the quantum Liouville equation cannot be solved.[149] Instead, various approximations to the QDO can be made which result in various forms of path integral based methods including linearized quantum dynamics, centroid Hamiltonian dynamics, and CMD.[150] In the case of CMD the centroid phase space variables are propagated quasi-classically by virtue of the following *ansatz:*

$$\hat{\delta}_c(t; x_c, p_c) \approx \hat{\delta}_c(x_c(t), p_c(t)) \tag{5.10}$$

$$m\dot{x}_c(t) = p_c(t);\ p_c(t) \approx F_c\big(\dot{x_c}(t)\big) = F_{CMD}(t) \tag{5.11}$$

The approximation made here assumes that the QDO is the same at $t = 0$ as at later times except for the placement of the centroids. This mean field-like assumption is reasonable for cases in which linear response theory approximates the dynamics of the system well, and for systems that have strong regression to equilibrium behavior. This perspective reveals that CMD is likely to be most

accurate for systems at equilibrium and in which coherent (purely quantum) aspects of the dynamics are not likely to have enough time to influence the system significantly before the correlations die out or de-cohere.

The key to CMD is thus to numerically calculate the effective equilibrium forces felt by the path centroids in one way or another. Sampling of the non-centroid imaginary time path integral modes at each centroid position determines an effective potential on which to propagate the centroids in a classical-like fashion. However, fully sampling these modes at each timestep in a simulation is usually very computationally expensive and so it is arguably this feature of CMD that has precluded its application to certain problems over the years. Numerical implementations of CMD instead attempt to adiabatically separate the centroid and non-centroid imaginary time path integral (Matsubara) modes, of which the zero-frequency mode is the centroid. Adiabatic separation is achieved by setting the fictitious masses of the non-centroid modes to be much lower than that of the centroid, which is set to the physical mass of the particle. Then, one attaches thermostats to the non-centroid modes to help more rapidly sample them. While the adiabatic approximation enables "on-the-fly" calculation of the centroid effective force, generating centroid trajectories in CMD still involves significant computational overhead.

As an alternative to CMD, Manolopoulos and co-workers subsequently introduced ring polymer molecular dynamics (RPMD) as another approximate quantum dynamics approach,[151–153] which shared certain key aspects of the spirit of CMD. In RPMD, the fictitious masses of the ring polymer beads are set to the physical mass of the particle, and each bead is evolved as a dynamical variable with PIMD. One then makes the *ad hoc* argument that the MD sampling time in PIMD (as measured by integrator timesteps) is related to the actual real time of the quantum dynamics. The RPMD approach thus removes the requirement that any sort of centroid force averaging be

carried out as in CMD. It should be noted that a subsequent analysis[154] showed that RPMD has no clear connection to real time quantum dynamics, but the methodology remains popular among users given its ease of use. In addition, an analysis of an approximate, but to date impractical form of quantum dynamics called "Matsubara Dynamics" has suggested that RPMD and CMD can be related to that approximation through further approximations.[155] Also of note is that RPMD is numerically faster than CMD, but the increase in speed is relatively small in comparison to classical MD, which is significantly faster than both.

While both RPMD and CMD can capture a range of quantum effects such as incoherent tunneling and zero-point quantization, they are not without their drawbacks. For example for certain potential energy functions and at low enough temperatures, CMD may exhibit a "curvature problem", which has a tendency to red shift certain vibrational frequencies for some systems,[156] This behavior can make CMD less suitable for spectrum prediction for certain systems, although this issue generally vanishes at room temperature or higher.[157] On the other hand, RPMD suffers from a spurious resonance problem, in which the centroid dynamics becomes coupled to the harmonic oscillations of the ring polymer, introducing artificial resonances into the spectra.[158] Thermostatted RPMD (TRPMD), in which Langevin thermostats are attached to the ring polymer internal modes, has been introduced as an *ad hoc* "fix" for this problem.[159]

Beyond these issues, both RPMD and CMD remain computationally relatively expensive due to the need to represent each physical atom in a simulation with dozens or more ring particle replicas (beads). Employing these methods on systems containing many thousands of atoms or more can be infeasible for all but the shortest simulations. For CMD, it has been found to be possible in some cases to generate accurate CMD dynamics with only a partial adiabatic separation of the internal ring polymer modes from the centroid mode.[160] Partially adiabatic CMD (PACMD)

simulations can handle larger timesteps as well, allowing them to reach similar levels of efficiency to similar to RPMD. In the case of RPMD, however, a clever ring polymer contraction scheme can be used to increase computational efficiency.[161] In this method, the short range interactions of the system can be treated using a full ring polymer, while longer range interactions can be approximated by interactions calculated across a smaller number of beads (somethings even just the centroid mode). In order to further bridge the gap between classical MD and PIMD-based methods, approaches have been proposed[162,163] to directly evolve the centroids of quantum particles along a learned centroid force field, but these methods have used pairwise tabulated potentials in their effective quantum force fields which can limit their accuracy in capturing the NQEs.

To quickly summarize, CMD is by now a venerable approach to estimating certain quantum effects in finite temperature systems, for better or for worse. After nearly 30 years it has largely stood the test of time as a valuable approximation. Yet, the demanding nature of the calculation of the effective centroid force (usually done "on the fly") has in some ways held back the method from wider use and applicability (e.g., in comparison to RPMD). To shed light on a path to overcome this central challenge – and to capitalize on the rapidly evolving developments in machine learning (ML) – in this paper we introduce Machine-Learned CMD (ML-CMD) and demonstrate its (arguably remarkable) features in increasing the computational speed and overall efficiency (time to solution) for CMD simulations. The method employs a deep neural network (DNN) trained on PIMD data to act as a force field which calculates the effective centroid forces based on configurations of the path centroids alone. This method retains most of the efficiency of a pairwise force field without its corresponding constraints. ML-CMD employs the DeePMD

kit,[110] which has been applied to both *ab initio* data as well as classical atomistic MD in the past to efficiently predict forces and energies of complicated systems.

DeePMD is a general use DNN method for learning atomic forces and energies based on two main components. The first is a descriptor network which converts the local environment, analogous to a neighbor list in classical MD, of a particle into translationally, rotationally, and permutationally invariant embedding. This network then passes these embedded features to a second fitting network which considers this environment to predict atomic contributions to energy or force.[164,165] The method can learn a force field for a completely generic representation of a system using mapped forces, including those from PIMD trajectories.[113] In the case of ML-CMD, the training dataset consists of a PIMD trajectory with forces projected to the centroids of the imaginary time paths using the mean square error in the forces as a loss function. The dataset is then used to learn the centroid forces directly, thus front-loading the work of deriving the effective centroid forces. Naturally, this approach benefits simulations of larger systems the most, but even for simple applications it can result in significantly faster results. We show later in this paper that ML-CMD can be applied to low temperature para-hydrogen as well as room temperature liquid water in order to calculate quantum time correlation functions to a great degree of accuracy, as well as significantly greater efficiency (between one and two orders of magnitude faster) against comparable path integral based methods. The ML-CMD models are also easy to train, require no more additional simulation than CMD, and can be deployed quickly enough to justify their use over CMD and RPMD even for simple systems with no simulation data readily available.

5.2 Methods

We trained two ML-CMD models to test the method's ability to capture static and dynamic properties of systems with significant NQEs. The first is the Silvera-Goldman model[166] of para-hydrogen at 14 K and a density of ρ=0.0235 Å^{-3}. A total of 180 particles were simulated, each corresponding to an entire H_2 molecule. This spherical approximation is justified as hydrogen is in the rotational ground state at the chosen state point. The second system contained 233 water molecules simulated using the qSPC/fw force field at 300 K and atmospheric pressure.[167] The datasets used for DeePMD were generated using i-PI and LAMMPS software packages.[100,168] Each system was simulated using normal mode PIMD (NMPIMD) with 32 replicas per particle for 250 picoseconds. A total of 8000 frames were taken from the final 200 picoseconds of each simulation, and the coordinates and forces were mapped to the centroid resolution.

The para-hydrogen model was trained using an embedding DNN with three hidden layers containing 10, 20 and 40 neurons, no timestep, and 46 nearest neighbors were considered as the local environment for the descriptor network. The fitting network was composed of three hidden layers of 240 neurons each, with a timestep. The learning rate schedule was exponential, with a starting rate of 5x10^{-3}, and ending rate of 1.76x10^{-7}, and 5000 decay steps. The model was trained for 500,000 iterations before validation and testing. The water model was trained using an identical model except for the following differences: The first is the number of neighbors considered for the local environment of each atom was increased to 60 hydrogens and 30 oxygens. Second, the model was trained for 100,000 iterations instead of 500,000. Additionally, the water training was batched with a batch size of 10. Both models were trained using DeePMD's se_e2_a descriptor which uses pairwise distances and embeds both radial and angular information about the system into the network. In order to obtain the most efficient models possible, the training length and local

environment size were systematically reduced until dynamical fidelity was impacted, thus resulting in a model which balances accuracy with speed. Input files for DeePMD training for both water and para-hydrogen are included in the Supporting Information.

The resulting force fields were tested against TRPMD and PACMD considering accuracy of static and dynamic properties as well as integration efficiency. All ML-CMD simulations were carried out in LAMMPS with the DeePMD force field add-on using a 0.5 fs timestep. For both RPMD and PACMD, each system was simulated using i-PI and LAMMPS using a 0.25 fs timestep for PACMD and a 0.5 fs timestep for TRPMD and 32 replicas per atom. PACMD frequencies were chosen according to the following equation

$$\Omega = P^{P(P-1)}/\beta\hbar, \tag{5.12}$$

which for the para-hydrogen system is 349 cm^{-1} and for the water system is 7481 cm^{-1}.[158] In order to accurately compute diffusion and time correlation functions for these systems, a parallel scheme developed by Pérez *et al*. was used.[169] A total of 64 individual frames were selected from an initial PIMD simulation to generate uncorrelated starting points as shown in Figure 5.1. Each starting point was then used to launch two 6 ps TRPMD, PACMD or ML-CMD trajectories, for a total of 128 simulations for each method. Velocity autocorrelation functions (VACFs) were computed from the final 5 ps of each of these trajectories and averaged to obtain a converged result. Self-diffusion constants were then obtained from the zero-frequency Fourier transform of the Kubo transformed VACF. It is noteworthy that the PIMD trajectory used to start each short simulation

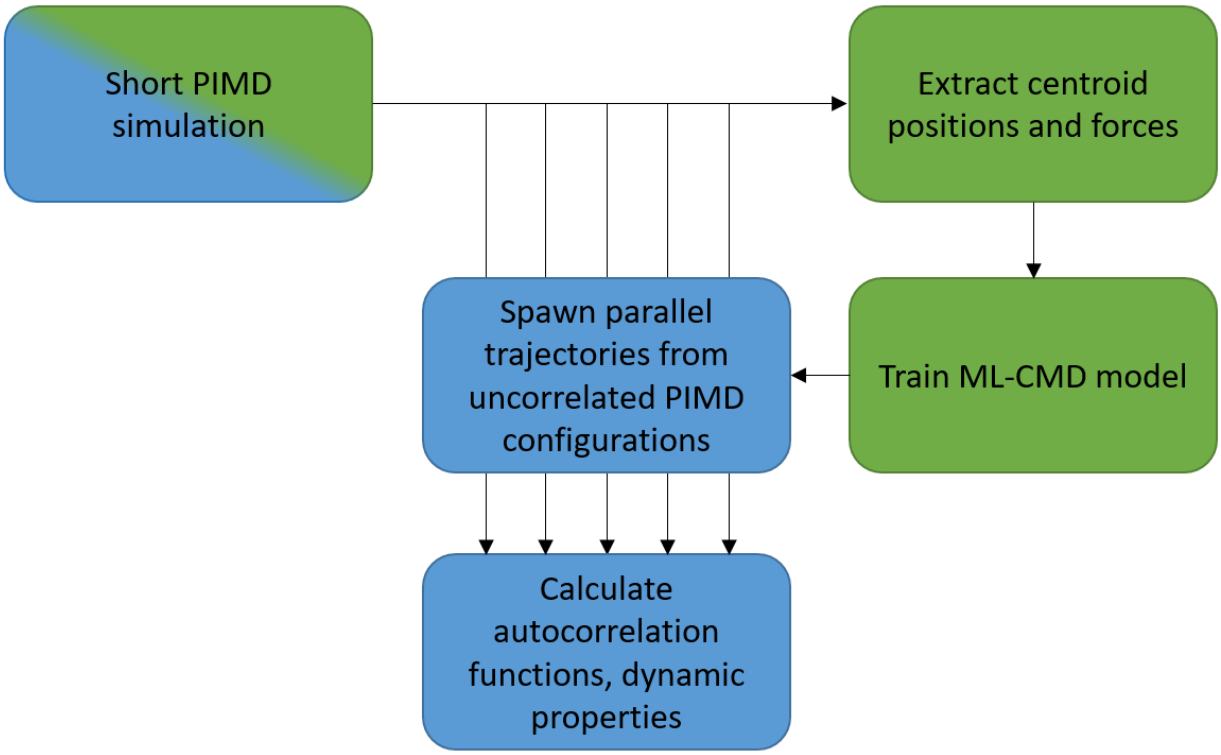

Figure 5.1. Workflow for calculating autocorrelation functions using path integral MD methods and for generating ML-CMD models. Blue panels correspond to necessary steps for CMD and RPMD using uncorrelated configurations. Green panels are additional necessary steps for creating a ML-CMD model. The initial PIMD simulation step is shared among both, eliminating the cost of data collection for training ML-CMD force fields when calculating autocorrelation functions.

can double as a dataset for ML-CMD. In cases where such a scheme is used, one saves even more time as this simulation performs two critical steps instead of just one.

When measuring efficiency, all methods were tested using the same computational environment. Each test simulation was run on 32 cores of a 40 core Cascade-Lake compute node to allow for one core per replica in the full PI simulations. All simulations were carried out in LAMMPS. The total time used to calculate autocorrelation functions was calculated as the sum of all required simulations and training ("time to solution"). For ML-CMD this includes the 250 ps PIMD simulation used to generate the dataset which was also used as the starting points for the autocorrelation function calculations, the DeePMD network training time, and the total time to run the 768 total ps of simulation required for the autocorrelation functions. For PACMD and TRPMD, this includes the short PIMD trajectory required to generate starting configurations and the time to simulate the 768 ps of data for the autocorrelation functions.

5.3 Results and Discussion

Both ML-CMD para-hydrogen and water results show excellent agreement with TRPMD and PACMD. Figures 5.2 and 5.3 show radial distribution functions for all three models as well as PIMD. Of note are the water peaks corresponding to the O-H bond and H-O-H angles, which ML-CMD captures very accurately despite there being no explicit bonded or angular interactions in the force field. It is, however, not surprising that the static properties of ML-CMD align with full PI methods as PIMD is well known to be ideal for sampling the equilibrium properties positions of

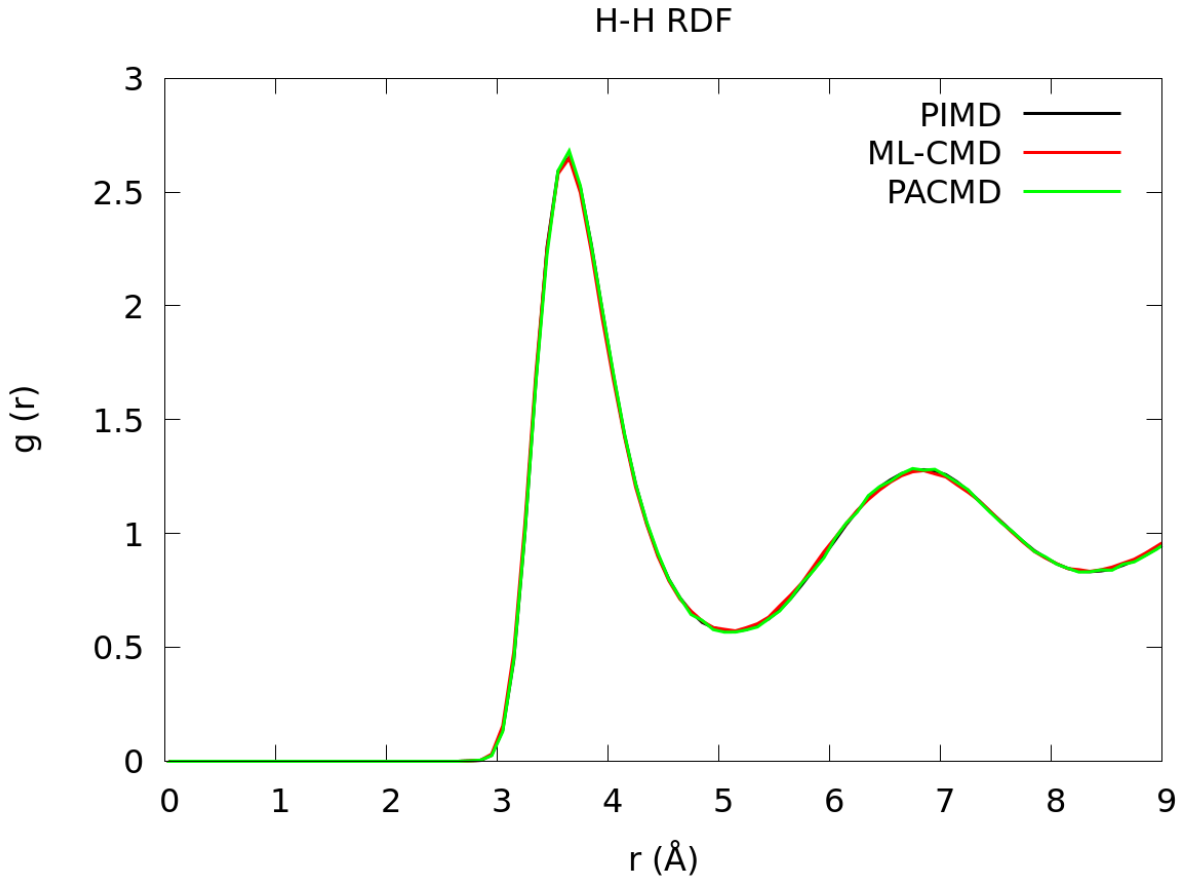

Figure 5.2. Plot of the radial distribution function (g(r)) for 14 K para-hydrogen. PIMD, ML-CMD, and PACMD are compared. As the Silvera-Goldman potential represents one hydrogen molecule as a single particle no peak corresponding to the H-H bond is present.

quantum particles. In addition, a previous attempt at calculating effective centroid forces from PIMD trajectories for CMD have shown similar levels of accuracy for para-hydrogen.[162]

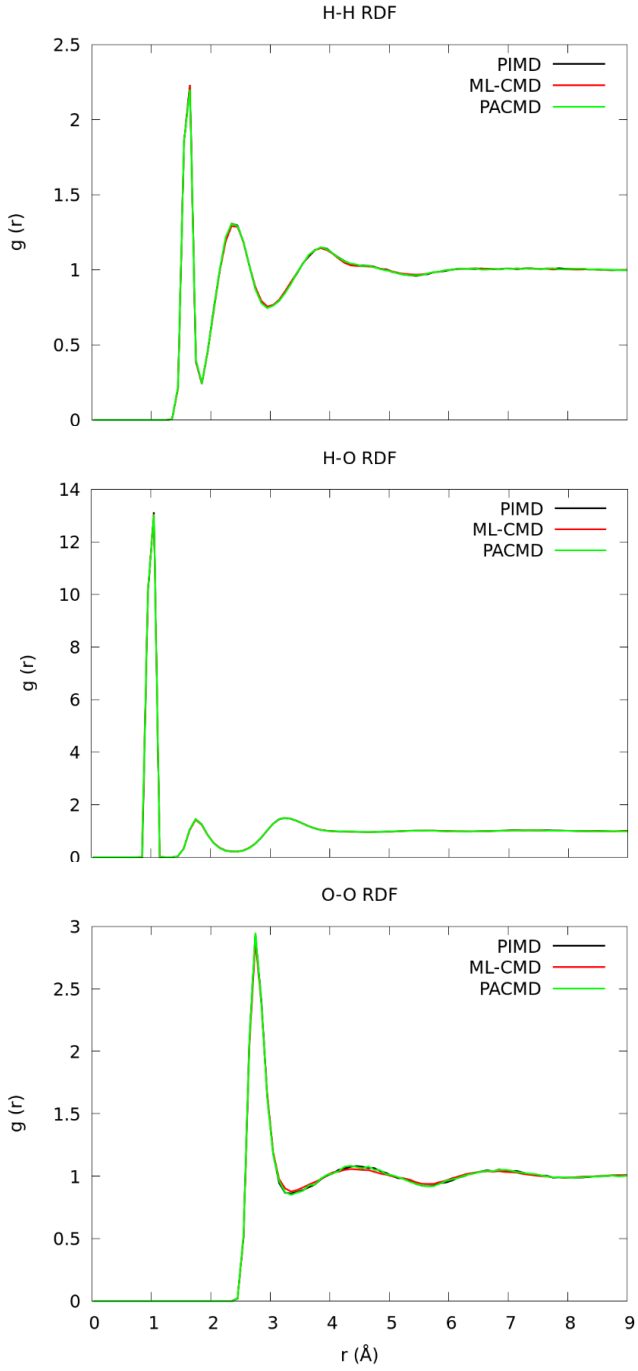

Figure 5.3. Plot of the radial distribution function (g(r)) for Hydrogen-Hydrogen (top), Hydrogen-Oxygen (middle) and Oxygen-Oxygen (bottom) of 300 K water. PIMD, ML-CMD, PACMD, and TRPMD are compared.

The approximate quantum dynamics of the ML-CMD models are also well in line with the PACMD and TRPMD results. Figure 5.4 shows the VACF for all three types of simulation for para-hydrogen as well as CMD simulations with greater adiabatic separation. ML-CMD generally agrees better with PACMD than TRPMD, which is expected as the latter method is not designed to approximate centroid dynamics. CMD simulations in which the adiabatic separation is not

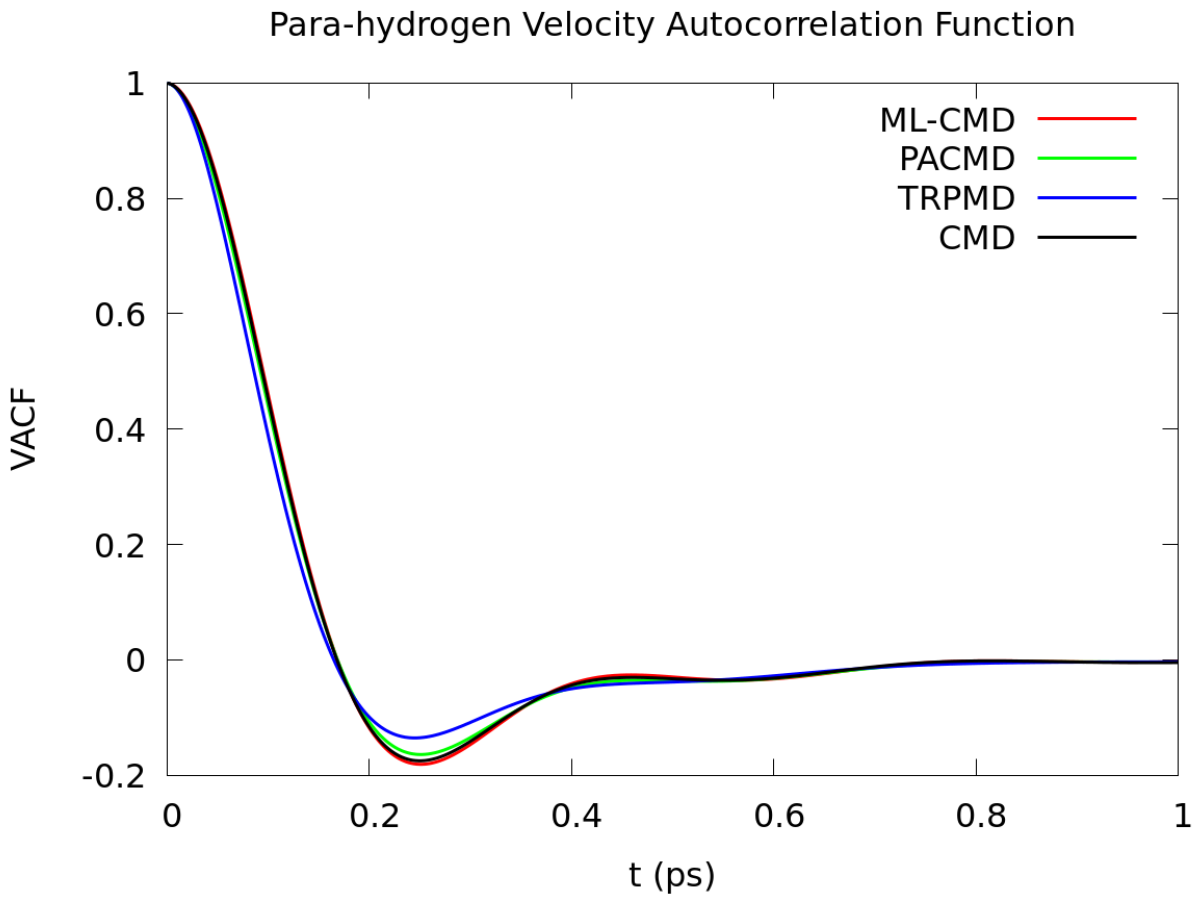

Figure 5.4. A plot of the normalized velocity autocorrelation functions (VACFs) for 14 K para-hydrogen using ML-CMD, PACMD, CMD, and TRPMD simulations. Each VACF was averaged over 128 individual trajectories.

partial show even better agreement with ML-CMD. This suggests that for systems with significant NQEs ML-CMD is not only faster than PA-CMD, but more accurate as well.

Figure 5.5 shows the velocity autocorrelation functions of water. In this case, ML-CMD matches PACMD and TRPMD nearly perfectly. The fact that RPMD and PACMD converge better for room temperature water than for 14 K para-hydrogen is not surprising. RPMD and CMD are known to converge as the system approaches the classical limit, e.g., heavier nuclei or higher temperatures.[160]

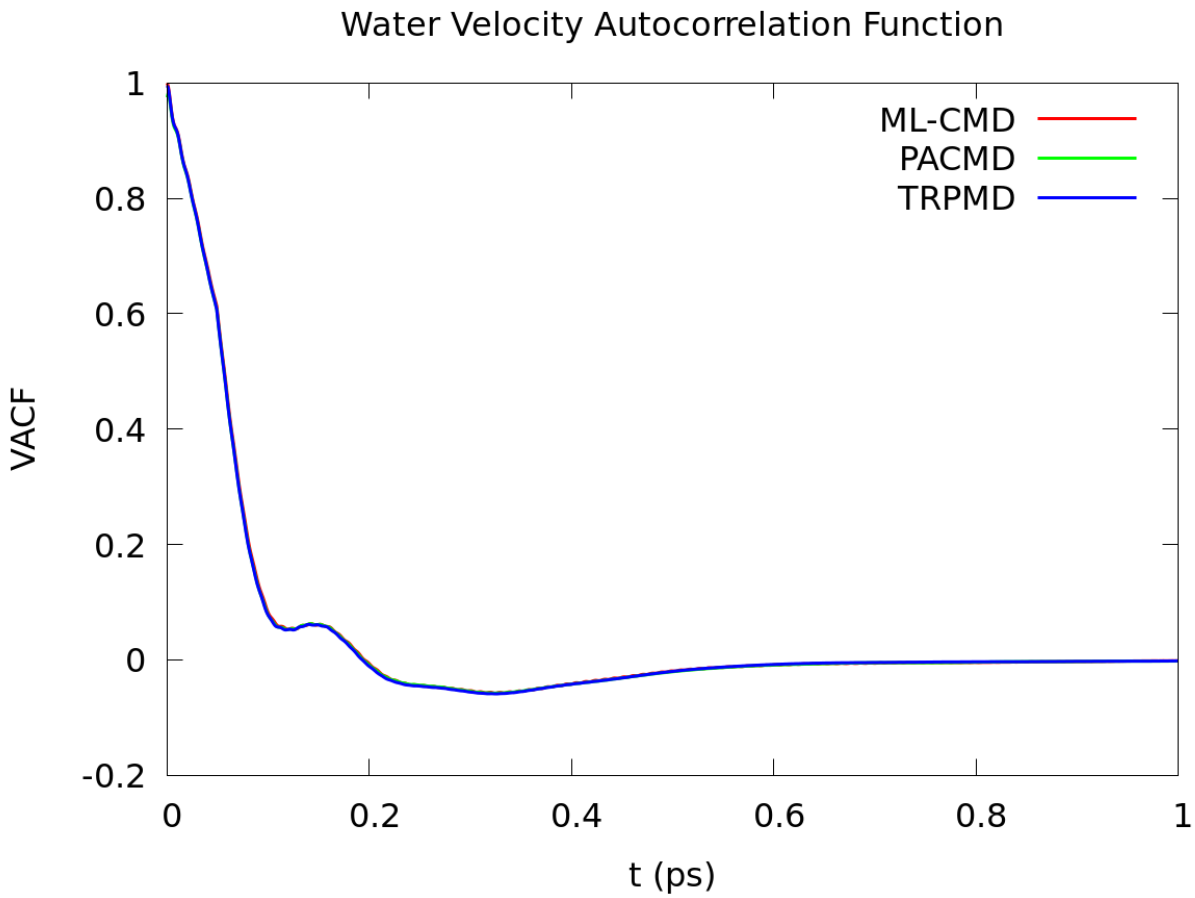

Figure 5.5. A plot of the normalized velocity autocorrelation functions (VACFs) for 300 K using ML-CMD, PACMD and TRPMD simulations. Each VACF was averaged over 128 individual trajectories.

Table 5.1 next shows the calculated self-diffusion constants for both water and para-hydrogen for all three types of simulation. ML-CMD shows excellent agreement in both cases; it

appears to be slightly more diffusive although this is well within the margin of error. The neural network architecture of ML-CMD's force field seems to directly contribute to the dynamical accuracy of the method. A previous attempt for accelerating CMD using force matching of the centroid forces resulted in less accurate values for self-diffusion at lower temperatures.[162]

Table 5.1. Self-diffusion constants for para-hydrogen and liquid water. All values were calculated using the zero frequency Fourier transform of the Kubo transformed velocity autocorrelation function.

System	**Para-hydrogen 14 K (Å^2/ps)**	**Water 300 K (Å^2/ps)**
ML-CMD	0.30 ± 0.03	0.32 ± 0.03
PACMD	0.29 ± 0.03	0.31 ± 0.03
TRPMD	0.28 ± 0.03	0.31 ± 0.04

An important factor in the training and validation of DNNs is the amount of data required to achieve a converged model. The ratio of training data to extrapolation informs not only the feasibility of the method, but also its quality. We trained 3 additional para-hydrogen models using 20, 60, and 100 ps of PIMD reference data to test this. Table 5.2 shows diffusion coefficients for each of these models. Figure 5.6 shows VACFs comparing the models to the model trained on the full 200 ps reference dataset.

Table 5.2. Diffusion coefficients for ML-CMD para-hydrogen trained on datasets of varying length compared to the base model, which was trained on 200 ps of PIMD simulation.

Reference simulation length (ps)	**Diffusion Coefficient ($Å^2$/ps)**
20	0.31 ± 0.03
60	0.30 ± 0.03
100	0.31 ± 0.03
200	0.30 ± 0.03

These results show that the method works with much less data than the full 200 ps trajectory. The diffusion constant values are well within a standard error of each other, and the autocorrelation functions match perfectly, even with 10 percent of the total training data used. Future applications of ML-CMD are thus likely to be deployed even more quickly and efficiently than the ones presented here.

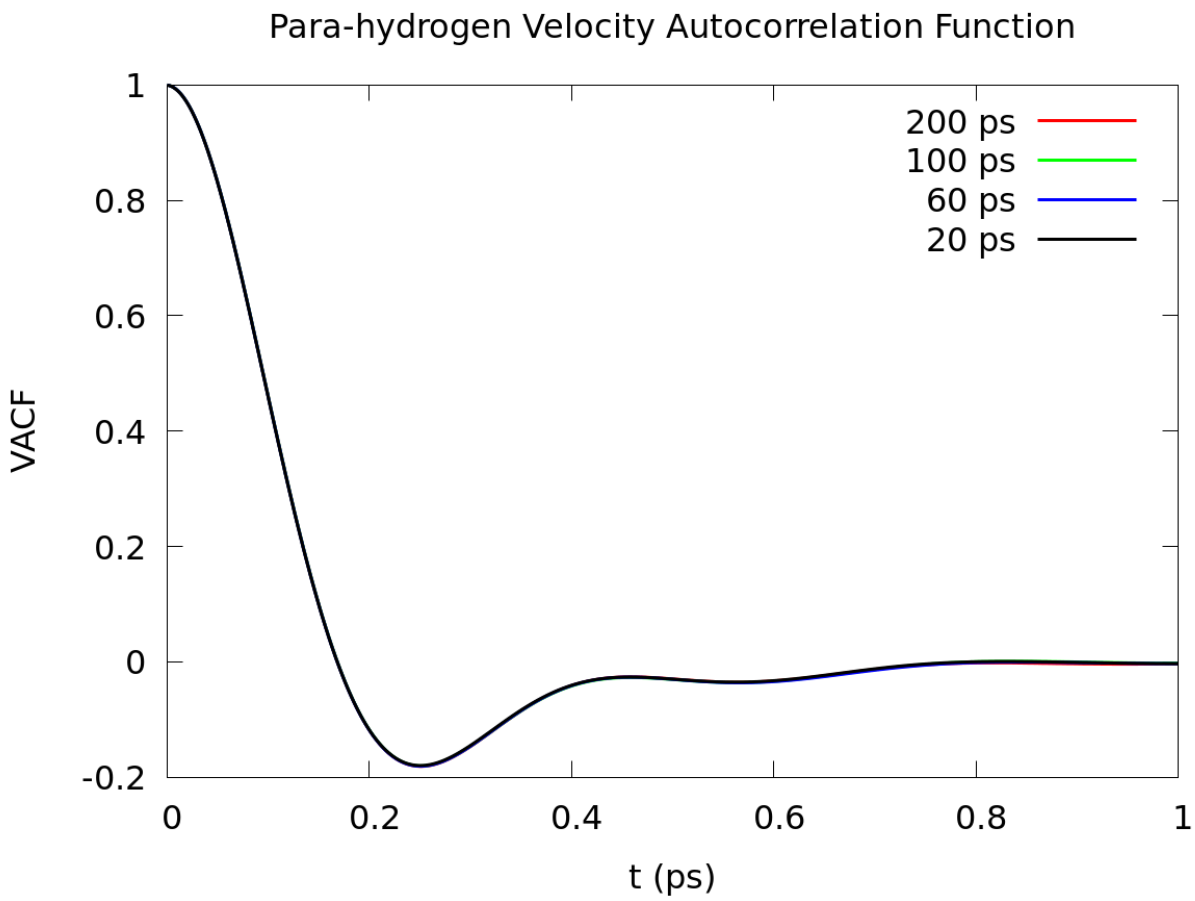

Figure 5.6. A plot of the normalized velocity autocorrelation functions (VACFs) for 14 K para-hydrogen using several trained ML-CMD models. Each model was trained with a different amount of reference PIMD data and VACFs were calculated in the same manner as the preceding figures.

Beyond accurate dynamics, the most important aspect of ML-CMD is its speed over methods such as full CMD, PACMD, RPMD, and TRPMD. In order to be useful as a replacement for these methods, it must be faster than them even when considering the time to collect the PIMD data and train the DNN force field. Table 5.3 shows integration speeds for the ML-CMD, PACMD, and TRPMD as well as the total time required to collect converged velocity autocorrelation functions. ML-CMD performs over 70 times faster than PACMD for para-hydrogen and over 20

times faster for water. Additionally, the time to data collection is shorter than both PI-based methods by a considerable amount, even when considering the amount of excess PIMD data used to train the models in the present case. These results are encouraging, as the correlation function calculations require only 768 total picoseconds of integration, less than four times the total amount of simulation time used to train the model in the first place. In cases where more or larger simulations are required such as calculating long time correlation functions for complex free energy surfaces, ML-CMD becomes even more efficient, as the initial simulations and model training only need to be performed once. Furthermore, each ML-CMD model can be shared for use in multiple simulations either individually or through a publicly available repository of models which can be contributed by anyone using the method.

Table 5.3. Integration speed and velocity autocorrelation function (VACF) calculation times for ML-CMD, PACMD and TRPMD. All measurements performed using 32 cascade-lake cores. VACF calculation time for PACMD and ML-CMD includes an initial PIMD simulation followed by 768 ps of integration. For ML-CMD, this also includes DeepMD neural network training time.

System	**Para-hydrogen speed (ps/hr)**	**Para-hydrogen VACF calculation time (hrs)**	**Water speed (ps/hr)**	**Water VACF calculation time (hrs)**
ML-CMD	734.4	21	144.5	49
PACMD	9.8	90	6.1	137
TRPMD	17.9	55	10.7	84

It is also worth discussing the flexibility of the machine learning algorithm chosen for ML-CMD. DeePMD is simple to deploy and yields excellent performance, but it is one of many similar DNN based methods which can predict forces and energies in molecular systems. One of the most important aspects of any of these methods is how they encode translational, rotational and

permutational invariances into the input features of the model. This allows for the network to work in scalar space with the same symmetries present in the real chemical environment. Recently, several models have been developed which replace these invariant networks with equivariant ones which can directly encode vectored information such as forces. These models have been shown to require far less data than symmetry invariant models, and can produce better results with up to one thousandth the number of training examples.[111,112] Applying such a method would naturally speed up the training of ML-CMD models and should be considered for future study.

5.4 Conclusions

In this work, we have presented ML-CMD, a machine learning approach for the calculation of the effective centroid potential in CMD simulations. Over the past 30 years, regular CMD has been limited by the need to repeatedly calculate the centroid effective force "on-the-fly" through adiabatic separation of the centroid and non-centroid imaginary time path integral modes. Such calculations represent a significant computational overhead for larger systems. Instead, in the present work we shown that by training a neural network to first learn the centroid effective potential from a PIMD simulation, one can greatly increase the efficiency and time to solution of a CMD simulation without sacrificing accuracy.

We have demonstrated that for both room temperature water and 14 K para-hydrogen ML-CMD provides highly accurate results which closely match both the static and dynamic properties given by full PACMD and TRPMD simulations. The ML-CMD simulations are also many times faster than either of the latter approaches, thereby extending the range of systems for which CMD will be applicable. While the initial PIMD simulations and DNN training steps somewhat constrain

the overall speed at which the ML-CMD models can be deployed, the PIMD simulations are also necessary to provide the initial conditions for trajectories in both PACMD and TRPMD.

This paper provides yet another example where machine learning promises to transform the field of molecular simulation, in this case by making CMD simulations feasible and accurate for a wider range of systems. Likewise, one can expect that future advances in machine learning, e.g., to better treat heterogeneous systems, rare events, etc., will also provide clear benefit to the ML-CMD approach developed in this work.

Chapter 6: Conclusions

When developing a bottom-up CG model there are always two considerations one must keep in mind. The first, which comes quite naturally, is to assess the accuracy of the models being created with respect to the reference dataset. Many benchmarks are commonplace, such as lipid membrane bending modulus, radial distribution functions, and qualitative analyses of model behavior. However, it is important to keep in mind that outside the limit of infinite sampling and perfect basis sets, which will never be accomplished in real life, all CG methods have their own advantages and disadvantages which cannot be escaped. Thus, recognizing the strengths and weaknesses of each CG method is important to the creation of usable CG models. Unfortunately, this information is not always readily available, and often the models which do not work can provide just as much guidance as those that do. This, in tandem with the fact that bottom-up CG methods are most useful when anyone without expertise can use them to create their own models, points to a deficiency in how information about CG modeling is disseminated. Best practices, tutorials and lessons learned should be included along with any CG modeling code base to maximize their usefulness.

The second consideration when designing CG models is assessing the practicality of the model itself. Coarse-graining is a means to an end- to simulate larger molecular systems for longer periods of time. To be useful a CG method must produce models which are efficient enough to justify their own existence including the time and effort required to generate them. Methods which involve weeks of optimization and multiple methodologies will never be more attractive to users than top-down CG force fields which one can get started with in mere minutes, regardless of how much more accurate they are. Additionally, models which integrate faster will be more attractive than slower ones. Integration speed is a combination of the overall number of CG sites the model

incorporates, the maximum timestep which still allows the model to run stably and the integration speed of the force field. Typically, models become faster as the CG resolution becomes lower, which is often though not always at odds with accuracy. Methods which are easy to use and deliver results in a short amount of time are the most practical.

With these considerations in mind, a simple framework for evaluating CG models can be made, where a model is classified according to a combination of its accuracy and practicality. There are four categories which any given model can be loosely put into. The first is the ideal case, where a model is accurate enough to use, and very practical to start using. These accurate-practical models should be strived for in all cases. The next case, inaccurate-impractical, are the worst CG models, which are very difficult to recommend using. While models of simple liquids and other common test systems rarely fall into this category, one unfortunately finds these types of models in many cases where the system at play is very complex or ill-suited to CG. The last two cases, accurate-impractical and inaccurate-practical, lie somewhere in between, and represent the pragmatic reality of the situation: where accuracy and practicality are inversely correlated. While one can often systematically improve the accuracy of the model, this usually involves more reference data collection, or the application of more complicated functional forms for the CG Hamiltonian which slow down dynamics and thus decreases practicality.

How does one decide whether practicality or accuracy are more important when balancing the two? Method and developers tend to highlight whichever element their models happen to be best at, which gives potential users a good heuristic to follow to find a method or model which works for their use case. In most cases though, issues of practicality are often glossed over in the literature. No one wants to say that their model requires extensive hyperparameter sweeping, or that the procedure used to build their model takes over a month to perform. The reason for this is

simple: no one would use such a model. This hints that practicality is far more important than most method developers give it credit for. While more consistency with underlying atomistic physics is always better, there is a large amount of tolerance for more phenomenological models which "get the job done" especially when learning to use these models is simple and they integrate quickly. This should come as no surprise to anyone familiar with MD force fields. Classical MD struggles greatly to simulate water due to its non-classical behavior and reactivity, yet most biophysical simulations utilize these inaccurate water models to represent most of the atoms in their system. It would be extremely impractical to suggest that a biophysical modeler go to the level of quantum mechanics to simulate water more accurately, and therefore it is only done when these quantum effects are critical.

The difference between AA-MD water models and CG-MD models is that the water models have been analyzed thoroughly, and their limitations are well known.[170–172] To that end, bottom-up CG method developers should not shy away from discussing the limitations of their methods. Any CG model, top-down or bottom-up, will necessarily have limitations, this is easily demonstrated by the ubiquity of the transferability and representability problems in CG models, but it is not necessary. All models inherently cannot capture the reality they emulate perfectly: even the most advanced quantum mechanical methods available today cannot analytically solve the Schrödinger equation for complex systems, so why hide that a CG model is less accurate in emulating AA-MD properties than one would like?

In chapter 2, I summarize Martini 3.0's difficulties with capturing thermodynamic transferability and macroscopic material properties of lipid bilayers. Analyses such as this one are not unique, and much has been said about top-down CG models in relation to the AA models they emulate. Naturally, the same scrutiny ought to be paid to bottom-up models. This type of criticism

will not only spur future efforts towards better models in the future but will also allow the CG modeling community to gauge the inherent limitations of bottom-up CG systems such as lipids, which as demonstrated in Chapter 3, fail in their own ways. Beyond this, combinations of analyses done on top-down and bottom-up CG models will provide insight into the capabilities of CG modeling in general. For example, the results of both chapter 2 and 3 taken together suggest that CG lipids as a rule do not exhibit transferability without explicit incorporation of temperature dependent potentials.

Beyond a larger discussion around limitations, it is also helpful to have a selection of previous models to either build off or to use as benchmarks to produce new models. In principle, generating large suites of bottom-up models should not be difficult, as it is possible to systematically create large numbers of models at varying resolutions and underlying optimization methods. These models could then be categorized, benchmarked, and published for use or reference by the rest of the CG modeling community. Unfortunately, there seems to be far more focus on generating new and more advanced methodologies than on utilizing existing methods. While method papers typically include at least a small set of test models which suggest their usefulness, these models are often far simpler than real life CG use cases, which often contain multiple species, large biomolecules, or anisotropy which can make models behave in unexpected ways.

Top-down CG models often do a much better job of this, partially due to how easy they can be to use. Martini models for practically any class of system have been used for published computational experiments, and the force field itself is the equivalent of a near complete library of bottom-up models. Bottom-up CG modelers should take a page from their book and focus more on applications which inform future models. This will also have the benefit of advertising these

methods as viable options that have been vetted and are reliable. It is often argued that bottom-up CG models should be more accurate than top-down ones, at least when compared to atomistic molecular dynamics. This is a fair claim to make, as a CG model is essentially an alternative to AA-MD. In addition, claims of accuracy are appealing as most bottom-up methods have been mathematically analyzed and guarantee a rigorous correspondence between resolutions, at least in principle. This also helps to justify their use, as they go beyond phenomenological models and incorporate physics from AA-MD. However, as has been demonstrated time and again, mathematical consistency with the AA resolution is only guaranteed in certain limiting cases, which are never truly met in practical use cases. In fact, if these conditions could be met, there would be no need for the models in the first place. Therefore, these justifications alone cannot convince CG modelers to use bottom-up methods, while real examples of functional bottom-up models can.

Returning to the accuracy-practicality framework described earlier, each chapter in this thesis sheds some light on how models in this category can be produced, critiqued, and improved. In Chapter 2, the Martini 3.0 force field's accuracy is analyzed, and found to be no better than previous Martini models. Despite this, Martini remains extremely popular. This can be largely attributed to the amount of work that has been done to ensure that Martini is as easy to use as possible. It is a robust force field that has been used to simulate systems of vastly higher complexity than even the most powerful bottom-up methodologies. It is no more difficult to run a Martini simulation than it is to run the corresponding AA system, and it runs much faster, especially when using the implicit solvent Dry Martini models. The most straightforward simulations described in this work were all Martini simulations- even bottom-up modelers cannot help but praise its practicality with this in mind, Martini falls into the inaccurate-practical category. While it may not

capture structural and thermodynamic properties that other models have the potential to, the ubiquity of Martini models reaffirms the importance of practical CG models.

Chapter 3 brings to light the difficulty in parameterizing merely stable lipid bilayers from atomistic data. The hydrophobic effect from the bulk solvent and the radial anisotropy of lipid bilayers in general both pose challenges that are difficult to address, even with a high amount of expertise. Molecular modelers who utilize CG models should not be expected to follow the procedures outlined therein, and by that metric even VS-REM and TT-VS-REM, which outperformed the implicit solvent and force matching methods, fail as models due to their impracticality. The VS-REM and TT-VS-REM models capture features of the corresponding AA models which Martini cannot. Namely, the RDFs, bending modulus, and entropy enthalpy decompositions. At the same time, poration of the bilayer at long simulation lengths and inability to capture the membrane self-assembly which lower resolution models can make it difficult to recommend as accurate. I believe these models for the most part fall into the inaccurate-impractical category for those reasons. Again, the real determination for whether these models are useful is their practicality. TT-VS-REM came about as a culmination of a variety of advanced CG methodologies to a system which was difficult to coarse-grain, but this process was largely based in discovery. At each step, the next model was hypothesized to fix accuracy issues. Many of these issues were indeed solved in the end, but the process for doing so may not be the same for other systems. Virtual solvent particles may be much harder to apply to protein systems, and it is also unclear how to resolve virtual solvent particles in situations where the association of other biomolecules to a lipid bilayer. Beyond this, methods for generating similar models for other lipids at other resolutions look very different, especially with regards to REM optimization. It is almost impossible to conceive that someone who is not a specialist in CG methodology would be able to

come up with these procedures on the fly and create a model that was to their liking without giving up and simply using AA-MD.

Chapter 4 introduces machine learning as a solution for CG accuracy problems. DNNs with their immense regressive capabilities are a hot topic in CG method development and for good reason. These force fields break out of the pairwise limitation which most CG models are constrained to and are thus able to accurately reproduce structural features of water, a notoriously difficult system, with very little tuning. However, DNN methods remain largely impractical today due to their slow integration speed. This is a natural consequence of incorporating many-body interactions into a CG model. While this firmly places the models in this chapter in the accurate-impractical category, the equivariant neural networks applied to the system have addressed the other problem endemic to machine learning for CG force fields: their data requirements. While other ML based CG force fields require many microseconds of data to learn a force field for simple systems such as alanine dipeptide, just a fraction of a nanosecond of sampling can generate a highly accurate water model if force equivariance is accounted for explicitly in the particle-wise embedding scheme.

Chapter 5 introduces Machine Learned Centroid Molecular Dynamics (ML-CMD), a second ML based method for "coarse-graining" the path integral representation of an atom down to a single site. This method is largely similar to that of Chapter 4, and even lacks the equivariant embedding scheme which aids in making those models more practical. Despite this, ML-CMD is both an accurate and highly practical model. The reason for this is simple: while DNNs are much slower than pairwise force fields for AA-MD, they are much faster than most methods which incorporate quantum mechanics, including path integral based methods. While these methods do not involve any electronic structure calculations, the need to represent each atom as a ring of

dozens of particles means that the level of "coarse-graining" achieved by ML-CMD is much higher than the other models presented. Additionally, the forces on the path centroids in traditional centroid molecular dynamics (CMD) are calculated via repeated sampling of the non-centroid modes which are removed. Lastly, the best practice in path integral simulations is to perform a short path integral molecular dynamics (PIMD) simulation from which to draw uncorrelated initial configurations and velocities for additional dynamical simulations (see Figure 4.1). This initial simulation is identical to one used to train the ML-CMD force field, which cuts out a substantial amount of extra time which is normally required for typical CG model generation. This combination of factors makes ML-CMD an ideal accurate-practical CG model.

The results of these models raise another important aspect of CG model development that often goes understudied: the CG mapping scheme. Results from Chapter 3 suggest that higher resolution lipid models struggle to capture behaviors that models mapped to a lower resolution can. For the most part, CG mapping is done via chemical intuition rather than algorithmic assignment. For example, in CG phospholipids, it is desirable to assign the phosphate group to its own site. This allows for the calculation of properties of height fluctuation spectrums that correspond perfectly with atomistic data. This is because the center of mass of the phosphate bead coincides with the location of the phosphorus atom. It also makes sense from a chemical perspective: the negative charge in a phospholipid is localized within the phosphate group, and thus this charge can be incorporated into the potential of the phosphate bead alone. In other cases, the choice is not so clear. As an example, Martini suffers from degenerate mappings of lipid tails. DMPC lipids and DLPC lipids are treated identically within the Martini force field, even though DMPC tails contain two more carbons.

There are several methods that have been developed which show promise in systematic mapping assignment[173–175] which range in their applicability. Essential dynamics coarse-graining (ED-CG) utilizes sequence-based clustering to optimally assign mappings but is limited in scope to large sequence-based molecules (namely proteins). It also relies on the user inputting the total number of desired CG sites and as such it cannot be used to select the optimal CG resolution without sweeping over the number of CG sites and observing the level of fluctuations that are captured by each one.[175] This method has shown use in the generation of models of large proteins, where it is able to identify individual protein domains and motifs in many cases and assign them individual CG sites. A key limitation of ED-CG is its sequence-based mapping assignment. Each bead is composed of a continuous sequence of amino acids which means that domains that are formed from more than one segment of discontinuous sequences are often poorly captured. At the same time, this ensures that models resulting from these mappings are capable of unfolding.

Supervised ML has also shown promise for CG mapping. Autoencoders are DNNs which compress input data down to a smaller bottleneck before expanding this feature back to a larger resolution. If the network is trained to reproduce input data, this can be exploited for the simultaneous generation of CG mappings and force fields. The first half of the network, called the encoder, can be thought of as a mapping scheme which transforms atomistic data down to a lower CG resolution. This network will have powerful regressive capacity and will naturally learn a mapping which captures as much of the AA information as possible. The last half of the network, called the decoder, can return these CG coordinates back to the AA resolution, and can be used as a backmapping function.[173] From there, the CG distributions can be used to match an effective CG Hamiltonian. This method also relies on user chosen mapping resolutions, as the width of the final layer of the encoder is fixed. In addition, the mapped CG coordinates for a generic autoencoder

would have no physical meaning, and only mean something in the context of the neural network itself. Therefore, care must be taken to ensure that the result of the encoding half of the network exists within Cartesian space for both the MD engine to interpret it, as well as for the mapping to allow for CG consistency with the AA resolution.

A common theme here is that choosing the CG resolution is a difficult task to perform, and it is not a trivial task to develop an algorithm to perform this. There is the obvious consideration that speed is directly related to the resolution of the model, and thus the selection of the resolution can be decided by the minimum required integration speed for the model to serve its purpose. However, this is muddied by systems where the relationship between accuracy and resolution is not clear, such as implicit solvent lipids. There is also an inherent difficulty in comparing CG models at different resolutions. One can often select a series of calculable observables one wishes to compare, such as bending moduli, which are not inherently dependent on the CG resolution. There may be other desirable observables which cannot be separated from the resolution of the model itself. For example, higher resolution models may capture the entropy or root-mean-square fluctuations of a model, but this is in part due to the nature of these observables itself. A single site representation of a molecule will necessarily have no fluctuations or conformational entropy and is therefore not necessarily the best metric. One might instead choose qualitative features of the model, such as self-assembly properties, as a metric of quality, but these are difficult to quantitatively measure in the first place.

On the topic of machine learning, the buzz surrounding neural network force fields has been largely hampered by their slow integration speed. This is often addressed via “batched” simulations, where the same network can be used as a force field for multiple parallel simulations, just as neural network training samples can be batched together to speed up training. This is useful

for calculating free energy landscapes but is not particularly helpful when the goal of a CG simulation is to see long-time processes such as macromolecular assembly. Despite this, I have a very positive outlook about machine learning in the field of CG modeling. The accuracy gains from neural networks are unparalleled, and the fact that machine learning is so popular across many fields of science lead me to believe that there will be future breakthroughs which may make neural networks better suited to the goal of accelerating molecular simulations. There is also much potential for improved CG methods which employ ML to generate accurate non-neural network-based force fields which have no penalty to their integration speed. Given how flexible machine learning has proven to be, it is likely that there are many undiscovered applications in CG which will be discovered in the near future, which can only benefit the field.

References

(1) Chandler, D.; Percus, J. K. Introduction to Modern Statistical Mechanics. *Phys. Today* **1988**, *41* (12), 114–118. https://doi.org/10.1063/1.2811680.

(2) Harayama, T.; Riezman, H. Understanding the Diversity of Membrane Lipid Composition. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (5), 281–296. https://doi.org/10.1038/nrm.2017.138.

(3) Sunshine, H.; Iruela-Arispe, M. L. Membrane Lipids and Cell Signaling. *Curr. Opin. Lipidol.* **2017**, *28* (5), 408–413. https://doi.org/10.1097/MOL.0000000000000443.

(4) Yang, C.; Svitkina, T. Filopodia Initiation. *Cell Adh. Migr.* **2011**, *5* (5), 402–408. https://doi.org/10.4161/cam.5.5.16971.

(5) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

(6) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069–4074. https://doi.org/10.1063/1.449071.

(7) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45* (14), 1196–1199. https://doi.org/10.1103/PhysRevLett.45.1196.

(8) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab Initio Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc. Natl. Acad. Sci.* **2005**, *102* (19), 6654–6659. https://doi.org/10.1073/pnas.0500193102.

(9) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55* (22), 2471–2474. https://doi.org/10.1103/PhysRevLett.55.2471.

(10) Harrison, R. L.; Granja, C.; Leroy, C. Introduction to Monte Carlo Simulation; 2010; pp 17–21. https://doi.org/10.1063/1.3295638.

(11) Manavalan, B.; Kuwajima, K.; Lee, J. PFDB: A Standardized Protein Folding Database with Temperature Correction. *Sci. Rep.* **2019**, *9* (1), 1588. https://doi.org/10.1038/s41598-018-36992-y.

(12) Jung, J.; Nishima, W.; Daniels, M.; Bascom, G.; Kobayashi, C.; Adedoyin, A.; Wall, M.; Lappala, A.; Phillips, D.; Fischer, W.; Tung, C.; Schlick, T.; Sugita, Y.; Sanbonmatsu, K. Y. Scaling Molecular Dynamics beyond 100,000 Processor Cores for Large-scale Biophysical Simulations. *J. Comput. Chem.* **2019**, *40* (21), 1919–1930. https://doi.org/10.1002/jcc.25840.

(13) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.;

Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51* (7), 91–97. https://doi.org/10.1145/1364782.1364802.

(14) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *J. Chem. Phys.* **2019**, *151* (7), 070902. https://doi.org/10.1063/1.5109531.

(15) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18* (10), 5759–5791. https://doi.org/10.1021/acs.jctc.2c00643.

(16) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

(17) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824. https://doi.org/10.1021/jp071097f.

(18) Español, P. Statistical Mechanics of Coarse-Graining; 2004; pp 69–115. https://doi.org/10.1007/978-3-540-39895-0_3.

(19) Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42* (1), 73–93. https://doi.org/10.1146/annurev-biophys-083012-130348.

(20) Stieffenhofer, M.; Bereau, T.; Wand, M. Adversarial Reverse Mapping of Condensed-Phase Molecular Structures: Chemical Transferability. *APL Mater.* **2021**, *9* (3), 031107. https://doi.org/10.1063/5.0039102.

(21) Shell, M. S.; Scott, M.; Ĵ, S. The Relative Entropy Is Fundamental to Multiscale and Inverse Thermodynamic Problems Coarse-Graining Errors and Numerical Optimization Using a Relative Entropy Framework On the Representability Problem and the Physical Meaning of Coarse-Grained Models The Relative Entropy Is Fundamental to Multiscale and Inverse Thermodynamic Problems. *J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys. J. Chem. Phys.* **2008**, *1291* (10), 144108–244114. https://doi.org/10.1063/1.2992060.

(22) Chaimovich, A.; Shell, M. S. Relative Entropy as a Universal Metric for Multiscale Errors. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2010**, *81* (6), 1–4. https://doi.org/10.1103/PhysRevE.81.060104.

(23) Shell, M. S. COARSE-GRAINING WITH THE RELATIVE ENTROPY; 2016; pp 395–441. https://doi.org/10.1002/9781119290971.ch5.

(24) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr,

J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18* (4), 382–388. https://doi.org/10.1038/s41592-021-01098-3.

(25) Chakraborty, S.; Wagh, K.; Gnanakaran, S.; López, C. A. Development of Martini 2.2 Parameters for N -Glycans: A Case Study of the HIV-1 Env Glycoprotein Dynamics. *Glycobiology* **2021**, *31* (7), 787–799. https://doi.org/10.1093/glycob/cwab017.

(26) López, C. A.; Sovova, Z.; van Eerden, F. J.; de Vries, A. H.; Marrink, S. J. Martini Force Field Parameters for Glycolipids. *J. Chem. Theory Comput.* **2013**, *9* (3), 1694–1708. https://doi.org/10.1021/ct3009655.

(27) Vazquez-Salazar, L. I.; Selle, M.; de Vries, A. H.; Marrink, S. J.; Souza, P. C. T. Martini Coarse-Grained Models of Imidazolium-Based Ionic Liquids: From Nanostructural Organization to Liquid–Liquid Extraction. *Green Chem.* **2020**, *22* (21), 7376–7386. https://doi.org/10.1039/D0GC01823F.

(28) Durumeric, A. E. P.; Charron, N. E.; Templeton, C.; Musil, F.; Bonneau, K.; Pasos-Trejo, A. S.; Chen, Y.; Kelkar, A.; Noé, F.; Clementi, C. Machine Learned Coarse-Grained Protein Force-Fields: Are We There Yet? *Curr. Opin. Struct. Biol.* **2023**, *79*, 102533. https://doi.org/10.1016/j.sbi.2023.102533.

(29) Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N. E.; Giorgino, T.; Husic, B. E.; Clementi, C.; Noé, F.; De Fabritiis, G. Machine Learning Coarse-Grained Potentials of Protein Thermodynamics. **2022**.

(30) Jin, J.; Pak, A. J.; Voth, G. A. Understanding Missing Entropy in Coarse-Grained Systems: Addressing Issues of Representability and Transferability. *J. Phys. Chem. Lett.* **2019**, *10* (16), 4549–4557. https://doi.org/10.1021/acs.jpclett.9b01228.

(31) Dunn, N. J. H.; Foley, T. T.; Noid, W. G. Van Der Waals Perspective on Coarse-Graining: Progress toward Solving Representability and Transferability Problems. *Acc. Chem. Res.* **2016**, *49* (12), 2832–2840. https://doi.org/10.1021/acs.accounts.6b00498.

(32) 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.* **2016**, *145* (4). https://doi.org/10.1063/1.4959168.

(33) Dannenhoffer-Lafage, T.; Wagner, J. W.; Durumeric, A. E. P.; Voth, G. A. Compatible Observable Decompositions for Coarse-Grained Representations of Real Molecular Systems. *J. Chem. Phys.* **2019**, *151* (13), 134115. https://doi.org/10.1063/1.5116027.

(34) Suga, K.; Umakoshi, H. Detection of Nanosized Ordered Domains in DOPC/DPPC and DOPC/Ch Binary Lipid Mixture Systems of Large Unilamellar Vesicles Using a TEMPO Quenching Method. *Langmuir* **2013**, *29* (15), 4830–4838. https://doi.org/10.1021/la304768f.

(35) Nyholm, T. K. M.; Lindroos, D.; Westerlund, B.; Slotte, J. P. Construction of a DOPC/PSM/Cholesterol Phase Diagram Based on the Fluorescence Properties of Trans - Parinaric Acid. *Langmuir* **2011**, *27* (13), 8339–8350. https://doi.org/10.1021/la201427w.

(36) Lütgebaucks, C.; Macias-Romero, C.; Roke, S. Characterization of the Interface of Binary Mixed DOPC:DOPS Liposomes in Water: The Impact of Charge Condensation. *J. Chem. Phys.* **2017**, *146* (4), 044701. https://doi.org/10.1063/1.4974084.

(37) Karal, M. A. S.; Mokta, N. A.; Levadny, V.; Belaya, M.; Ahmed, M.; Ahamed, M. K.; Ahammed, S. Effects of Cholesterol on the Size Distribution and Bending Modulus of Lipid Vesicles. *PLoS One* **2022**, *17* (1), e0263119. https://doi.org/10.1371/journal.pone.0263119.

(38) Sundquist, W. I.; Krausslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* **2012**, *2* (7), a006924–a006924. https://doi.org/10.1101/cshperspect.a006924.

(39) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497* (7451), 643–646. https://doi.org/10.1038/nature12162.

(40) Jiang, P.; Hansmann, U. H. E. Modeling Structural Flexibility of Proteins with Go-Models. *J. Chem. Theory Comput.* **2012**, *8* (6), 2127–2133. https://doi.org/10.1021/ct3000469.

(41) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (3), 225–248. https://doi.org/10.1002/wcms.1169.

(42) Leonard, A. N.; Wang, E.; Monje-Galvan, V.; Klauda, J. B. Developing and Testing of Lipid Force Fields with Applications to Modeling Cellular Membranes. *Chem. Rev.* **2019**, *119* (9), 6227–6269. https://doi.org/10.1021/acs.chemrev.8b00384.

(43) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. Multiscale Coarse-Graining and Structural Correlations: Connections to Liquid-State Theory. *J. Phys. Chem. B* **2007**, *111* (16), 4116–4127. https://doi.org/10.1021/jp068549t.

(44) 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 Comput.* **2008**, *4* (5), 819–834. https://doi.org/10.1021/ct700324x.

(45) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5* (12), 3195–3210. https://doi.org/10.1021/ct900313w.

(46) Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, 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.* **2015**, *11* (1),

260–275. https://doi.org/10.1021/ct500477k.

(47) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108* (2), 750–760. https://doi.org/10.1021/jp036508g.

(48) Ingólfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; de Vries, A. H.; Tieleman, D. P.; Marrink, S. J. Lipid Organization of the Plasma Membrane. *J. Am. Chem. Soc.* **2014**, *136* (41), 14554–14559. https://doi.org/10.1021/ja507832e.

(49) Stevens, J. A.; Grünewald, F.; van Tilburg, P. A. M.; König, M.; Gilbert, B. R.; Brier, T. A.; Thornburg, Z. R.; Luthey-Schulten, Z.; Marrink, S. J. Molecular Dynamics Simulation of an Entire Cell. *Front. Chem.* **2023**, *11*. https://doi.org/10.3389/fchem.2023.1106495.

(50) Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* **2019**, *119* (9), 6184–6226. https://doi.org/10.1021/acs.chemrev.8b00460.

(51) Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. T. Two Decades of Martini: Better Beads, Broader Scope. *WIREs Comput. Mol. Sci.* **2023**, *13* (1). https://doi.org/10.1002/wcms.1620.

(52) Winger, M.; Trzesniak, D.; Baron, R.; van Gunsteren, W. F. On Using a Too Large Integration Time Step in Molecular Dynamics Simulations of Coarse-Grained Molecular Models. *Phys. Chem. Chem. Phys.* **2009**. https://doi.org/10.1039/b818713d.

(53) Hadley, K. R.; McCabe, C. On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties. *J. Phys. Chem. B* **2010**, *114* (13), 4590–4599. https://doi.org/10.1021/jp911894a.

(54) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6* (6), e1000810. https://doi.org/10.1371/journal.pcbi.1000810.

(55) Bennett, W. F. D.; Tieleman, D. P. Water Defect and Pore Formation in Atomistic and Coarse-Grained Lipid Membranes: Pushing the Limits of Coarse Graining. *J. Chem. Theory Comput.* **2011**, *7* (9), 2981–2988. https://doi.org/10.1021/ct200291v.

(56) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J. Chem. Theory Comput.* **2021**, *17* (2), 1170–1180. https://doi.org/10.1021/acs.jctc.0c00638.

(57) Jin, J.; Schweizer, K. S.; Voth, G. A. Understanding Dynamics in Coarse-Grained Models. I. Universal Excess Entropy Scaling Relationship. *J. Chem. Phys.* **2023**, *158* (3), 034103. https://doi.org/10.1063/5.0116299.

(58) Lu, L.; Voth, G. A. The Multiscale Coarse-Graining Method. VII. Free Energy

Decomposition of Coarse-Grained Effective Potentials. *J. Chem. Phys.* **2011**, *134* (22), 224107. https://doi.org/10.1063/1.3599049.

(59) Hakobyan, D.; Heuer, A. Comparing an All-Atom and a Coarse-Grained Description of Lipid Bilayers in Terms of Enthalpies and Entropies: From MD Simulations to 2D Lattice Models. *J. Chem. Theory Comput.* **2019**, *15* (11), 6393–6402. https://doi.org/10.1021/acs.jctc.9b00390.

(60) Brandt, E. G.; Braun, A. R.; Sachs, J. N.; Nagle, J. F.; Edholm, O. Interpretation of Fluctuation Spectra in Lipid Bilayer Simulations. *Biophys. J.* **2011**, *100* (9), 2104–2111. https://doi.org/10.1016/j.bpj.2011.03.010.

(61) Watson, M. C.; Brandt, E. G.; Welch, P. M.; Brown, F. L. H. Determining Biomembrane Bending Rigidities from Simulations of Modest Size. *Phys. Rev. Lett.* **2012**, *109* (2), 028102. https://doi.org/10.1103/PhysRevLett.109.028102.

(62) Helfrich, W. Elastic Properties of Lipid Bilayers: Theory and Possible Experiments. *Zeitschrift für Naturforsch. C* **1973**, *28* (11–12), 693–703. https://doi.org/10.1515/znc-1973-11-1209.

(63) Canham, P. B. The Minimum Energy of Bending as a Possible Explanation of the Biconcave Shape of the Human Red Blood Cell. *J. Theor. Biol.* **1970**, *26* (1), 61–81. https://doi.org/10.1016/S0022-5193(70)80032-7.

(64) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(65) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(66) Noid, W. G.; Chu, J. W.; 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.* **2008**. https://doi.org/10.1063/1.2938860.

(67) 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.* **2008**. https://doi.org/10.1063/1.2938857.

(68) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(69) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.;

Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35* (27), 1997–2004. https://doi.org/10.1002/jcc.23702.

(70) Johannes, L.; Pezeshkian, W.; Ipsen, J. H.; Shillcock, J. C. Clustering on Membranes: Fluctuations and More. *Trends Cell Biol.* **2018**, *28* (5), 405–415. https://doi.org/10.1016/j.tcb.2018.01.009.

(71) Jarin, Z.; Pak, A. J.; Bassereau, P.; Voth, G. A. Lipid-Composition-Mediated Forces Can Stabilize Tubular Assemblies of I-BAR Proteins. *Biophys. J.* **2021**, *120* (1), 46–54. https://doi.org/10.1016/j.bpj.2020.11.019.

(72) Khelashvili, G.; Kollmitzer, B.; Heftberger, P.; Pabst, G.; Harries, D. Calculating the Bending Modulus for Multicomponent Lipid Membranes in Different Thermodynamic Phases. *J. Chem. Theory Comput.* **2013**, *9* (9), 3866–3871. https://doi.org/10.1021/ct400492e.

(73) Krishna, V.; Noid, W. G.; Voth, G. A. The Multiscale Coarse-Graining Method. IV. Transferring Coarse-Grained Potentials between Temperatures. *J. Chem. Phys.* **2009**, *131* (2), 024103. https://doi.org/10.1063/1.3167797.

(74) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

(75) Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **2014**, *159* (5), 995–1014. https://doi.org/10.1016/j.cell.2014.10.051.

(76) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. X-Ray Crystallography over the Past Decade for Novel Drug Discovery – Where Are We Heading Next? *Expert Opin. Drug Discov.* **2015**, *10* (9), 975–989. https://doi.org/10.1517/17460441.2015.1061991.

(77) Gallagher-Jones, M.; Glynn, C.; Boyer, D. R.; Martynowycz, M. W.; Hernandez, E.; Miao, J.; Zee, C.-T.; Novikova, I. V.; Goldschmidt, L.; McFarlane, H. T.; Helguera, G. F.; Evans, J. E.; Sawaya, M. R.; Cascio, D.; Eisenberg, D. S.; Gonen, T.; Rodriguez, J. A. Sub-Ångström Cryo-EM Structure of a Prion Protofibril Reveals a Polar Clasp. *Nat. Struct. Mol. Biol.* **2018**, *25* (2), 131–134. https://doi.org/10.1038/s41594-017-0018-0.

(78) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-Scale Molecular Dynamics Simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*; ACM: New York, NY, USA, 2009; pp 1–11. https://doi.org/10.1145/1654059.1654126.

(79) 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.* **2008**, *128* (24), 244115. https://doi.org/10.1063/1.2938857.

(80) 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.* **2010**. https://doi.org/10.1021/ct900643r.

(81) Larini, L.; Lu, L.; Voth, G. A. The Multiscale Coarse-Graining Method. VI. Implementation of Three-Body Coarse-Grained Potentials. *J. Chem. Phys.* **2010**, *132* (16), 164107. https://doi.org/10.1063/1.3394863.

(82) Stillinger, F. H.; Weber, T. A. Computer Simulation of Local Order in Condensed Phases of Silicon. *Phys. Rev. B* **1985**, *31* (8), 5262–5271. https://doi.org/10.1103/PhysRevB.31.5262.

(83) Shell, M. S. The Relative Entropy Is Fundamental to Multiscale and Inverse Thermodynamic Problems. *J. Chem. Phys.* **2008**, *129* (14), 144108. https://doi.org/10.1063/1.2992060.

(84) Chaimovich, A.; Shell, M. S. Coarse-Graining Errors and Numerical Optimization Using a Relative Entropy Framework. *J. Chem. Phys.* **2011**, *134* (9), 094112. https://doi.org/10.1063/1.3557038.

(85) Sezgin, E.; Levental, I.; Mayor, S.; Eggeling, C. The Mystery of Membrane Organization: Composition, Regulation and Roles of Lipid Rafts. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (6), 361–374. https://doi.org/10.1038/nrm.2017.16.

(86) Izvekov, S.; Voth, G. A. Multiscale Coarse-Graining of Mixed Phospholipid/Cholesterol Bilayers. *J. Chem. Theory Comput.* **2006**, No. 2. https://doi.org/10.1021/ct050300c.

(87) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109* (7), 2469–2473. https://doi.org/10.1021/jp044629q.

(88) Lu, L.; Voth, G. A. Systematic Coarse-Graining of a Multicomponent Lipid Bilayer. *J. Phys. Chem. B* **2009**. https://doi.org/10.1021/jp809604k.

(89) Martí, J.; Csajka, F. S. Flip-Flop Dynamics in a Model Lipid Bilayer Membrane. *Europhys. Lett.* **2003**, *61* (3), 409–414. https://doi.org/10.1209/epl/i2003-00192-1.

(90) Sapay, N.; Bennett, W. F. D.; Tieleman, D. P. Molecular Simulations of Lipid Flip-Flop in the Presence of Model Transmembrane Helices. *Biochemistry* **2010**, *49* (35), 7665–7673. https://doi.org/10.1021/bi100878q.

(91) Pak, A. J.; Dannenhoffer-Lafage, T.; Madsen, J. J.; Voth, G. A. Systematic Coarse-Grained Lipid Force Fields with Semiexplicit Solvation via Virtual Sites. *J. Chem. Theory Comput.* **2019**, *15* (3), 2087–2100. https://doi.org/10.1021/acs.jctc.8b01033.

(92) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-

Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.

(93) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(94) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(95) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97* (1), 50–58. https://doi.org/10.1016/j.bpj.2009.04.013.

(96) Terry, C. A.; Fernández, M.-J.; Gude, L.; Lorente, A.; Grant, K. B. Physiologically Relevant Concentrations of NaCl and KCl Increase DNA Photocleavage by an N-Substituted 9-Aminomethylanthracene Dye. *Biochemistry* **2011**, *50* (47), 10375–10389. https://doi.org/10.1021/bi200972c.

(97) Hilt, D. E.; Seegrist, D. W. *Ridge, a Computer Program for Calculating Ridge Regression Estimates /*; Dept. of Agriculture, Forest Service, Northeastern Forest Experiment Station,: Upper Darby, Pa :, 1977. https://doi.org/10.5962/bhl.title.68934.

(98) Liu, P.; Shi, Q.; Dauḿ, H.; Voth, G. A. A Bayesian Statistics Approach to Multiscale Coarse Graining. *J. Chem. Phys.* **2008**, *129* (21). https://doi.org/10.1063/1.3033218.

(99) Pasenkiewicz-Gierula, M.; Baczynski, K.; Markiewicz, M.; Murzyn, K. Computer Modelling Studies of the Bilayer/Water Interface. *Biochim. Biophys. Acta - Biomembr.* **2016**, *1858* (10), 2305–2321. https://doi.org/10.1016/j.bbamem.2016.01.024.

(100) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171. https://doi.org/10.1016/j.cpc.2021.108171.

(101) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* **2001**, *105* (19), 4464–4470. https://doi.org/10.1021/jp010238p.

(102) Protestio, R. Is Henderson's Theorem Practically Useful? *J. Unsolved Quest.* **2013**, *3* (1), 13–15.

(103) Sahrmann, P. G.; Loose, T. D.; Durumeric, A. E. P.; Voth, G. A. Utilizing Machine Learning to Greatly Expand the Range and Accuracy of Bottom-Up Coarse-Grained Models through Virtual Particles. *J. Chem. Theory Comput.* **2023**.

https://doi.org/10.1021/acs.jctc.2c01183.

(104) Lu, L.; Dama, J. F.; Voth, G. A. Fitting Coarse-Grained Distribution Functions through an Iterative Force-Matching Method. *J. Chem. Phys.* **2013**, *139* (12), 121906. https://doi.org/10.1063/1.4811667.

(105) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (80-.).* **2011**, *334* (6055), 517–520. https://doi.org/10.1126/science.1208351.

(106) Lee, H.-S.; Tuckerman, M. E. Dynamical Properties of Liquid Water from Ab Initio Molecular Dynamics Performed in the Complete Basis Set Limit. *J. Chem. Phys.* **2007**, *126* (16), 164501. https://doi.org/10.1063/1.2718521.

(107) Freedman, H.; Laino, T.; Curioni, A. Reaction Dynamics of ATP Hydrolysis in Actin Determined by Ab Initio Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3373–3383. https://doi.org/10.1021/ct3003282.

(108) Cao, J.; Voth, G. A. The Formulation of Quantum Statistical Mechanics Based on the Feynman Path Centroid Density. I. Equilibrium Properties. *J. Chem. Phys.* **1994**, *100* (7), 5093–5105. https://doi.org/10.1063/1.467175.

(109) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet – A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148* (24), 241722. https://doi.org/10.1063/1.5019779.

(110) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-Kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178–184. https://doi.org/10.1016/j.cpc.2018.03.016.

(111) Haghighatlari, M.; Li, J.; Guan, X.; Zhang, O.; Das, A.; Stein, C. J.; Heidar-Zadeh, F.; Liu, M.; Head-Gordon, M.; Bertels, L.; Hao, H.; Leven, I.; Head-Gordon, T. NewtonNet: A Newtonian Message Passing Network for Deep Learning of Interatomic Potentials and Forces. **2021**. https://doi.org/arXiv:2108.02913.

(112) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13* (1), 2453. https://doi.org/10.1038/s41467-022-29939-5.

(113) Li, C.; Voth, G. A. Using Machine Learning to Greatly Accelerate Path Integral Ab Initio Molecular Dynamics. *J. Chem. Theory Comput.* **2022**, *18* (2), 599–604. https://doi.org/10.1021/acs.jctc.1c01085.

(114) Gasteiger, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. **2020**. https://doi.org/arXiv:2003.03123.

(115) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15* (6), 3678–3693.

https://doi.org/10.1021/acs.jctc.9b00181.

(116) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. **2021**. https://doi.org/arXiv:2102.03150.

(117) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5* (5), 755–767. https://doi.org/10.1021/acscentsci.8b00913.

(118) Husic, B. E.; Charron, N. E.; Lemm, D.; Wang, J.; Pérez, A.; Majewski, M.; Krämer, A.; Chen, Y.; Olsson, S.; de Fabritiis, G.; Noé, F.; Clementi, C. Coarse Graining Molecular Dynamics with Graph Neural Networks. *J. Chem. Phys.* **2020**, *153* (19), 194101. https://doi.org/10.1063/5.0026133.

(119) Durumeric, A. E. P.; Voth, G. A. Adversarial-Residual-Coarse-Graining: Applying Machine Learning Theory to Systematic Molecular Coarse-Graining. *J. Chem. Phys.* **2019**, *151* (12), 124110. https://doi.org/10.1063/1.5097559.

(120) Chan, H.; Cherukara, M. J.; Narayanan, B.; Loeffler, T. D.; Benmore, C.; Gray, S. K.; Sankaranarayanan, S. K. R. S. Machine Learning Coarse Grained Models for Water. *Nat. Commun.* **2019**, *10* (1), 379. https://doi.org/10.1038/s41467-018-08222-6.

(121) Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev.* **1958**, *65* (6), 386–408. https://doi.org/10.1037/h0042519.

(122) Fukushima, K. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biol. Cybern.* **1980**, *36* (4), 193–202. https://doi.org/10.1007/BF00344251.

(123) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.

(124) Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph Convolutional Networks: A Comprehensive Review. *Comput. Soc. Networks* **2019**, *6* (1), 11. https://doi.org/10.1186/s40649-019-0069-y.

(125) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. **2022**. https://doi.org/arXiv:2204.05249.

(126) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. **2018**.

(127) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. **2017**. https://doi.org/arXiv:1704.01212.

(128) Bidone, T. C.; Polley, A.; Jin, J.; Driscoll, T.; Iwamoto, D. V.; Calderwood, D. A.; Schwartz, M. A.; Voth, G. A. Coarse-Grained Simulation of Full-Length Integrin Activation. *Biophys. J.* **2019**, *116* (6), 1000–1010. https://doi.org/10.1016/j.bpj.2019.02.011.

(129) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95* (9), 4183–4192. https://doi.org/10.1529/biophysj.108.139733.

(130) Hadley, K. R.; McCabe, C. Coarse-Grained Molecular Models of Water: A Review. *Mol. Simul.* **2012**, *38* (8–9), 671–681. https://doi.org/10.1080/08927022.2012.671942.

(131) Jin, J.; Han, Y.; Pak, A. J.; Voth, G. A. A New One-Site Coarse-Grained Model for Water: Bottom-up Many-Body Projected Water (BUMPer). I. General Theory and Model. *J. Chem. Phys.* **2021**, *154* (4), 044104. https://doi.org/10.1063/5.0026651.

(132) Chaimovich, A.; Shell, M. S. Anomalous Waterlike Behavior in Spherically-Symmetric Water Models Optimized with the Relative Entropy. *Phys. Chem. Chem. Phys.* **2009**, *11* (12), 1901. https://doi.org/10.1039/b818512c.

(133) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001. https://doi.org/10.1103/PhysRevLett.120.143001.

(134) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; E, W. End-to-End Symmetry Preserving Inter-Atomic Potential Energy Model for Finite and Extended Systems. **2018**.

(135) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511–519. https://doi.org/10.1063/1.447334.

(136) Hoover, W. G.; Holian, B. L. Kinetic Moments Method for the Canonical Ensemble Distribution. *Phys. Lett. A* **1996**, *211* (5), 253–257. https://doi.org/10.1016/0375-9601(95)00973-6.

(137) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE, 2016; pp 770–778. https://doi.org/10.1109/CVPR.2016.90.

(138) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3* (6), 2034–2045. https://doi.org/10.1021/ct700127w.

(139) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *32*, NA-NA. https://doi.org/10.1002/jcc.21367.

(140) Tuckerman, M. E. Ab Initio Molecular Dynamics: Basic Concepts, Current Trends and Novel Applications. *J. Phys. Condens. Matter* **2002**, *14* (50), 1297–1356.

(141) Carloni, P.; Rothlisberger, U.; Parrinello, M. The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems. *Acc. Chem. Res.* **2002**, *35* (6), 455–464. https://doi.org/10.1021/ar010018u.

(142) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103* (2), 227–249. https://doi.org/10.1016/0022-2836(76)90311-9.

(143) Feynman, R. P.; Hibbs, A. R. *Quantum Mechanics and Path Integrals*; 1965.

(144) Berne, B. J.; Thirumalai, D. On the Simulation of Quantum Systems: Path Integral Methods. *Annu. Rev. Phys. Chem.* **1986**, *37* (1), 401–424. https://doi.org/10.1146/annurev.pc.37.100186.002153.

(145) Markland, T. E.; Ceriotti, M. Nuclear Quantum Effects Enter the Mainstream. *Nat. Rev. Chem.* **2018**, *2* (3), 0109. https://doi.org/10.1038/s41570-017-0109.

(146) Cao, J.; Voth, G. A. A New Perspective on Quantum Time Correlation Functions. *J. Chem. Phys.* **1993**, *99* (12), 10070–10073. https://doi.org/10.1063/1.465512.

(147) Cao, J.; Voth, G. A. The Formulation of Quantum Statistical Mechanics Based on the Feynman Path Centroid Density. II. Dynamical Properties. *J. Chem. Phys.* **1994**, *100* (7), 5106–5117. https://doi.org/10.1063/1.467176.

(148) Cao, J.; Voth, G. A. The Formulation of Quantum Statistical Mechanics Based on the Feynman Path Centroid Density. I. Equilibrium Properties. *J. Chem. Phys.* **1994**, *100* (7), 5093–5105. https://doi.org/10.1063/1.467175.

(149) Jang, S.; Voth, G. A. Path Integral Centroid Variables and the Formulation of Their Exact Real Time Dynamics. *J. Chem. Phys.* **1999**, *111* (6), 2357–2370. https://doi.org/10.1063/1.479514.

(150) Jang, S.; Voth, G. A. A Derivation of Centroid Molecular Dynamics and Other Approximate Time Evolution Methods for Path Integral Centroid Variables. *J. Chem. Phys.* **1999**, *111* (6), 2371–2384. https://doi.org/10.1063/1.479515.

(151) Craig, I. R.; Manolopoulos, D. E. Chemical Reaction Rates from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2005**, *122* (8), 084106. https://doi.org/10.1063/1.1850093.

(152) Miller, T. F.; Manolopoulos, D. E. Quantum Diffusion in Liquid Para -Hydrogen from Ring-Polymer Molecular Dynamics. *J. Chem. Phys.* **2005**, *122* (18), 184503. https://doi.org/10.1063/1.1893956.

(153) Miller, T. F.; Manolopoulos, D. E. Quantum Diffusion in Liquid Water from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2005**, *123* (15), 154504.

https://doi.org/10.1063/1.2074967.

(154) Jang, S.; Sinitskiy, A. V; Voth, G. A. Can the Ring Polymer Molecular Dynamics Method Be Interpreted as Real Time Quantum Dynamics? *J. Chem. Phys.* **2014**, *140* (15), 1–11. https://doi.org/10.1063/1.4870717.

(155) Hele, T. J. H.; Willatt, M. J.; Muolo, A.; Althorpe, S. C. Communication: Relation of Centroid Molecular Dynamics and Ring-Polymer Molecular Dynamics to Exact Quantum Dynamics. *J. Chem. Phys.* **2015**, *142* (19), 191101. https://doi.org/10.1063/1.4921234.

(156) Witt, A.; Ivanov, S. D.; Shiga, M.; Forbert, H.; Marx, D. On the Applicability of Centroid and Ring Polymer Path Integral Molecular Dynamics for Vibrational Spectroscopy. *J. Chem. Phys.* **2009**, *130* (19), 194510. https://doi.org/10.1063/1.3125009.

(157) Paesani, F.; Voth, G. A. A Quantitative Assessment of the Accuracy of Centroid Molecular Dynamics for the Calculation of the Infrared Spectrum of Liquid Water. *J. Chem. Phys.* **2010**, *132* (1), 014105. https://doi.org/10.1063/1.3291212.

(158) Habershon, S.; Fanourgakis, G. S.; Manolopoulos, D. E. Comparison of Path Integral Molecular Dynamics Methods for the Infrared Absorption Spectrum of Liquid Water. *J. Chem. Phys.* **2008**, *129* (7), 074501. https://doi.org/10.1063/1.2968555.

(159) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to Remove the Spurious Resonances from Ring Polymer Molecular Dynamics. *J. Chem. Phys.* **2014**, *140* (23), 234116. https://doi.org/10.1063/1.4883861.

(160) Hone, T. D.; Rossky, P. J.; Voth, G. A. A Comparative Study of Imaginary Time Path Integral Based Methods for Quantum Dynamics. *J. Chem. Phys.* **2006**, *124* (15), 154103. https://doi.org/10.1063/1.2186636.

(161) Markland, T. E.; Manolopoulos, D. E. An Efficient Ring Polymer Contraction Scheme for Imaginary Time Path Integral Simulations. *J. Chem. Phys.* **2008**, *129* (2), 024105. https://doi.org/10.1063/1.2953308.

(162) Hone, T. D.; Izvekov, S.; Voth, G. A. Fast Centroid Molecular Dynamics: A Force-Matching Approach for the Predetermination of the Effective Centroid Forces. *J. Chem. Phys.* **2005**, *122* (5), 054105. https://doi.org/10.1063/1.1836731.

(163) Fletcher, T.; Zhu, A.; Lawrence, J. E.; Manolopoulos, D. E. Fast Quasi-Centroid Molecular Dynamics. *J. Chem. Phys.* **2021**, *155* (23), 231101. https://doi.org/10.1063/5.0076704.

(164) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001. https://doi.org/10.1103/PhysRevLett.120.143001.

(165) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; E, W. End-to-End Symmetry Preserving Inter-Atomic Potential Energy Model for Finite and Extended Systems. In *Advances in Neural Information Processing Systems*; Garnett, S. B. and H. W. and H. L. and K. G. and

N. C.-B. and R., Ed.; 2018.

(166) Silvera, I. F.; Goldman, V. V. The Isotropic Intermolecular Potential for H 2 and D 2 in the Solid and Gas Phases. *J. Chem. Phys.* **1978**, *69* (9), 4209–4213. https://doi.org/10.1063/1.437103.

(167) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A. An Accurate and Simple Quantum Model for Liquid Water. *J. Chem. Phys.* **2006**, *125* (18), 184507. https://doi.org/10.1063/1.2386157.

(168) Kapil, V.; Rossi, M.; Marsalek, O.; Petraglia, R.; Litman, Y.; Spura, T.; Cheng, B.; Cuzzocrea, A.; Meißner, R. H.; Wilkins, D. M.; Helfrecht, B. A.; Juda, P.; Bienvenue, S. P.; Fang, W.; Kessler, J.; Poltavsky, I.; Vandenbrande, S.; Wieme, J.; Corminboeuf, C.; Kühne, T. D.; Manolopoulos, D. E.; Markland, T. E.; Richardson, J. O.; Tkatchenko, A.; Tribello, G. A.; Van Speybroeck, V.; Ceriotti, M. I-PI 2.0: A Universal Force Engine for Advanced Molecular Simulations. *Comput. Phys. Commun.* **2019**, *236*, 214–223. https://doi.org/10.1016/j.cpc.2018.09.020.

(169) Pérez, A.; Tuckerman, M. E.; Müser, M. H. A Comparative Study of the Centroid and Ring-Polymer Molecular Dynamics Methods for Approximating Quantum Time Correlation Functions from Path Integrals. *J. Chem. Phys.* **2009**, *130* (18), 184105. https://doi.org/10.1063/1.3126950.

(170) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C. A Systematic Study of Water Models for Molecular Simulation: Derivation of Water Models Optimized for Use with a Reaction Field. *J. Chem. Phys.* **1998**, *108* (24), 10220–10230. https://doi.org/10.1063/1.476482.

(171) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006**, *124* (2), 024503. https://doi.org/10.1063/1.2136877.

(172) Chen, M.; Ko, H.-Y.; Remsing, R. C.; Calegari Andrade, M. F.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P.; Wu, X. Ab Initio Theory and Modeling of Water. *Proc. Natl. Acad. Sci.* **2017**, *114* (41), 10846–10851. https://doi.org/10.1073/pnas.1712499114.

(173) Wang, W.; Gómez-Bombarelli, R. Coarse-Graining Auto-Encoders for Molecular Dynamics. *npj Comput. Mater.* **2019**, *5* (1), 125. https://doi.org/10.1038/s41524-019-0261-5.

(174) Potter, T. D.; Barrett, E. L.; Miller, M. A. Automated Coarse-Grained Mapping Algorithm for the Martini Force Field and Benchmarks for Membrane–Water Partitioning. *J. Chem. Theory Comput.* **2021**, *17* (9), 5777–5791. https://doi.org/10.1021/acs.jctc.1c00322.

(175) 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.* **2008**, *95* (11), 5073–5083. https://doi.org/10.1529/biophysj.108.139626.