THE UNIVERSITY OF CHICAGO

CONTRIBUTIONS TOWARDS THE THEORY OF THE BOTTOM-UP COARSE-GRAINING OF COMPLEX MOLECULES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTORATE OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
JAMES FARRIS DAMA

CHICAGO, ILLINOIS
JUNE 2016

Copyright © 2016 by James Farris Dama

All Rights Reserved

To my grandfather Farris Moore Elgee.

Let us beware of thinking that the world is a living being. Where would it stretch? What would it feed on? How could it grow and procreate? After all, we know roughly what the organic is; are we then supposed to reinterpret what is inexpressibly derivative, late, rare, accidental, which we perceive only on the crust of the earth, as something essential, common, and eternal, as those people who call the universe an organism? This nauseates me. Let us beware even of believing that the universe is a machine; it is certainly not constructed to one end, and the word "machine" pays it far too high an honor. Let us beware of assuming in general and everywhere anything as elegant as the cyclical movements of our neighboring stars; even a glance at the Milky Way raises doubts whether there are not much coarser and more contradictory movements there, as well as stars with eternally linear paths, etc. The astral order in which we live is an exception; this order and the considerable duration that is conditioned by it have again made possible the exception of exceptions: the development of the organic. The total character of the world, by contrast, is for all eternity chaos, not in the sense of a lack of necessity but a lack of order, organization, form, beauty, wisdom, and whatever else our aesthetic anthropomorphisms are called. Judged from the vantage point of our reason, the unsuccessful attempts are by far the rule; the exceptions are not the secret aim, and the whole musical mechanism repeats eternally its tune, which must never be called a melody–and ultimately even the phrase "unsuccessful attempt" is already an anthropomorphism bearing a reproach. But how could we reproach or praise the universe!

-Friedrich Nietzsche, *The Gay Science*, ed. Bernard Williams, trans. Josefine Nauckhoff

.

I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false.

-Ludwig Wittgenstein, *On Certainty*, ed. G. E. M. Anscombe & G. H. von Wright

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank fate, first of all, for all of the countless accidents and necessities that have preceded this and brought about me and milieu. Second, the innumerable men and women who preceded me and make the human world what it is today. But it would be inhuman to continue with such a broad approach; I will not.

I should thank Gregory Voth and the entire Voth group first among my professional loyalties. Greg has given me incredible rein and opportunities for exploration at the foundations of the field, at the same time giving powerful direction and vision. It is hard to overstate his influence and it would be hard to overstate the importance of frequent conversations within the group. Officemates John Savage, John Grime, and Rui Sun kept me sane—or boundedly insane—and among my other collaborators and coauthors I am especially indebted to Grant Rotskoff, Anton Sinitsky, Jacob Wagner, Aram Davtyan, Andrew White, and Glen Hocky. Among the younger cohort, Jaehyeok Jin, Alek Durumeric, and Yining Han are quickly becoming close collaborators as well, and their prints are all over the latest sections of this manuscript. I have to particularly thank the group for their patience with my opportunistic and wide-ranging interests that seldom stayed focused on one (visible) topic for very long, and for not laughing too much when I insist there is a program hiding behind all of my individual projects, even if that might take another ten or twenty years to complete.

My strongest influences have not been limited to the Voth group of course. Michele Parrinello and Hans Andersen deserve special mention. Michele's hospitality in Lugano, and his group's, gave rise to a shared research project I wouldn't have dreamed of when beginning graduate school, and he provided invaluable advice throughout. In his group, I have also to thank Federico Giberti and Gareth Tribello especially for initial introductions to metadynamics and Omar Valsson, Patrick Shaffer, and Pratyush Tiwary for later conversations. Hans, on the other hand, may be the primary reason I got into the theory of bottom-up coarse-graining: he and John Grime gave compelling presentations my first year that provided the initial inspiration for all of my later work in coarse-graining. Our conversations

since have only reinforced his influence.

My interest in enhanced sampling began with Aaron Dinner and Jonathan Weare; those now long-abandoned multigrid umbrella sampling notes we shared early on informed all my later work, and I have continued to benefit from colocation with their shared research ever since. Jonathan deserves particular thanks for his patience with my bloody determination to write a real proof despite being a chemist. Benoît Roux shared indispensable insight into both coarse-graining and enhanced sampling of biomolecules at key moments.

It would also be hard to overstate the importance of those fortuitous encounters at conferences that often prove so fertile. In no particular order, Pep Espanol, Wei Yang, Mike Schnieder, Fabrizio Marinelli, Mahmoud Moradi, Jose Faraldo-Gomez, Bernie Brooks, Jerome Henin, Chris Chipot, Gabriele Stoltz, Greg Bowman, Rafael Delgado-Buscalioni, Benjamin Jourdain, Suri Vaikuntanathan, Garnet Chan, Tim Berkelbach, Patrick Shaffer, and many others have made the research feel worthwhile by their interest and their good suggestions. The scientists of the Los Alamos theoretical division, especially Art Voter, Danny Perez, and the 2014 summer visitor cohort, have also been wonderful influences.

Among friends and family, it would be difficult to know who to thank for what given that I shouldn't make these acknowledgments the length of a second dissertation themselves. My mother and father, of course, have supported me and raised me, and my brothers can't be ignored for their opposing influences towards practicality and idealism. My grandfather deserves his dedication; he left a deep mark on my character for the better. He is missed. To my friends it wouldn't be at all my style to be very particular in a public document: if you friends ever want more thanks it's my pleasure to offer it in embarrassingly heartfelt detail in person—including all I have and have not kept in contact with over these years. In many cases, just having a memory or two has made all the difference.

ABSTRACT

The coarse-grained modeling of molecular systems is a burgeoning field, yet significant challenges remain in interpreting these models and understanding how to build new ones for new systems. Bottom-up coarse-graining provides methodology for building new coarse-grained models systematically and theoretical foundations for understanding how to interpret coarse-grained models more generally. However, it has limitations, especially when it comes to modeling systems of complex molecules. This thesis describes research that begins with work towards deeper understanding of the bottom-up coarse-graining of many relatively simple molecules, then branches towards work on the coarse-graining of few complex molecules, and finally concludes by pursuing work on the coarse-graining of many moderately complex molecules. The results of these threads are 1) new theory of coarse-grained equilibrium thermodynamics with powerful tools for its computational investigation, 2) deep theoretical insight into nonequilibrium adaptation, proven by a flurry of methodological applications in enhanced sampling, and 3) expanded paradigms for the renormalization of complex molecular systems that provide foundations for otherwise impossible modeling challenges.

CHAPTER 1

UNSCIENTIFIC PREFACE

When I applied to graduate school, I had a single eventual goal on my mind: to better understand the molecular details underlying continuum field theories, especially continuum fluid transport. I had learned just enough about renormalization under Zhen-Gang Wang and Olexei Motrunich at Caltech to know that it was difficult even to derive continuum field theories from microscopic field theories, as in one of Professor Wang's papers on the renormalization of polymer field theories,[349] much less to derive continuum theories from the true underlying molecular-scale physics.

I was fascinated for several reasons. First, there seemed to be abundant chemical applications in polymer physics, fluid transport, and biological materials chemistry. There is ample research on universality classes in today's physics, but relatively little research on design guidance for how to optimize behavior within a given universality class; understanding the actual micro-macro correspondence precisely rather than appealing to universality might lead to a trove of design insights. More personally, I have always been fascinated by how smooth and discrete appearances emerge out of many-scaled roughness not just in physics but in personal life—for instance, how do Marxist class analysis and free market equilibria seem so compelling as social continuum theories to so many despite their opaque microfoundations? Or how do group social processes that seem jammed and glassy suddenly give way to consensus when the group is faced with an external threat?

At the time of my transition to graduate school, this problem of abstraction seemed to be one of the sort that requires sending out deep roots and growing a broad trunk in ways that others may never follow or endorse until the tree bears fruit. Seeking to go as deep as possible with the fewest preconceptions possible, I chose Professor Voth as an advisor because in our interview he emphasized finding out what was missing in coarse-grained models, not just how to apply them or make the most of the theory we already have.

Indeed, as I worked, I became only more convinced that the types of models being used

in coarse-graining are cripplingly incomplete. They are neither transferable nor capable of fitting many subtleties correctly. For instance, the pair-correlation-based Lum-Chandler-Weeks[205] theory will work for pure hydrophobicity but not for hydrogen bonding in protein solvation, where specificity is key.[38, 176] In the coarse-graining of water molecules to one bead, pair potentials cannot reproduce the liquid density maximum,[68] and though three body potentials can reproduce the liquid thermodynamics well,[230] they do it by compromising on structural accuracy,[185, 199]—and the proton ordered and disordered ice phases, of course, are missed entirely.

Therefore I developed a primary interest in determining how models can correspond to one another, what makes them distinguishable, and how they can be systematically refined. This provides one major theme of this thesis, whether it arises in the guise of particle interaction complexity in Chapters 3-5, the dynamic complexity of collective particle motions in Chapters 6-10, or the possibility of combining discrete state transitions with smooth particle interactions in Chapters 11-13. My interest was also deeply shaped by reading Wittgenstein, Anscombe, Heidegger, and Braver beginning in my second year. Their relentless emphases on watching how language and models are actually used and how reasons are actually given inspired me to look the same way at coarse-grained modeling and enhanced sampling.

Professor Voth's coarse-graining work emphasizes the ideal of correspondence between looking at a coarse-grained model and looking at a fine-grained model through a coarse-grained lens. This is a very particular kind of correspondence that has numerous implications for the kinds of questions that can be asked of the coarse-grained models and the kind of uses to which they can be put. First, it immediately shows why most such models will not be transferable (seen in Chapter 3), how to directly pose the question of how transferable they are (explored in Chapter 4), and how that issue of transferability changes how other questions can be asked of the model, i.e., how the definitions of experimentally observed properties in the model must therefore differ from the definitions of the same in more standard atomistic models (Chapter 5).

Like any model a coarse-grained model is not a real thing, but a specific kind of representation with specific uses, which it gets from linked practices of interpreting experiments. Conceiving of the model this way frees the theorist from seeing common atomistic model behavior as an ideal, revealing coarse-grained models as they actually are instead of consigning them to appear as poor, too complicated and haphazardly interpreted imitations of atomistic models. Coarse-grained models intentionally conceal the fundamentals of a situation, concealing what grounds them: the authors cited above provide expert descriptions of the language of approximations that do not allow recourse to fundamental grounds, with Braver's *Groundless Grounds* being most clearly a book on exactly this point.

However, it is impossible to convey exactly how such philosophy has been important without also briefly explaining how I do theory by analogy and by metaphor. My published work is very direct and mathematical, but it is not how I think. However unfashionable it might be to admit, I rely immensely on using anthropomorphism and similar techniques to articulate my problems. For the confidence to embrace this tendency I have to thank at minimum Julie Kornfield, who taught her classes that she liked to understand transport and rheology by imagining herself as a fluid element or polymer being sheared, and Bill Goddard, who taught quantum by teaching us what electrons "like" best and had us students act out bonding topologies by literally joining hands to form the shapes ourselves.

This worked very well for me! Contra injunctions against anthropomorphism as insufficiently rigorous, I find that attributing things like 'preference,' 'intention,' and 'frustration' to chemical systems to be powerful and suggestive language.

This is the reason for my Nietzsche epigraph above, a joke of sorts. It is the first half of a long tirade against anthropomorphism, but it is a tirade against "thinking that nature is human." However, after years of reading Nietzsche, one learns to be wary even of anthropomorphizing man himself—man is a special case of the animal like the living (biological) is a special case of the dead (chemical).

This naturally combines with Anscombe's remarks in *Intention* that intention need not

be something that actually goes on in our heads, but is simply a way to answer the question "why?" asked in a particular sense. There is no reason not to apply anthropomorphic language to chemicals and molecules so long as one is careful not to actually impute mental activity to them. The proper Wittgensteinian hesitates to impute mental activity even to people! (Or to deny it.) In other words, rather than attributing any sort of humanity to the molecules, what I am really doing when I anthropomorphize them by describing their preferences, intentions, and frustrations is revealing how little humanity is required to generate complexities isomorphic to the complexities of human life that initially inspired the language of preference, intention, and frustration. Instead of anthropomorphizing nature, this naturalizes mankind.

Each of the following chapters represents approximately one published paper and substantially reproduces the material in that paper. Each, therefore, is written professionally and does not contain the colorful metaphor that inspired and guided the work. However, for the sake of making this thesis really representative of my thinking, I've put small, linked and growing metaphors before each one that link the phenomena in chemical systems, concerning molecules, to phenomena in social systems, concerning people. These little thoughts are how I actually did the work—even how I actually worked out the rough forms of the equations, though I don't expect to be able to explain that here to anyone who does not already know those tricks themselves.

I hope that these will also illustrate somewhat how the philosophy above keeps these analogies from simply going off the rails and becoming fluffy nonsense. There is always another detail to weave into the metaphor, but for a rich enough analogy the metaphors can be extended remarkably far and, even more importantly, the failures of initially intuitive metaphors are highly informative.

CHAPTER 2

INTRODUCTION

2.1 Multiscale Physics and the Nano-Meso Barrier

The last half-century has been an age of multiscale physics, from boundary layer theory[35] rocket aerodynamics to renormalization[356, 139] emulating neural networks.[355, 225] In contrast to the initial apparent universality of the laws of Galileo, Newton, or Maxwell, since the maturation of quantum physics we have begun to see all physical laws as having specific domains of application: specific scales on which they make sense and others on which they do not. At the most fundamental level quantum mechanics and general relativity stubbornly resist being made compatible, much less becoming unified. Scientists can therefore either study a single scale or multiple scales—and they can also study the bridges between them, the subject of this thesis.

There are many successful theories for bridging scales, from boundary layer theory[35, 72] to homogenization[37, 258] to renormalization group[252, 356, 139, 72] to the heterogeneous multiscale method.[110, 1] All have a central idea in common: singular perturbation theory.[35, 72] This is the mathematics describing cases in which one set of laws seem to hold when the deviation from an ideal case is small, but increasing distance from the ideal case causes qualitative change in the apparent laws. A corollary of this is that when one tries to patch up the ideal case's laws with epicycles, the epicycles explode in number and importance as more are added—the approximation does not improve with additional corrections, but actually gets worse to the point of meaninglessness.

For my part, I am most interested in molecular physics: how chemistry accounts for the rich diversity of colors, textures, and dynamics characteristic of oceans and corals, clouds and birds. Typically, the fluids and biological systems I am most interested fall under the scientific heading of soft condensed matter. For better or worse, soft condensed matter strongly resists perturbative analyses around ideal cases: even the best microscopic perturbative theories we

currently have for molecular soft matter, e.g., the reference interaction site model[71, 31], are not always reliable; the best cases that are reliable, e.g. the Weeks-Chandler-Andersen theory,[350, 13] are frustratingly difficult to generalize.

For this reason, we must often resort to direct molecular dynamics (MD) simulation:[131, 7, 262] we hardly simplify the system into the laws of an ideal case at all, but rather calculate the properties from highly complex and general models. We write any perturbation theories in terms of these complex, minimally idealized MD reference states, and unlike in the Weeks-Chandler-Andersen theory where each reference calculation gives information about a whole class of fluids, each reference calculation in our normal cases is likely to give information only about a particular system in a particular system state or, when doing alchemical transformations and the like, a narrow range of system states.

Direct MD simulation allows us to answer many questions at the nanometer scale. We have good physical understanding of the quantum physics governing the atomic motions at this scale. However, MD of atoms is not currently feasible on large scales and some systems are heterogeneous in a way that cannot easily be modeled in terms of a decomposition into smaller systems. Examples include polymer fuel cell membranes, hydrogels, and large protein complexes.

When nanoscale properties are insufficient to model a phenomenon, but molecular detail remains important, this is the so-called mesoscale[136, 276] between atomic physics and continuum field theories. To investigate mesoscale physics, one can lump atoms together into effective particles with effective interactions and perform coarse-grained MD on the lumped particles or one can apply molecular field theory models and density functional theories that ignore very small-scale structure without compromising larger patterns of molecular organization.[130] However, the parameters of CG MD and field theory models can be difficult to connect directly to chemistry,[349] so the models are predictive and useful in guiding chemistry only insofar as one has a good intuition for how different chemistries will affect the parameters of the model. This can be a subtle problem.

Therefore there are many ongoing efforts to bridge between the nanoscale and mesoscale by developing better mesoscale CG models and stronger understandings of how nanoscopic chemistry gives rise to coarse-grained mesoscopic physics. At stake may be the material design principles of the future, deep understanding of cell-scale biology, and—perhaps—fundamentally new paradigms for understanding connections between atomistic models and field theories.

2.2 This Thesis

The current thesis is part of the effort to bridge between the nanoscale and mesoscale. I've spent my graduate career building a foundation for understanding the connection between these scales by understanding how to derive coarse-grained models exactly equivalent (in precise senses explained later) to atomistic models, called bottom-up coarse-graining.[207, 276, 154, 296, 246, 235] This paradigm has been explored for decades now, from the early methods simply based on iterating until things looked right[207] to modern variational and machine-learning approaches, which define principled error metrics and then seek models to minimize those errors.[154, 296, 246, 235, 241]

A common criticism of bottom-up models is that they are not as accurate in matching experiment as mesoscopic models that are parameterized to match experiment directly.[292, 322, 216] However, these other models typically have many parameters and von Neumann's famous joke about fitting elephants should be kept in mind. Such models often fit known values but make incorrect predictions about other system details. As importantly, studying bottom-up coarse-graining does not only create models, but also guides theoretical understanding. It reveals connections between fine-grained and coarse-grained scales even when the generated models do not function directly as predictors of experimental observations.

My interest has above all been in this connection between scales. To that end I have developed new theory for calculating coarse-grained force fields between many simple particles and for properly interpreting those models, new theory for using out of equilibrium sampling

of complex molecules to characterize their coarse-grained collective motions at equilibrium, and new theory for deriving models of many molecules that have internal dynamics that cannot be represented in terms of coarse-grained particle motion.

Together, these projects pull out two features of the connection that I hope will lead to a general reconsideration of the project of bottom-up coarse-graining. First, the connection between mesoscopic models and fine-grained models is considerably more complex than usually appreciated. However, second, that complexity can cancel itself out in surprisingly robust and designable ways that have often been stumbled upon by accident (accounting for much of the early successes). Bottom-up coarse-graining has not been as simple to use as its inventors initially hoped, but the reasons for that may prove revolutionary in time.

The first three body chapters, Chapters 3-5, concern the coarse-graining of systems of many simple molecules by means of variational force matching, i.e., the multiscale coarse-graining (MS-CG) method.[154, 153, 248, 246, 247, 201] This method picks out coarse-grained models that minimize the squared norm of the difference between coarse-grained model's predicted forces and observed forces from a corresponding fine-grained model that have been passed through a corresponding coarse-grained lens. If the method is allowed to choose from among all possible coarse-grained models of the chosen resolution, then one can guarantee that the coarse-grained model's equilibrium statistics are identical to the equilibrium statistics of the fine-grained model observed through the coarse-graining lens. All of this is described explicitly in Chapter 3, which discusses MS-CG, its derivation, its implementation, and several applications in detail.

One serious drawback of coarse-grained models is that they can lose their connections to the underlying physics of a fine-grained model. Even if one knows how interactions should be changed at a coarse-grained level to achieve a desired behavior, it may still be unclear how to design chemistries or system states that change the coarse-grained interactions in the desired way. To address this limitation I derived new computationally efficient formulae for the sensitivity of bottom-up coarse-grained force fields to parameters of its underlying

fine-grained model, described in Chapter 4.

On the other side, the coarse-grained models also may not correspond to experiment in the same way that atomistic models do. For instance coarse-grained force fields, unlike the quantum mechanics governing interactions between atoms, can often be very sensitive to changes in pressure and temperature. They should not be interpreted purely as potential energies, but may also include entropy and enthalpy, changing how internal energy and entropy are measured in the system. Chapter 5 describes such issues in detail, discussing not just thermodynamics but observables of all types including ones that might seem to be entirely unavailable at the coarse-grained level.

The central five chapters, Chapters 6-10, discuss efficient coarse-graining of complex systems down to just their few slowest, most complex motions by means of the method metadynamics.[181, 28, 27] Metadynamics accelerates these slow motions by forcing systems away from their previously visited configurations, countering metastability, but in a controlled way that allows for exact characterization of the unbiased equilibrium statistics of the system's slowest motions after the fact. It offers orders of magnitude faster characterization than direct simulation in the best cases, but has suffered from significant drawbacks to the point it was long believed to be fundamentally approximate rather than exact.[27, 313] This thesis presents fixes for many of these drawbacks.

Chapters 6-10 first introduce the metadynamics method, then provide proof that it converges exactly and under which conditions,[88] then discuss three methodological improvements in turn: an alteration that eliminates tradeoffs between configuration space exploration rate and convergence efficiency,[89] an alteration that improves stability by allowing for bounded exploration rather than complete exploration,[87] and an alteration that prevents false convergence by introducing a new control sequence for pushing the system away from previously visited configurations.

Chapter 7's proof of convergence provides rationale for analysis of the method via a quasi-equilibrium approximation that is unprecedented in the chemistry literature, introducing

exceptionally powerful stochastic approximation proof methods[179] that directly guided the method development in Chapters 8-10. Each methodological improvement in those chapters takes a specific feature of Chapter 7's proof and turns it into a guide to improvement rather than simply a technicality of the proof.

Finally, Chapters 11-13 describe a class of models designed to braid the threads described in Chapters 3-5 and 6-10. When each molecule of a system has complex, slow motions of Chapters 6-10 can be described as transitions between just a few discrete states per molecule, but there are many such molecules, Chapter 11 provides means to coarse-grain these systems from the bottom up into models with many coarse-grained entities that contain discrete states.

Chapter 11 describes only the theory of fitting force fields with discrete states, leaving Chapters 12 and 13 to describe how to actually implement such models and simulate them in two polar opposite limits: in Chapter 12, the discrete states change infrequently but promptly, creating a surface hopping dynamics, and in Chapter 13 the states equilibrate so rapidly that the system can always be assumed to keep a quasi-equilibrium state distribution. Each has its own peculiarities of implementation and parameterization, discussed in the relevant chapters.

The theory of Chapter 11 explicitly considers the case that machinery like described in Chapters 6-10 could be used to generate input force observations for a bottom-up force matching coarse-graining procedure, and the general principles of Chapters 3-5 also translate remarkably simply to the context of Chapters 11-13. In many cases, we are able to use the same codes for both with no modification because the theory is truly fundamental and general.

Finally, Chapter 14 concludes the thesis and describes current directions of research that will continue on as I complete my work here and afterwards.

CHAPTER 3

COARSE-GRAINING MANY SIMPLE MOLECULES: VARIATIONAL FORCE-MATCHING

The following chapter reprises material originally prepared as a chapter for a book reviewing biomolecular coarse graining. It was written primarily 2011-2012 and only slightly revisited 2013, and though it is missing some of the newer references in the field, it is surprising how little the issues it identifies have changed.

This section describes how to use overall forces on each of many groups of particles to determine the average forces each group exerts on each other. The average forces define a coarse-grained potential for the coarse-grained particle groups. The groups of particles are permanently bonded, so I'll imagine them as families—possibly bound to other families through marriage as well. If we look at the forces as preferences, then what we are doing is figuring out the preferences of families for particular types of neighbors from their preferences for variously mixed neighborhoods. This echoes a classic model for racial segregation popularized by Thomas Schelling.

There are two basics ways to impute preferences for neighbors based on neighborhood preferences. One is to look at actual neighborhood occupation, assuming people live where they would like to live, the method of revealed preference. Another is to ask them which marginal changes to their neighborhoods they would tolerate, the method of aggregated preference. In economic practice, these often do not agree and in politics this is often attributed to cognitive dissonance or other similar dishonesty. For instance, white people may not say they don't want to live by black neighbors, even in a survey, though in practice they don't live by black neighbors. One way to resolve this is to say these white people are in fact prejudiced as individuals, just lying or in denial. Another, more subtle way that Schelling illustrated beautifully, is to note that though individual white families' preferences for white neighbors may be weak, when white families cluster together their small preferences

are amplified, leading to overall disaggregation. I.e., there is structural racism.

The imputation of individual strong racism above relied on an implicit mean-field model that assumed white and black families would be well-mixed to begin with, and this is why it leads one astray in the case of structural racism. The method of revealed preference is likely to give very misleading answers in the presence of strong cooperative effects if the model of the underlying preferences is not accurate.

The same phenomenon occurs when comparing distribution matching coarse-graining methods (the revealed preference analog) to force matching coarse-graining methods (the marginal preference analog). When the underlying model of preference makes incorrect assumptions, force matching is more likely to reveal the problem and less likely to lead research into confident misdiagnoses.

3.1 Motivation

The quality of a coarse-grained model is often defined by its ability to reproduce the statistical distribution of coarse-grained degrees of freedom from a fine-grained model it is meant to represent.[207, 276, 296, 285, 246] However, targeting a specific distribution directly by iteratively updating the potentials conjugate to that distribution, as in the most accurate structure-based coarse-graining algorithms,[207, 276, 296] has important drawbacks. The first is that finding such a match does not indicate whether or not the coarse-graining scheme is reasonable; the structure-based method may provide an excellent reproduction of the target distributions but achieve that match through a different physical mechanism than a more complete model would.[185, 286] In other words, it is not diagnostic. The second is that targeting a probability is targeting a quantity that, through normalization, is dependent on global properties of the system; changing one probability requires changing another–often many others.[207, 276] In other words, structure matching is nonlocal.

The fact that structure-matching methods are not diagnostic and are nonlocal has motivated the development of another class of methods that is both local and diagnostic, the

force-matching methods. In force-matching methods, a coarse-grained model is not made by targeting the values of a specific reduced distribution, but rather the gradient of the logarithm of the total distribution. By targeting the gradient of the logarithm, the normalization factor for the distribution is rendered irrelevant and the method becomes local. Additionally, force-matching methods are easy to cast in ways that are diagnostic– one can hypothesize a model potential which is not conjugate to the target distributions, attempt to match the forces, and show that the candidate potential is unable to describe the coarse-grained system adequately.[185, 286, 371] Furthermore, the ways in which the model fails can be used to design improved models.[185, 371, 318, 74, 21, 147, 317]

Being diagnostic gives force-matching methods an important advantage in coarse-grained model development and interpretation. Instead of targeting a distribution and picking from the coarse-grained potentials conjugate to those distributions as in a structure-based method, in force-matching methods one first chooses a criterion of quality which involves more than just the target distribution, for instance minimizing a variational residual[154, 213, 335, 308] or satisfying a generalized Yvon-Born-Green (gYBG) hierarchy equation,[74, 235, 248] and then one decides on the particular functional form for trial coarse-grained potentials. Criteria are chosen such that they ensure that a coarse-grained potential model perfect with respect to these criteria gives rise to a model with statistics entirely consistent with those of their counterparts in the fine-grained model, but also so as not to overly privilege fitting the target distributions conjugate to the test potential when choosing among imperfect models. Thus, the fitting procedure leaves room for the model to fail to reproduce those target distributions if it has been poorly chosen.

Locality, meanwhile, gives force-matching methods an important advantage when coupling coarse-grained and fine-grained simulations, namely, that only the forces for visited coarse-grained configurations are needed and thus that the coarse-grained force field can be generated on the fly, visited configuration by visited configuration, as in Car-Parrinello *ab initio* molecular dynamics.[62] By using force matching, one can couple a fine-grained

simulation to a coarse-grained one without requiring any explicit definition for the coarse-grained force field, and without requiring any structure-bias-removal steps when the coupled simulation is used for free energy calculations.[281, 332, 214]

The definition of a force-matching approach as given above, matching gradients of a distribution rather than its values directly, is quite general and is well established in the literature. [64, 112, 306, 80] This chapter will discuss general considerations of the application of force matching to coarse-graining of molecular systems, in which two recently developed methods, the multiscale coarse-graining (MS-CG) method[154, 153, 246, 247] and temperature accelerated molecular dynamics (TAMD),[213, 4, 212] eloquently demonstrate how to use force matching as a practical, efficient, and systematic approach to the coarse-graining of chemical systems.

The MS-CG method is a variational force-matching procedure first performed by Izvekov and Voth for lipid bilayers and liquid-state systems.[154, 153, 155, 156] The essential idea of the method consists of two steps. First, one constructs a quadratic residual describing the extent to which the forces in a candidate coarse-grained model deviate from the same forces calculated in an all-atom simulation. Next, one minimizes the residual by solving a system of linear equations obtained by differentiating the residual with respect to all parameters in the test force field. Solving these equations results in a set of coarse-grained force field parameters that, if the space of test force fields is complete, will ensure perfect consistency between the two models.[246] The method was motivated as the transposition of a force matching method to define molecular mechanics force fields from *ab initio* data developed by Ercolessi and Adams[112] into a new problem space.

Though the MS-CG variational principle was initially postulated without careful justification, work by Noid et al. soon showed that it had roots in rigorous statistical mechanics and a connection to integral equation theories of the liquid state.[248, 246] Further work has refined the numerical implementation of the method, leading to major improvements in performance.[21, 247, 196, 203, 201] In particular, using smoother splines to represent single-

parameter components of coarse-grained force fields can turn the force-matching equations into a simpler linear least squares problem and regularization and use of specialized solvers can improve the stability and condition number of the force-matching problem.[196, 201] Additionally, while early work focused on modeling coarse-grained systems using only isotropic pair potentials and one-parameter bonded interactions in the canonical ensemble, more recent work has introduced new basis sets and residuals for three-body interactions,[185] isobaric-isothermal ensembles,[95] and for other extended ensembles,[234] including ensembles with distinguishable substates that cannot be represented in terms of coarse-grained particle configurations.[90]

Challenges remain, particularly in developing transferable potentials,[147, 317, 234, 94, 177, 186] overcoming the problem of sampling enough coarse-grained configurations in a fine-grained simulation to ensure that all configurations are described by a reasonable force field,[147, 317, 196, 186, 98] and determining which basis sets and mappings will result in effective models.[348, 371, 195, 21, 203, 147, 95, 185, 186, 96, 97]

Temperature accelerated molecular dynamics (TAMD), on the other hand, is an enhanced sampling method meant to enable researchers to explore coarse-grained phase space[214, 4] or compute coarse-grained potentials of mean force[213, 212, 231] efficiently, without requiring either a discretization of phase space as in umbrella sampling[172] or a uniform sampling of coarse-grained phase space as in metadynamics.[27] TAMD accomplishes this by sampling from the coarse-grained potential of mean force at the physical temperature as if at a higher effective temperature. Enforcing a separation of time scales between the evolution of the coarse-grained variables and the fine-grained variables, the force coarse-grained variables feel on their time scale of evolution is the mean force for their instantaneous configuration at the physical temperature of interest. After a fictitious-temperature thermostat is added to the coarse-grained dynamics, they will explore the PMF of the coarse-grained variables at whatever fictitious temperature has been chosen.

It is very similar in spirit to the older adiabatic free energy dynamics (AFED)[281] and

constrained adiabatic free energy surface (CAFES)[332] methods and has helped inspire the newer driven adiabatic free energy dynamics (d-AFED) method.[5] While enforcing the separation of time scales acts to decrease sampling, increasing the temperature more than makes up for that effect so that the method results in order of magnitude improvements in computational efficiency. Furthermore, later work has demonstrated that the steps of sampling and mean force computation can be separated to result in an efficient, embarrassingly parallel method that allows reconstruction of an accurate PMF by means of variational force matching from mean forces sampled from a poorly-defined ensemble.[213]

This chapter will introduce the theoretical and numerical aspects of force matching methods used to study coarse-graining of structureless point particle systems to coarser structureless point particle systems and discuss important results demonstrating their strengths and weaknesses. Section II will cover the statistical mechanics behind these methods, deriving the force matching equations and two variational principles in the isothermal-isochoric (canonical) ensemble and the isothermal-isobaric ensemble. Section III will give an overview of techniques used to implement force matching in practice, including techniques that represent necessary practical modifications to the pure variational theory to account for known defects. Section IV will then discuss the application of MS-CG to the coarse-graining of liquids and peptides and TAMD to the coarse-graining of proteins. Finally, Section V concludes by summarizing the previous sections and suggesting future directions of research.

3.2 Force Matching in the Abstract

3.2.1 Model definitions in the canonical ensemble

Following the work of Noid et al.,[246] we consider two levels of detail, first a fine-grained system of n structureless point particles with Cartesian positions $\mathbf{r}^n$ and momenta $\mathbf{p}^n$ and

a separable Hamiltonian

$$h(\mathbf{r}^n, \mathbf{p}^n) = t(\mathbf{p}^n) + u(\mathbf{r}^n) \tag{3.1}$$

with t the kinetic energy and u the configurational energy, second a coarse-grained system of N structureless point particles with Cartesian positions $\mathbf{R}^N$ and momenta $\mathbf{P}^N$, and free energy

$$H(\mathbf{R}^N, \mathbf{P}^N) = T(\mathbf{P}^N) + U(\mathbf{R}^N) \tag{3.2}$$

with T and U representing kinetic and configurational free energies.

Assuming that the two models propagate ergodically with constant temperature $T = k_B^{-1}\beta^{-1}$, they are characterized by separable probability distributions

$$p(\mathbf{r}^n, \mathbf{p}^n) = p_p(\mathbf{p}^n)p_r(\mathbf{r}^n) \sim e^{-\beta t(\mathbf{p}^n)}e^{-\beta u(\mathbf{r}^n)} \tag{3.3}$$

and

$$P(\mathbf{R}^N, \mathbf{P}^N) = P_P(\mathbf{P}^N)P_R(\mathbf{R}^N) \sim e^{-\beta T(\mathbf{P}^N)}e^{-\beta U(\mathbf{R}^N)} \tag{3.4}$$

for the fine-grained and coarse-grained models, respectively. Finally, each model is endowed with a gradient with respect to all configurational degrees of freedom, $\nabla_r = \partial/\partial\mathbf{r}^n$ and $\nabla_R = \partial/\partial\mathbf{R}^N$.

With both models defined, one next defines a linear transformation $M : \mathbf{r}^n \to \mathbf{R}^N$ defined so that $\mathbf{R}_I = \sum_i c_{Ii} r_i \equiv c_I^n \circ \mathbf{r}^n$ for some set of coefficients c_{Ii} for each I. To ensure that both models behave the same under translations of the full system, one requires that $|c_I^n|_1 \equiv \sum_i c_I i = 1$ for all I. While many derivations of MS-CG consistency further specify that each particle in the fine-grained system must be involved in the definition of only at

most one coarse-grained position in this transformation, and the rest have assumed that each coarse-grained position has at least one fine-grained particle specific to it, this chapter's derivation will relax those assumptions and require only that M have a right inverse. In practice, the most common mapping operators define each coarse-grained position as the center of mass or center of geometry of a group of fine-grained particles, with all groups mutually exclusive. Relaxing the assumptions on the mapping to only require that M have a right inverse can be thought of as allowing the groups to overlap instead of being mutually exclusive, though still requiring that the groups' coefficients are linearly independent. With this mapping operator defined, it becomes simple to write down the equation for consistency between the two models in configuration space: $P_R(\mathbf{R}^N) = \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) p_r(\mathbf{r}^n)$.

The next section will describe how this consistency in terms of probabilities can be rewritten in terms of forces; at this point it is worth mentioning that the derivation will be only a special case of a much more general relationship between structure-matching and force-matching that also holds for nonlinear mappings between non-Cartesian coordinate systems. The reader interested in more general coarse-graining schemes is encouraged to examine the work of Cicotti et al.[80] or Wong and York[361] for a thorough account of important generalizations that are beyond the scope of this chapter.

3.2.2 Force matching from distribution consistency

As described in the introduction, attempting to fit distributions directly has important drawbacks which have motivated the development of an alternative set of consistency equations written in terms of forces. This section will go through a derivation of that alternative set for the models defined in section IIA, starting from the configurational consistency equation

$$P_R(\mathbf{R}^N) = \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) p_r(\mathbf{r}^n) \tag{3.5}$$

and ending with the force-matching equation

$$\nabla_R U(\mathbf{R}^N) = \langle M^+(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N} \tag{3.6}$$

where $\langle f(\mathbf{r}^n)\rangle_{\mathbf{R}^N} \equiv \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) p_r(\mathbf{r}^n) f(\mathbf{r}^n)$ for any f.

One begins by writing the consistency equation between configurational probabilities in terms of the energy,

$$Z^{-1} e^{-\beta U(\mathbf{R}^N)} = z^{-1} \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)}, \tag{3.7}$$

then taking the logarithm of both sides to separate the multiplicative partition functions into an additive free energy offset,

$$U(\mathbf{R}^N) = -\beta^{-1} \ln \left[\int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)} \right] + \beta^{-1} \ln(z/Z), \tag{3.8}$$

and next taking the gradients of both sides to remove that constant, nonlocal factor, leaving a local equation,

$$\nabla_R U(\mathbf{R}^N) = -\beta^{-1} \int d\mathbf{r}^n \nabla_R \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)} / Z_{\mathbf{R}^N}, \tag{3.9}$$

where $Z_{\mathbf{R}^N} \equiv \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)}$. Making the mapping operator explicit, one has

$$\frac{\partial}{\partial \mathbf{R}_I} \delta(c_I^n \circ \mathbf{r}^n - \mathbf{R}_I) = -\frac{1}{c_{Ii}} \frac{\partial}{\partial \mathbf{r}_i} \delta(c_I^n \circ \mathbf{r}^n - \mathbf{R}_I) \tag{3.10}$$

for any nonzero c_{Ii}, and more generally,

$$\frac{\partial}{\partial \mathbf{R}_I} \delta(c_I^n \circ \mathbf{r}^n - \mathbf{R}_I) = -\sum_i \frac{d_{Ii}}{c_{Ii}} \frac{\partial}{\partial \mathbf{r}_i} \delta(c_I^n \circ \mathbf{r}^n - \mathbf{R}_I) \tag{3.11}$$

for any set of nonzero c_{Ii} if $\sum_i d_{Ii} = 1$ over the i with nonzero c_{Ii}, which implies

$$\frac{\partial}{\partial \mathbf{R}_I} U(\mathbf{R}^N) = -\beta^{-1} \int d\mathbf{r}^n \left(-\sum_i \frac{d_{Ii}}{c_{Ii}} \frac{\partial}{\partial \mathbf{r}_i} \delta(M(\mathbf{r}^n) - \mathbf{R}^N) \right) e^{-\beta u(\mathbf{r}^n)} / Z_{\mathbf{R}^N}, \tag{3.12}$$

which can be integrated by parts to

$$\frac{\partial}{\partial \mathbf{R}_I} U(\mathbf{R}^N) = -\beta^{-1} \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) \left(\sum_i \frac{d_{Ii}}{c_{Ii}} \frac{\partial}{\partial \mathbf{r}_i} e^{-\beta u(\mathbf{r}^n)} \right) / Z_{\mathbf{R}^N}, \tag{3.13}$$

so long as

$$\sum_i \frac{d_{Ii}}{c_{Ii}} \frac{\partial}{\partial \mathbf{r}_i} \delta(c_J^n \circ \mathbf{r}^n - R_J) = 0 \text{ for all } I \neq J \tag{3.14}$$

or, more compactly,

$$\sum_i \frac{d_{Ii}}{c_{Ii}} c_{Ji} = \delta_{IJ}. \tag{3.15}$$

This system of equations implies that the transpose of the d_{Ii}/c_{Ii} matrix is a right inverse of M, so call it M^+. So as long as M has a right inverse, then, the integration by parts can be done and the result can be written compactly as

$$\nabla_R U(\mathbf{R}^N) = \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) \left(M^+(\nabla_r u(\mathbf{r}^n)) \right) e^{-\beta u(\mathbf{r}^n)} / Z_{\mathbf{R}^N}, \tag{3.16}$$

$$\nabla_R U(R^N) = \langle M^+(\nabla_r u(\mathbf{r}^n)) \rangle_{\mathbf{R}^N}, \tag{3.17}$$

where, as a reminder, $\langle f(\mathbf{r}^n) \rangle_{\mathbf{R}^N} \equiv \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) p_r(\mathbf{r}^n) f(\mathbf{r}^n)$ for any f.

This is the end point of the derivation, the force-matching equation, and it defines the coarse-grained potential uniquely up to an additive constant that is irrelevant for calculating any probability distribution for the model. The reader is invited to verify that this is a special case of formula 4 of Cicotti et al.[80] for constant $b_j(x)$.

3.2.3 Variational force matching

In the previous section, the force-matching equations were treated as an exact formula for the many-body coarse-grained potential of mean force. They could be used directly because they were applied to the calculation of reaction pathways and trajectories: one-dimensional curves in higher-dimensional spaces. As the dimensionality of the desired PMF increases, the cost of computing the surface from these exact formulas becomes exorbitant and impractical. An alternative, approximate method is needed is to interpolate sparsely, inhomogeneously sampled forces into smooth PMFs when they are too expensive to compute directly. A prominent approach in the literature is to use variational formulations of the exact equations.

This section introduces two closely-related force-matching variational principles, first the MS-CG variational principle of Voth and colleagues,[246, 154] and second the later variational principle of Maragliano and Vanden-Eijnden.[213] (Though see Stecher's comments.[308])

The MS-CG variational method initially proposed by Izvekov and Voth[154] consists in finding the coarse-grained force field $\mathbf{G}^N(\mathbf{R}^N) = \nabla_R W(\mathbf{R}^N)$ (where $W(\mathbf{R}^N)$ is the corresponding potential of mean force) that minimizes a force residual among all trial force fields; that residual is

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \frac{1}{3NT}\sum_{t_k}^{T}\sum_{I} ||\mathbf{G}_I(M(\mathbf{r}^n(t_k))) - (d_I/c_I)^n \circ \nabla_r u(\mathbf{r}^n(t_k))||^2, \tag{3.18}$$

where configurations $\mathbf{r}^n(t_k)$ are sampled from the canonical ensemble of the fine-grained model using an ergodic dynamics, and T is the total number of samples. In the long-time limit this is

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle ||\mathbf{G}_I(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))||^2 \rangle. \tag{3.19}$$

Splitting the expectation of the residual over $u(\mathbf{r}^n)$ into an expectation over $U(\mathbf{R}^N)$ of the

expectations of the residual over $u(\mathbf{r}^n)$ at fixed $\mathbf{R}^N$, this becomes

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle\langle||\mathbf{G}_I(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))||^2\rangle_{\mathbf{R}^N}\rangle. \tag{3.20}$$

Now introduce the true force field defined by

$$\mathbf{F}^N(\mathbf{R}^N) = \langle M^+(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N} \tag{3.21}$$

so that

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle\langle||\mathbf{G}^N(M(\mathbf{r}^n)) - \mathbf{F}^N(M(\mathbf{r}^n)) + \mathbf{F}^N(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))||^2\rangle_{\mathbf{R}^N}\rangle \tag{3.22}$$

$$\begin{aligned}\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = {} & \langle||\mathbf{G}^N(M(\mathbf{r}^n)) - \mathbf{F}^N(M(\mathbf{r}^n))||^2\rangle \ \ldots \\ & \ldots \ + \langle\langle\mathbf{G}^N(M(\mathbf{r}^n)) - \mathbf{F}^N(M(\mathbf{r}^n))\rangle_{R^N} \ \ldots \\ & \ldots \cdot \langle\mathbf{F}^N(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N}\rangle \ \ldots \\ & \ldots \ + \langle||\mathbf{F}^N(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))||^2\rangle\end{aligned} \tag{3.23}$$

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle||\mathbf{G}^N(\mathbf{R}^N) - \mathbf{F}^N(\mathbf{R}^N)||^2\rangle + \chi^2[\mathbf{F}^N(\mathbf{R}^N)] \tag{3.24}$$

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] \geq \chi^2[\mathbf{F}^N(\mathbf{R}^N)], \tag{3.25}$$

proving that the residual does in fact give rise to a variational principle. Note that this derivation works for any two-norm $||\cdot||^2$, though the two-norm used in practice has, in every case but one, been the norm induced by an unweighted dot product. Later in this text, for residuals in other ensembles, it will be natural to consider a residual that could be written as

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle\Big(\mathbf{G}^N(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))\Big) \cdot W \cdot \Big(\mathbf{G}^N(M(\mathbf{r}^n)) - M^+(\nabla_r u(\mathbf{r}^n))\Big)\rangle \tag{3.26}$$

for some arbitrary, symmetric positive-definite N by N matrix W.

The Maragliano and Vanden-Eijnden force-matching variational principle,[213] on the other hand, refers to a simulation of a coarse-grained model and the mean forces at the visited coarse-grained configurations. One aims to find a trial force field $\mathbf{G}^N(\mathbf{R}^N)$ that minimizes the residual

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \frac{1}{3NT}\sum_{t_k}^{T}\sum_{I}||\mathbf{G}_I(M(\mathbf{r}^n(t_k))) - \langle(d_I/c_I)^n \circ \nabla_r u(\mathbf{r}^n)\rangle_{\mathbf{R}^N(t_k)}||^2, \quad (3.27)$$

where the $\mathbf{R}^N(t_k)$ are sampled from an arbitrary, possibly unknown distribution $P_R(\mathbf{R}^N) \sim e^{-\beta V(\mathbf{R}^N)}$ for an unknown V. As before, the residual has a minimum for the true force field. In the long-time limit, one has

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle||\mathbf{G}^N(\mathbf{R}^N) - \langle M^+(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N}||^2\rangle_V(R^N) \quad (3.28)$$

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \langle||\mathbf{G}^N(\mathbf{R}^N) - \mathbf{F}^N(\mathbf{R}^N)||^2\rangle_{V(\mathbf{R}^N)} \quad (3.29)$$

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] \geq 0 = \chi^2[\mathbf{F}^N(\mathbf{R}^N)] \quad (3.30)$$

The minimum is unique so long as $V(\mathbf{R}^N)$ is positive almost everywhere and bounded almost everywhere. Other residuals with a similar form can be justified in the same way and have shown superior performance by some metrics.[231]

MS-CG has the important advantage that, though the coarse-grained configuration specific expectations enter into the theory, they need not be evaluated directly in practice. It has an important disadvantage in that it requires sampling from a canonical ensemble of the fine-grained coordinates. Each entails a cost, and the relative difficulty of sampling the canonical ensemble versus performing the constrained expectations will determine whether MS-CG or the Maragliano and Vanden-Eijnden method is more efficient for each particular application.

3.2.4 Transferability, other ensembles, auxiliary variables

The previous sections described the derivation of coarse-grained potentials for a single state point in a single class of ensembles, an isothermal-isochoric (NVT) ensemble. However, it is desirable to define state-dependent, transferable coarse-grained force fields and force fields for alternative ensembles as well. The cases that have received the most attention in the literature are transferability across temperature[147, 177, 204] and coarse-graining in the isothermal-isobaric (NPT) ensemble.[95] Transferability across composition changes in mixtures has also received a great deal of attention,[203, 234, 94, 186, 61] but will remain outside of the scope of this chapter.

Transferability has typically been examined in one of two ways. Either coarse-grained models are developed at different state points and then compared, or a single coarse-grained model is compared to the fine-grained model at multiple state points. The first has been accomplished by brute-force model development for each state point,[204] which is a straightforward application of previously discussed ideas, and also by trajectory reweighting schemes at one state point from a simulation at another by using a modified residual.[177] For temperature, one develops a model at state point NVT' from a simulation at NVT using the MS-CG-type residual

$$\chi^2[\mathbf{G}^N(\mathbf{R}^N)] = \frac{\sum_{t_k}^{T} e^{-(\beta'-\beta)u(\mathbf{r}^n(t_k))} \sum_I ||\mathbf{G}_I(M(\mathbf{r}^n(t_k))) - (d_I/c_I)^n \circ \nabla_r u(\mathbf{r}^n(t_k))||^2}{3N \sum_{t_k}^{T} e^{-(\beta'-\beta)u(\mathbf{r}^n(t_k))}} \tag{3.31}$$

that converges to the usual MS-CG $\chi^2[\mathbf{G}^N(\mathbf{R}^N)]$ for the NVT' ensemble in the long-time limit. Accurate calculation of an exponential expectation requires excellent sampling, though, limiting the applicability of the method. In practice, one must typically choose $T > T'$ to ensure that all configurations relevant at T' are sampled; however, this rule of thumb will fail in many systems, especially those that order upon heating. In the second method to demonstrate transferability, one derives a coarse-grained force field and then sim-

ulates it at a different state point, a procedure that is inherently ill-defined. This is easily demonstrated in the example of temperature transferability: here, one typically uses a T-derived force field to drive dynamics subject to a T' thermostat. Effectively, one is treating the coarse-grained force field as if it were an internal energy surface. One could as easily scale the T-derived force field by T'/T and use that to drive the dynamics, treating it as if it were an entirely entropic surface. Despite this drawback, the relative ease of this second method has helped it see use; a later chapter will explore this in more detail.

Coarse-graining in other ensembles is a problem closely related to transferability, since the force field of one ensemble, say NPT, can be written as an interpolation between force fields of another ensemble, in this case NVT, over a range of the conjugate variables that have been exchanged between the models, in this case V. For the sake of concreteness, this chapter will begin with a discussion of coarse-graining in the NPT ensemble and then discuss, briefly, the methods used to derive MS-CG variational principles for more general ensembles.

When deriving coarse-graining formulas in the NPT ensemble, one considers the same fine-grained and coarse-grained models as before, with the same Hamiltonians, only now the fine-grained model also has a variable volume v and the coarse-grained model has a variable volume V. The sets of positions are scaled so that the fine-grained particles are described by coordinates $\mathbf{x}_i = v^{-1/3}\mathbf{r}_i$ and the coarse-grained particles are described by coordinates $\mathbf{X}_i = V^{-1/3}\mathbf{R}_i$. Next, the fine-grained Hamiltonian is assumed to be translation invariant, isotropic, and dependent on v only through an isotropic scaling of the coordinates, $u(\mathbf{x}^n; v) = u(v^{1/3}\mathbf{x}^n)$. Finally, the mapping operator M is extended to send $v \to V$. Then, defining the full set of fine-grained variables and the full set of coarse-grained variables, we can write the per-configuration Gibbs-like free energy in the fine-grained model as

$$g(\gamma^{n+1}) = u(v^{1/3}\mathbf{x}^n) - pv - \beta^{-1} n \log v \tag{3.32}$$

and the per-configuration energy in the coarse-grained model as

$$G(\Gamma^{N+1}) = U(V^{1/3}\mathbf{X}^N; V) - pV - \beta^{-1} N \log V. \tag{3.33}$$

Each model has a configuration and volume probability distribution from these weights,

$$p_\gamma(\gamma^{n+1}) \sim e^{-\beta g(\gamma^{n+1})} \tag{3.34}$$

$$P_\Gamma(\Gamma^{N+1}) \sim e^{-\beta G(\Gamma^{N+1})}. \tag{3.35}$$

One next defines a map between the models analogous to the original MS-CG map,

$$M : \gamma^{n+1} \to \Gamma^{N+1}, \text{ such that} \tag{3.36}$$

$$\text{for all } I \leq N,\ \Gamma_I = \mathbf{X}_I = \sum_i c_{Ii}\mathbf{x}_i = \sum_i c_{Ii}\gamma_i \tag{3.37}$$

$$\text{and } \Gamma^{N+1} = V = v = \gamma^{n+1}, \tag{3.38}$$

so that consistency is simply

$$P_\Gamma(\Gamma^{N+1}) \sim \int d\gamma^{n+1} \delta(M(\gamma^{n+1}) - \Gamma^{N+1}) p_\gamma(\gamma^{n+1}) \tag{3.39}$$

$$e^{-\beta G(\Gamma^{N+1})} \sim \int d\gamma^{n+1} \delta(M(\gamma^{n+1}) - \Gamma^{N+1}) e^{-\beta g(\gamma^{n+1})}, \tag{3.40}$$

and, defining $\nabla_\gamma = \partial/\partial\gamma^{n+1}$ and $\nabla_\Gamma = \partial/\partial\Gamma^{N+1}$, then after performing steps precisely analogous to those in the derivation of canonical-ensemble force matching, one finds

$$\nabla_\Gamma G(\Gamma^{N+1}) = \langle M^+(\nabla_\gamma g(\gamma^{n+1}))\rangle_{\Gamma^{N+1}}, \tag{3.41}$$

a force-matching equation for a Gibbs-like work function rather than a Helmholtz-like work function. Also analogously to the previous section, this gives rise to an MS-CG-like varia-

tional principle for test force fields $H^{N+1}(\Gamma^{N+1})$,

$$\chi^2[H^{N+1}(\Gamma^{N+1})] = \frac{1}{3T}\sum_{t_k}^{T}\sum_{IJ}\left(\mathbf{H}_I(M(\gamma^{n+1}(t_k))) - (d_I/c_I)^n \circ \nabla_\gamma u(\gamma^{n+1}(t_k))\right) W_{IJ} \ldots$$
$$\ldots * \left(\mathbf{H}_I(M(\gamma^{n+1}(t_k))) - (d_I/c_I)^n \circ \nabla_\gamma u(\gamma^{n+1}(t_k))\right) \tag{3.42}$$

where W is some symmetric positive-definite $N+1$ by $N+1$ matrix. Choosing

$$W_{IJ} = \delta_{IJ}/N \quad \text{for } I, J \leq N, \text{ and } W_{N+1,N+1} = a \tag{3.43}$$

results in the variational principle derived by Das and Andersen.[95] It should also be clear how to construct a residual in the style of Maragliano and Vanden-Eijnden from these force-matching equalities. Clearly, trial force fields for this residual must be constructed to allow dependence on both V and the $\mathbf{X}_I$, but so long as they are and the trial functions are linear combinations of basis functions, the same algorithms to be discussed in the coming sections for canonical variational force matching can be applied to the optimization.

Though the procedure outline above is specific to the concrete case of the NPT ensemble, the general procedure readily generalizes to ensembles including external fields coupled to other degrees of freedom than volume, to anisotropically stressed systems, and to the case in which the fine-grained free energy surface is not perfectly transferable. A summary of the general procedure would be: add new variables, perform a change of variables to remove coupling, define a linear map between the full sets of newly decoupled variables, define distribution-based consistency between the distributions of these variables in the two models, take the gradient with respect to all of the new variables of the log of that consistency equation, integrate by parts using a right-inverse of the mapping to obtain an equation for perfect force matching, and then define a generic quadratic residual describing the deviation from perfect force matching. However, the difficulty of defining appropriate basis sets for these problems and performing variational optimization for these more-complex models

impedes progress in this direction.

3.3 Implementation of Force Matching

3.3.1 Direct use of the exact force matching equations

As described earlier, this definition has advantages when defining a coarse-grained force field on the fly, for low-dimensional subspaces, and from poorly sampled or ill-defined coarse-grained configuration ensembles. To use these equations in this way, however, requires numerical schemes for evaluating the conditional expectations $\langle M^{+}(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N}$, an old problem which has gone by the name of 'Blue Moon sampling' in the literature.[80]

These are calculated using either harmonic (soft) or holonomic (hard) constraints to keep $\mathbf{R}^N$ fixed. For the general case where coarse-grained variables correspond to nonlinear transformations of the fine-grained variables, these two methods of calculating the constrained expectation differ even in the limit of infinitely stiff harmonic constraints.[123] However, when the Jacobian of the transformation is constant, as it is for the linear transformations used in the coarse-graining schemes we consider here, many subtleties associated with using general coordinate systems disappear. In the end it is sufficient to use stiff springs and to average the restoring force of the springs:

$$\langle M^{+}(\nabla_r u(\mathbf{r}^n))\rangle_{\mathbf{R}^N} = \lim_{\kappa\to\infty} \langle \kappa \cdot (\mathbf{R}_I - c_I^n \circ \mathbf{r}^n\rangle_{\mathbf{R}^N} \tag{3.44}$$

where κ is the stiffness of the harmonic restraint. The actual values of κ required for the restraint to be effectively infinitely stiff will be problem-dependent, depending strongly on the temperature and the curvature of the underlying fine-grained potential. Sampling the restrained ensemble is typically less expensive than sampling the full model, but can prove more difficult if the coarse-grained configurations hide multiple metastable fine-grained states which are separated on the given coarse-grained configuration hypersurface, but connected

in the full fine-grained model. In this case, multistate potentials of mean force may become necessary in practice.[90]

A local method is ideally suited for calculating the PMF of the coarse-grained model along a single arbitrary curve, since it requires no information about points surrounding the curve and no discretization of the higher-dimensional space the curve lies in. Force matching has been used extensively for this purpose, but these applications lie beyond the scope of this chapter since they typically deal with reaction coordinates rather than coarse-grained particulate models. Force matching is also employed when a trajectory for a simulation is needed and the force field can be actively generated as the simulation progresses. An excellent example of this is the TAMD method introduced by Maragliano and Vanden-Eijnden. TAMD can be seen as an extended Lagrangian method for simulating a coarse-grained system in which an entire copy of the fine-grained system is added to the coarse-grained dynamics.

With the motion of the $\mathbf{R}^N$ very slow to ensure adiabatic separation between the two subsystems, the fine-grained system will essentially equilibrate at fixed $\mathbf{R}^N$ and the force on the $\mathbf{R}^N$ will be the averaged restraint force: from the previous subsection's discussion, this is the expected mean force. Thus, TAMD can generate trajectories defined with respect to the unique, consistent $U(\mathbf{R}^N)$ without requiring an explicit definition for it.

In particular, driving the coarse-grained and fine-grained variables with different effective temperatures allows the dynamics to explore of the surface $U(\mathbf{R}^N)$ from a fine-grained model at temperature $T_F G$ at any desired temperature $T_C G$. When using this sampling scheme, care must be taken to ensure that the time scale of the fine-grained system's relaxation is always faster than that of the coarse-grained motion, which is a highly problem-dependent constraint. It is necessary to design one's coarse-graining to ensure that the restrained fine-grained ensemble has a reasonable relaxation spectrum and to tailor the TAMD parameters to this relaxation; Maragliano and Vanden-Eijnden provide a discussion of the relevant asymptotics in their paper introducing the method.[214]

3.3.2 Trial force fields for variational force matching

The variational methods introduced in previous sections have the form of least squares problems. The form of the trial force fields determines whether they appear as linear or nonlinear least squares problems. If the optimizable parameters of the trial force fields are the coefficients in a linear combination of basis functions, the problem will be a linear least-squares problem. Otherwise, the problem will be nonlinear. Because linear least squares problems are much easier to solve, this chapter will consider that case exclusively, and because MS-CG has been experimented with most among the variational force matching coarse-graining methods, this section will focus almost exclusively on the implementation of MS-CG.

In the first case, given a linear combination force field with force field parameters $\tilde{\phi} = \phi_i$ controlling the strength of individual interaction potentials $V_i(\mathbf{R}^N)$, one has

$$W(\mathbf{R}^N) = \sum_i \phi_i V_i(\mathbf{R}^N) \equiv \tilde{\phi} * \tilde{V} \tag{3.45}$$

In the second, general case, one has

$$W(\mathbf{R}^N) = V(\mathbf{R}^N; \phi_i). \tag{3.46}$$

To obtain a tractable optimization problem and an accurate PMF, the set of basis functions must be chosen to be flexible enough to represent the physics of the problem and yet not so flexible as to lead to an exorbitantly expensive optimization. For this reason, the basis set must be chosen carefully using physical intuition to include only the interactions one expects to be needed and as little extra as possible.[247, 201, 203]

In practice, the interactions are typically chosen to be smooth spline functions of a single parameter, linear combinations of polynomial functions of a single parameter, or separable functions of multiple parameters in which the dependence on all but one of the parameters is

known, and the remaining parameter is described with a spline fit. Examples are the use of B-splines to fit non-bonded short-range isotropic pairwise potentials,[203] use of truncated Taylor series to fit an angle-based three-body bonded interaction,[371] and the use of a three-body interaction with Stillinger-Weber[310] distance dependence multiplied by a spline-fit angular portion.[185] Another case worth noting in this pattern comes from Das and Andersen's work on force matching in the isothermal-isobaric ensemble.[95] They chose their basis set by assuming that the full Gibbs work surface could be reproduced using a part that depends on volume only through the scaling of the $\mathbf{X}_I$, $G(mathbf{R}^N, \phi_i)$, which they fit using typical trial functions for this part of the force field in a canonical ensemble, and a part that depends only on volume, $G(V, \psi_i)$, which they expanded as a Taylor series around the expected volume. An exception to this pattern is found in the work of Maragliano and Vanden-Eijnden, who used radial basis functions to fit PMFs in two and four dimensions instead of using one-parameter-at-a-time spline functions; however, this method quickly becomes impractical with increasing dimension.[213, 231, 308] In all cases, there are two essential steps to determining the form of a basis set: first, a list of interactions and their parameters is chosen, and then functional forms for each of these are chosen. As will be discussed later in the context of peptide modeling, current understanding of these steps is very limited, and choosing basis sets remains one of the major challenges in applying the MS-CG method.

3.3.3 Numerical algorithms for variational force matching

Once a basis set is chosen, the variational minimization can be written as a linear least-squares problem,

$$\tilde{\phi} = \arg\min_{\tilde{\phi}} \frac{1}{3NT} \sum_{t_k}^{T} \sum_{I} ||\mathbf{G}_I(M(\mathbf{r}^n(t_k))) - (d_I/c_I)^n \circ \nabla_r u(\mathbf{r}^n(t_k))||^2, \tag{3.47}$$

which can also be viewed as finding a solution of

$$\nabla_R \tilde{V}(t_k) * \tilde{\phi} = f(t_k) \text{ for all } t_k \tag{3.48}$$

in the least squares sense, which can be written as a matrix equation

$$\bar{\bar{F}} * \tilde{\phi} = \bar{f}, \tag{3.49}$$

where $\bar{\bar{F}}$ is a $3NTM$ matrix recording the relative activity of each basis function on each momentum component of each particle in each sampled frame, $\tilde{\phi}$ is the M-vector of force field parameters, and $\bar{f}$ is a $3NT$-vector representing each component of the instantaneous force on each particle at each sampled frame. Since a typical application will involve many coarse-grained particles and many sample frames, and the basis sets used to fit the PMF may involve hundreds or thousands of optimizable parameters, the matrix $\bar{\bar{F}}$ can be very large– too large to store in RAM or even too large to store in a single hard drive.[201] This causes difficulties for many standard least-squares implementations, but because $\bar{\bar{F}}$ is typically sparse and typically $3NT \gg M$, schemes based on block-averaging, sequential accumulation, the normal equations can avoid these difficulties.[201] A second closely related problem exists in that, for choices of basis set in which some interactions are rarely sampled, the equations are ill-posed and the condition number of $\bar{\bar{F}}$ may be large.[196, 201] This motivates the use of least-squares algorithms that are relatively insensitive to condition number and the use of regularization schemes to account for the poor sampling by introducing a model for the noise of each sample.

The three algorithms used to solve this equation, block-averaging, sequential accumulation, and conversion to normal form, have particular strengths and weaknesses. In the block-averaging approach, one splits the samples into blocks, and then calculates $\tilde{\phi}$ for each block, and then averages. This is not equivalent to the original problem,[154, 247] but approximates it if the blocks are sufficiently large. It has the advantage that each block's

matrix may be sparse with fewer than M^2 nonzero entries, making it potentially the least memory-intensive of all three methods. However, if this is true then not all interactions have been sampled and the blocks are almost certainly not long enough for the approximate solutions to be close to the actual solution.

Sequential accumulation, on the other hand, is accomplished by splitting the problem frame by frame and then combining the frame-wise $\bar{\bar{F}}$ matrices using a series of QR decompositions. The resulting problem has the same residual, nonzero singular values, and condition number as the original problem, but is far less memory intensive.

The normal form of the least-squares problem, finally, is

$$\bar{\bar{F}}^T \bar{\bar{F}} \tilde{\phi} = \bar{\bar{F}}^T \bar{f}. \tag{3.50}$$

$\bar{\bar{F}}^T \bar{\bar{F}}$ can be calculated blockwise or frame by frame, resulting in memory savings comparable to the sequential accumulation method, and it has similar advantages. However, the nonzero singular values are changed from the initial problem, causing the condition number of the problem to square.

Because the condition number for these equations can be very large, it has proven fruitful to consider using Tikhonov regularization to improve the behavior of these methods.[196, 201] This scheme is equivalent to using Bayesian inference to predict $\tilde{\phi}$ under the assumption that the target force data and parameters are subject to additive Gaussian noise. This prior has not been empirically verified and is unlikely to represent the full physics of the problem, but nonetheless improves some aspects of the fitting. Using regularization amounts to solving

$$(\bar{\bar{F}}^T \bar{\bar{F}} + \lambda \tilde{\bar{I}}) \tilde{\phi} = \bar{\bar{F}}^T \bar{f} \tag{3.51}$$

for the normal equations. In either case, the regularization parameter λ is chosen by examining the noise or the data or singular values of the matrix, for instance by using the L-curve method.[140]

3.3.4 Methods to circumvent common difficulties

While variational force-matching gives a unique, optimal answer in the long time limit, as in most optimizations which rely on partial data to construct an approximation of an objective function, it is common to see MS-CG overfit the input data when there are many free parameters to optimize and insufficient data to optimize them reliably or to give physically unreasonable answers when there is no data to optimize a parameter at all.[371, 21, 154, 153, 306, 196, 201, 203, 238] This has given rise to two types of common modification to the MS-CG method. The first is selectively reducing the sensitivity of the force-field optimization to noise whenever overfitting is observed.[196, 201, 203] The second is using interpolation or alternative methods to fill in portions of the force field that MS-CG is unable fit because they have not been sampled.[371, 21, 154, 153, 238]

The most prominent systematic scheme for reducing the sensitivity of variational force matching to sample noise is the use of Tikhonov regularization discussed in an earlier section. While in principle this can be fully systematic, the regularization parameters are often chosen ad hoc in practice.[196, 201] Many additional ad hoc smoothing schemes have been tried in the literature and many more have been attempted. One that is very common, but rarely mentioned in the final results of an MS-CG study, is to fit a force field using one basis set, find overfitting, and then reduce the size of the basis set to refit. Another which is commonly used, but not connected to clearly enunciated physical motivation, is performing a second round of fitting by matching less-flexible functions to the spline curve after the actual force matching is complete. One example is approximation by negative powers of the separation distance for isotropic pair non-bonded interactions.[153] A third modification involves responding to overfitting by choosing smoother basis functions, for instance higher-order B-splines,[203] to incorporate the assumption that noise in the data will be disproportionately high-frequency. The common thread is always the use of physical intuition to decide what is overfitting and what is not, and then modification of either the trial force field or the results to remove what has been judged undesirable. In each of these cases, one can imagine extending these ad hoc

tactics into a systematic strategy, and this will likely be one of the major challenges facing future research in the field.

Attempting to overcome a complete absence of data is a qualitatively different problem, and one that is, unfortunately, ubiquitous in coarse-grained modeling.[371, 21, 147, 153, 154, 201, 186] Whereas in structure-matching methods, if a configuration is not visited then it will be given an excessively high potential and will be avoided in simulations of the resulting coarse-grained model, in regularized variational force-matching methods, if a configuration is not visited then it will be given no force, thus no barrier, and could very well be sampled in a simulation of the coarse-grained simulation. In some ways, this can be an advantage–the errors of a force-matched model may be more apparent, so force-matched models can be superior diagnostic tools but in other cases, where the errors are inevitable and uninteresting, it will be a disadvantage. A good example is the hard-core repulsion in pair non-bonded interactions. It is rare to sample near-core configurations, so nave, regularized force-matched models will often have soft-core interactions unless the hard-core repulsion can be interpolated from the well-sampled regions of the interaction. This interpolation can be done by fitting functions which interpolate from the well-sampled region into the unsampled region to a variationally optimal fit.[154] Finally, the problem can also be corrected by assuming that a particular force field should act in the unsampled region and simply including that in every single trial function.[21] A regularized optimization will then return that potential unchanged in completely unsampled regions. These force fields for the poorly sampled regions are typically chosen using other coarse graining methods, especially established analytic methods[21, 19] or Boltzmann inversion.[286, 371, 318, 147, 317, 203, 287] The practice of using assumed analytical forms in this way has been referred to as the hybrid analytic-systematic (HAS) method in the literature.

The last method, assuming some part of the force field is the same in every trial force field, is of more general use as well. When a model involves many different interactions, it can be impractical to fit them all at once. Instead, interactions may be fit sequentially–

for instance, by optimizing solvent-solvent interactions before considering solute-solvent and solvent-solvent interactions.[155] Alternatively, when certain interactions are described poorly by MS-CG force fields, they can be replaced with potentials derived by other methods before fitting the remaining interactions with MS-CG, which is especially common for intramolecular interactions. These largely ad hoc and poorly understood sequential and combined-method optimizations lead naturally to the idea of systematic hybrid methods, which are now just beginning to be explored. See especially the papers of Ruhle et al.[286] and Das, Lu, and Andersen[98] showing that a combination of force-matching and structure-matching can create models that seem better than either is capable of creating alone; it remains to be seen if this is a normal case and if these hybrid methods can be given a clear physical basis and interpretation.

3.4 Exemplary Results

3.4.1 Homogenous molecular liquid models

Homogenous liquids play an important role in MS-CG theory. As test cases, they have well-understood physics and experimental properties, and their analysis can draw on decades of liquid-state theory. As practical models, coarse-grained liquids can serve as simplified solvents in more complicated models, providing a middle ground between fully explicit and fully implicit solvent models. This subsection will briefly discuss two cases, methanol[154, 247, 155, 204] and water,[185, 247, 154, 201, 177] though many others, including ethanol,[74, 287] propane,[287] hexane,[286, 287] ionic liquids,[247, 346, 163] and homogeneous mixtures[238, 94, 195, 236, 61] have also been studied.

Methanol serves as an excellent example of a system amenable to force matching. The radial distribution function of a one-site center-of-mass MS-CG model of OPLS-AA methanol is in excellent agreement with that of the fine-grained model at 300 and 310 K,[154, 204] as illustrated for 300K in Figure 3.1, despite containing no explicit information about the

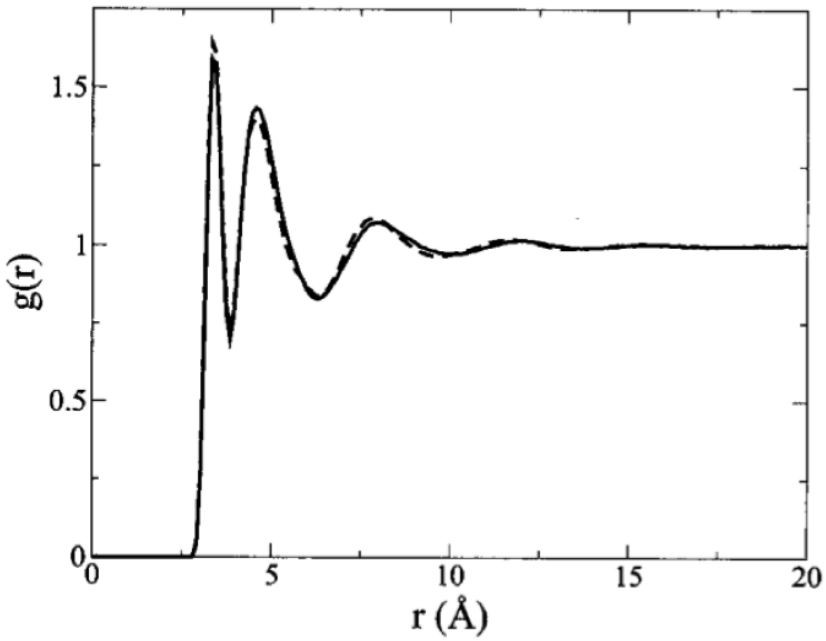

Figure 3.1: Comparison of atomistic (solid) and MS-CG (dashed) predictions for the radial distribution function of centers of mass of OPLS-AA methanol molecules at 300 K. Reproduced from Noid et al.[247]

dipole-dipole correlations that give rise to the attractive well of the interaction. This is expected, as the phase diagram of methanol is very much like that of a simple liquid and so methanol should be described well by a simple liquid model.

Water, on the other hand, has a more complex phase diagram, and thus it should be no surprise that accurate reproduction of fine-grained center-of-mass radial distribution functions for water requires a more complex model than a one-site model with only isotropic pair potentials. The history of one-site MS-CG water models is an excellent example of how force matching can be used as a diagnostic tool to reveal flaws in the assumptions of a model. Both iterative Boltzmann and inverse Monte Carlo pair potentials for a one-site coarse-grained model of SPC/E water[39] at 300 K define coarse-grained models that accurately reproduce the radial distribution function of the fine-grained model.[287] However, force-matched potentials lead to a coarse-grained model with far too little structure in the radial distribution function.[185, 287] The first two methods, however, do not accurately reproduce

the three-body correlations observed in the fine-grained system, suggesting that regardless of the success they seem to have in reproducing their target structural properties, they do not accurately capture the physics of the system. This is confirmed by an MS-CG approach with Stillinger-Weber-like three-body terms to the trial force fields; adding these results in a coarse-grained model with the proper amount of structure in the radial distribution function, as shown in Figures 3.2 and 3.3. This confirms that some important physics was missed in the previous model, and shows that a combination of two- and three-body interactions is sufficient to describe the physics behind the first few solvation shells of molecular center of mass pair and angular correlations in SPC/E water.

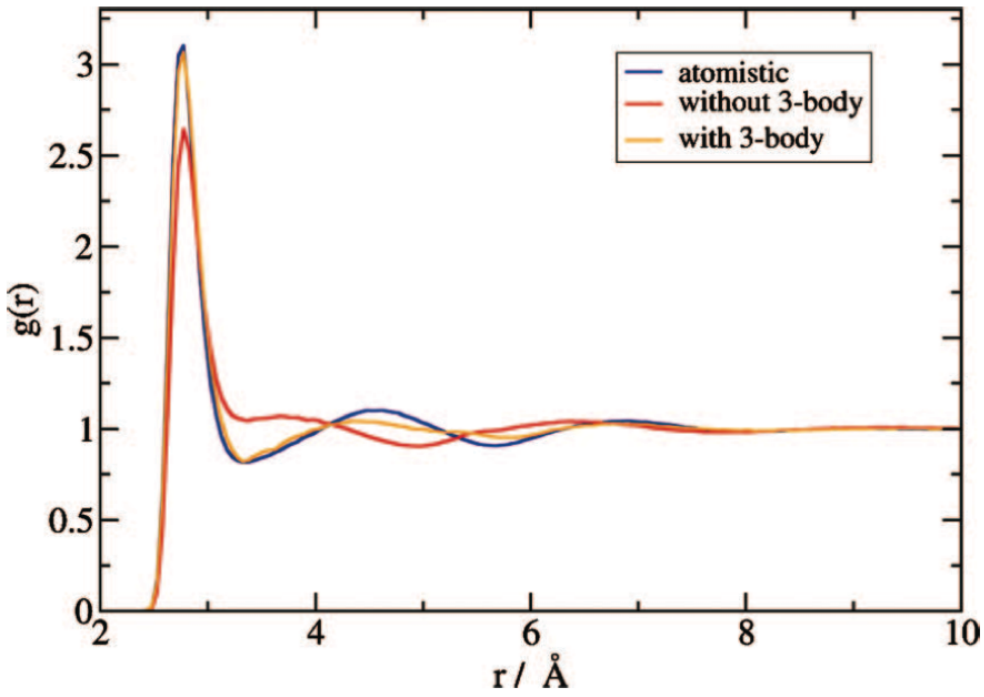

Figure 3.2: Comparison between predictions of the radial distribution function of the centers of mass of SPC/E water for three models, an all-atom model (blue), an MS-CG model including only isotropic pairwise interactions (red), and an MS-CG model including both isotropic pairwise and Stillinger-Weber-like three-body interactions (yellow). Reproduced from Larini et al.[185]

This case is almost too simple to show the advantage of MS-CG as a diagnostic tool, since in one-site water the three body correlations are an obvious distribution to check

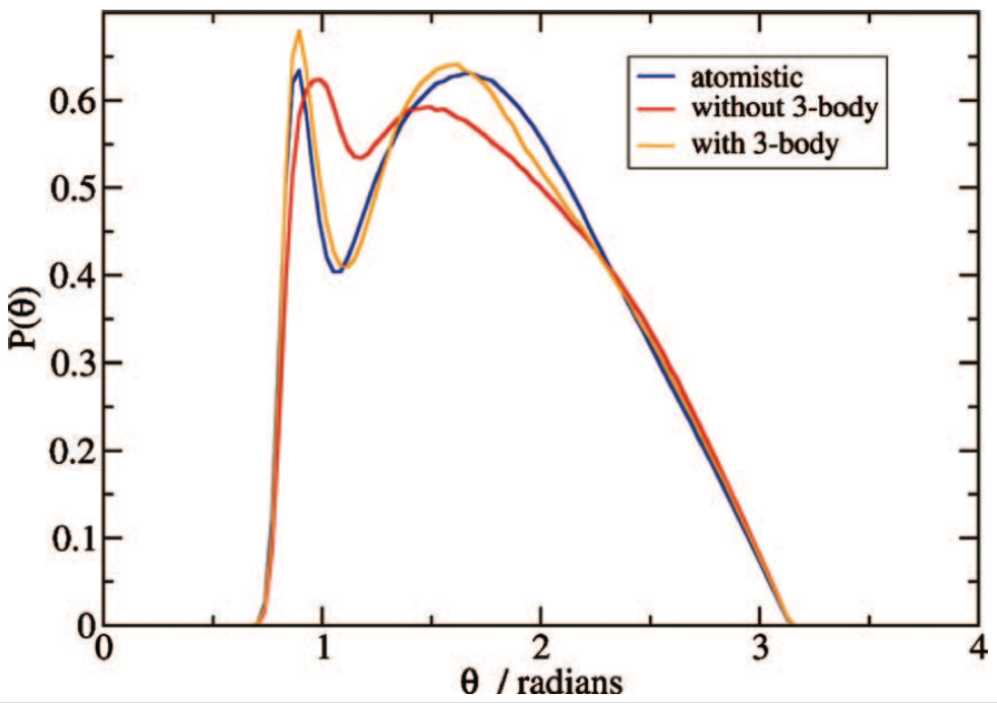

Figure 3.3: Comparison between predictions of the triplet-angle distribution function of the centers of mass of SPC/E water for three models, an all-atom model (blue), an MS-CG model including only isotropic pairwise interactions (red), and an MS-CG model including both isotropic pairwise and Stillinger-Weber-like three-body interactions (yellow). Reproduced from Larini et al.[185]

for the structure-matching methods. However, in cases that are more general, the missing interactions and correlations may not be obvious or even suspected. In these cases, the MS-CG method can automatically indicate if missing interactions are important, whereas a structure matching method may misleadingly appear successful just because the necessary tests for a rigorous validation are not known in advance. The next examples will show how MS-CG can reveal and clarify poorly understood model insufficiencies.

3.4.2 Peptide and protein models

Homogeneous liquids provide important tests of variational force matching coarse-graining methods, but biomolecular systems present more stringent and complex tests. Multiscale modeling of biomolecular systems is a major focus for the developers of MS-CG,[371, 318, 20,

147, 317, 153, 156, 186, 195, 19, 158, 237, 347] TAMD,[213, 4, 212] and, for that matter, all the methods discussed elsewhere in this volume. This section will briefly discuss application of MS-CG to liquid N-methyl acetamide (NMA),[186] providing a bridge with the previous section, then the use of MS-CG to calculate small peptide PMFs.[371, 318, 147, 317] After discussing MS-CG in these cases, the section will cover the use of TAMD to explore the coarse-grained conformational space of a large protein, GroEL,[4] showing how coupling coarse-grained and fine-grained systems can lead to significant performance gains, and finally the use of Maragliano and Vanden-Eijnden's variational force matching formula to derive an implicit protein model for the diffusion of carbon monoxide in myoglobin.[212]

The NMA molecule is touted as a minimal model for peptides because it is one of the simplest molecules containing a peptide bond with distinct E and Z configurations.[186] NMA exhibits two essential properties that present challenges to the straightforward application of MS-CG to peptide and protein molecules. These are, first, the problem of sampling all intramolecular degrees of freedom, in this case the peptide bond dihedral being difficult to sample adequately, and second, the problem of deciding which interactions to include in the basis set to properly represent the coupling between different degrees of freedom. Larini and Emma-Shea[186] investigate these problems by deriving MS-CG models of NMA at four different, successively coarsened resolutions shown in Figure 3.4. They derive their target force fields from simulations intentionally set up to generate incomplete sampling of the NMA peptide bond configurations, one all-cis NMA liquid and one all-trans NMA liquid, emphasizing and exacerbating the problem of acquiring good statistics, and then develop MS-CG force fields from each to show how the dihedral composition of the liquid affects other coarse-grained interactions.

Larini and Emma-Shea use a standard set of trial force fields for their modeling. They assume all intramolecular interactions can be written as a sum of one-parameter pair distance, angular, and dihedral interactions, while all non-bonded interactions are isotropic pairwise interactions. With these trial force fields, they report the following results. First, for the

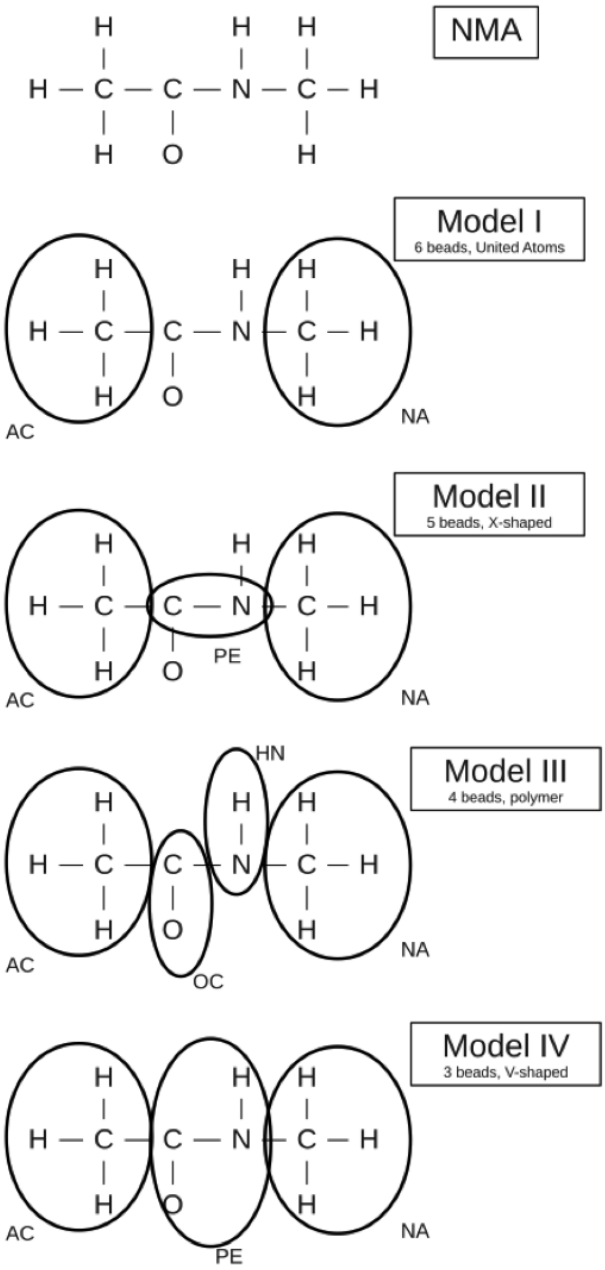

Figure 3.4: The five successively coarsened models considered in Larini and Emma-Shea's study[186] of N-methyl acetamide. Reproduced from Larini and Emma-Shea.[186]

least coarse of the coarse-grained models, MS-CG potentials from either the all-cis or all-trans liquid are transferable across both pure liquids. However, for every model coarser than that, they find that the all-cis and all-trans liquids give rise to different optimal force fields. For the least coarse model exhibiting this dependence, only the intermolecular interactions

between the C-N site and other sites depend significantly on the dihedral composition of the liquid, but the next most coarse sees additional coupling of an angular intramolecular interaction to the dihedral composition, and the most coarse sees yet more coupling. This steady progression demonstrates clearly that for aggressively coarsened MS-CG models of peptides to be effective, new basis sets capturing coupling between the currently standard interactions must be developed. Moreover, it also shows that poorly sampled data can be used to make MS-CG models that are effective in the sampled regions, but nowhere else. The next examples will also confirm this, and build on it by demonstrating that force fields that interpolate between MS-CG force fields derived for different sampling regions can be used for greater transferability between sampling regions.

While NMA is a minimal model for peptides because it contains a peptide bond, small peptides that exhibit secondary structure and folding transitions serve as minimal models for proteins. The first two peptides studied using MS-CG were the polyalanine Ala15 and the beta-hairpin V5PGV5, specifically because they are minimal models for alpha helicity and beta sheet formation, respectively.[371, 318] Several further studies have continued to focus on this type of model,[147, 317] while others have moved on to model more complex behavior, such as aggregation, in addition to structure.[347]

The studies of secondary structure have focused on three questions: what sorts of MS-CG model can reproduce the native state, whether a single MS-CG-based model can reproduce both the native state and non-native states, and whether a single MS-CG-based model can describe peptides with different sequences and secondary structure. The first question is most basic, simply asking what characteristics an MS-CG force field needs to reproduce a single folding basin of the target peptide's potential of mean force. The second and third ask about reproducing multiple basins with a single model, the questions of transferability in configuration space and transferability in chemical design space, respectively. The short answer to these questions seems to be that MS-CG models based on pre-computed Boltzmann inverted bonded interactions and optimized isotropic pairwise potentials can ac-

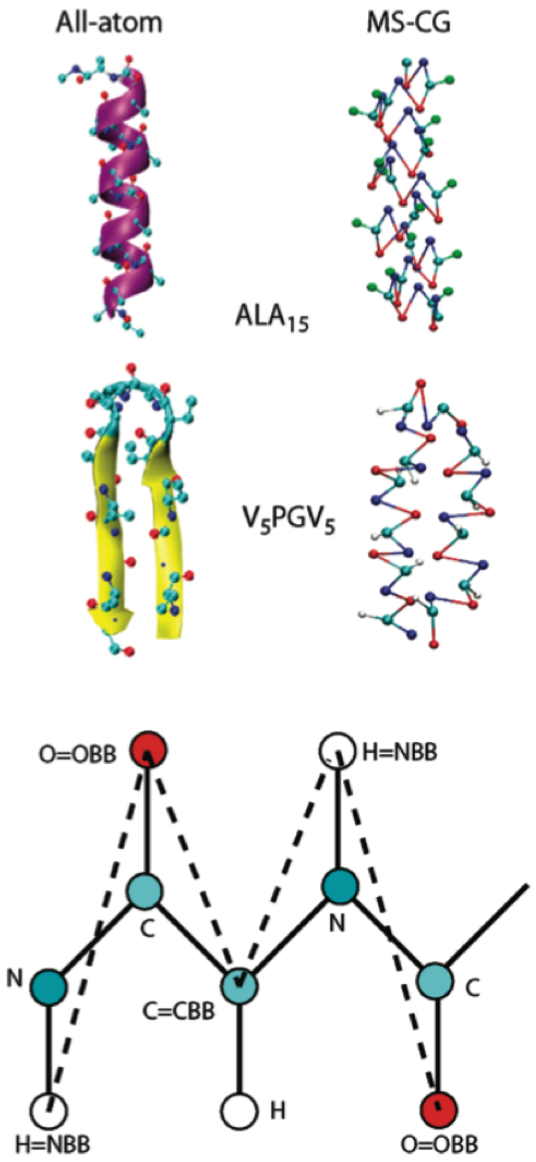

Figure 3.5: The most successful coarse-grained mapping scheme of Zhou et al.,[371] shown for two archetypal peptides (uppermost and middle) and a single glycine monomer (lowest). Reproduced from Thorpe et al.[317]

curately reproduce peptide native states, but that more complex basis sets which include coupling between molecular structural order parameters and standard interactions will be necessary to derive truly transferable MS-CG force fields for peptides.

The first question was answered in a pair of studies, the first of which demonstrated that four site per residue MS-CG models could reproduce native state structures for Ala15

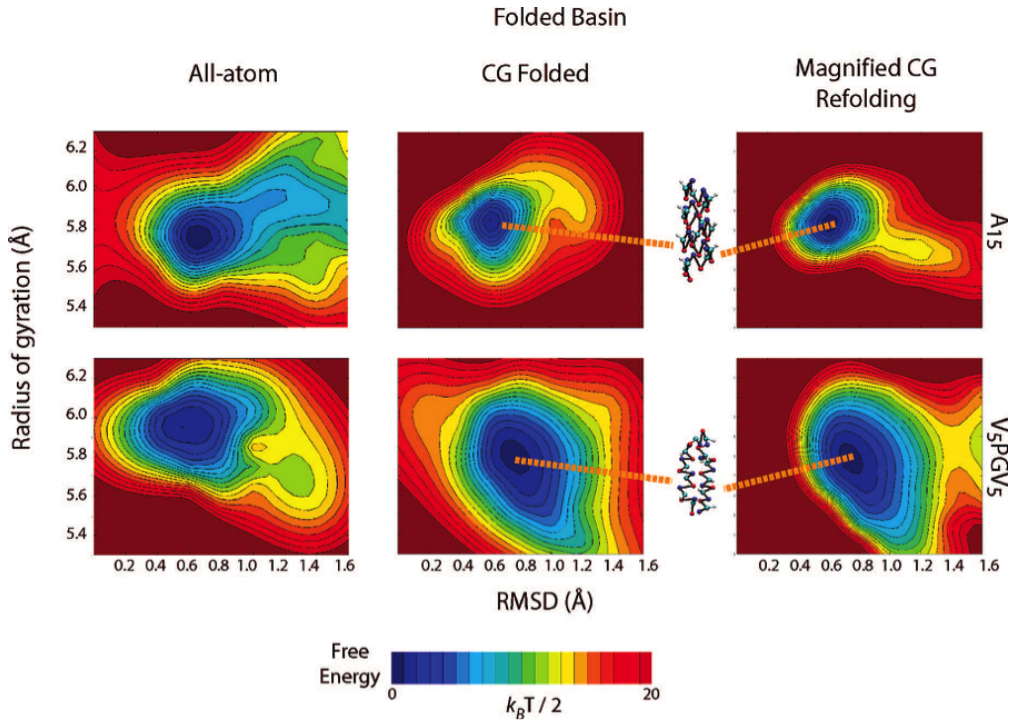

Figure 3.6: Comparison of the predicted folding free energy basins of Ala15 and V5PGV5 for two models of each, an all-atom CHARMM model and an MS-CG model using the mapping and parameterization schemes of Zhou et al.[371] The four leftmost free energy surfaces were calculated for peptides initialized in the folded state, while the two rightmost were calculated for peptides initialized in an unfolded state. The differences between the four rightmost surfaces indicate that these predictions do not correspond to true equilibrium free energy profiles. Reproduced from Thorpe et al.[317]

and V5PGV5 so long as nitrogen-bonded hydrogens on the backbone were apparent in the coarse-grained model,[371] the model illustrated in Figure 3.5, and the second of which demonstrated that these MS-CG models also fold into the native state from unfolded starting configurations.[318] These conclusions were evaluated by comparing the native state basins of each peptide in the coarse-grained and fine-grained simulations, as shown in Figure 3.6. These studies used the following coarse-graining strategy. After defining several mappings and defining particle types by residue and location in the residue, coarse-grained bond distances, angles, and dihedrals for each possible bonded interaction were histogrammed and then these histograms were used to define bonded potentials by Boltzmann inversion. Bonds and angles were assumed to have harmonic potentials, while dihedrals were expanded in a

cosine series. Next, these bonded interactions and an existing water-water potential were used as an invariant part of all trial force fields for an MS-CG optimization. The remainder of the trial force fields consisted of spline-basis isotropic pairwise interactions for each pair of types of coarse-grained particle, including water. The fine-grained models, meanwhile, were simulated using TIP3P wate[167] and the CHARMM27 force field[208] at 310 K. That force field incorrectly predicts a pi helical structure for Ala15, but since the aim of these studies was to show that MS-CG matches all-atom simulation predictions, this is immaterial to their conclusion that the native states were properly reproduced.

After those studies confirmed MS-CG could be used to coarse-grain the native states of these peptides, two further studies by Thorpe et al.[317] and Hills et al.[147] focused on the transferability of MS-CG peptide models. The work of Thorpe et al. used a similar strategy to the previously described studies, while Hills et al. used an implicit solvent scheme with lower backbone resolution (involving only the alpha carbons in the coarse-graining) and higher side chain resolution. Since it is more similar to the work described above and will require less introduction, the work of Thorpe et al. will be our focus. Their model follows the same pattern as the previous two studies, except that the OPLS force field[68] is used in the fine-grained simulations used to develop the MS-CG force fields, which makes the Ala15 model properly alpha helical, and the study extends to a number of other peptides including Leu15, CPEP, GB1, and several tryptophan zippers.

As might be expected, they find that the hydrogen bonding dependent interactions between the nitrogen-bonded hydrogen sites in the coarse-grained model and the carbonyl oxygen sites in the coarse grained models are not transferable between Ala15 and V5PGV5. The isotropic pairwise interactions between these sites are the result of orientational averaging of hydrogen bonds, and the orientations sampled in the two structures are vastly different. However, and perhaps surprisingly, by using a distance-dependent interpolation between the Ala15 and GB1 force fields for this interaction, Thorpe et al. manage to construct an interaction that can be used for any of the peptides studied without destroying the

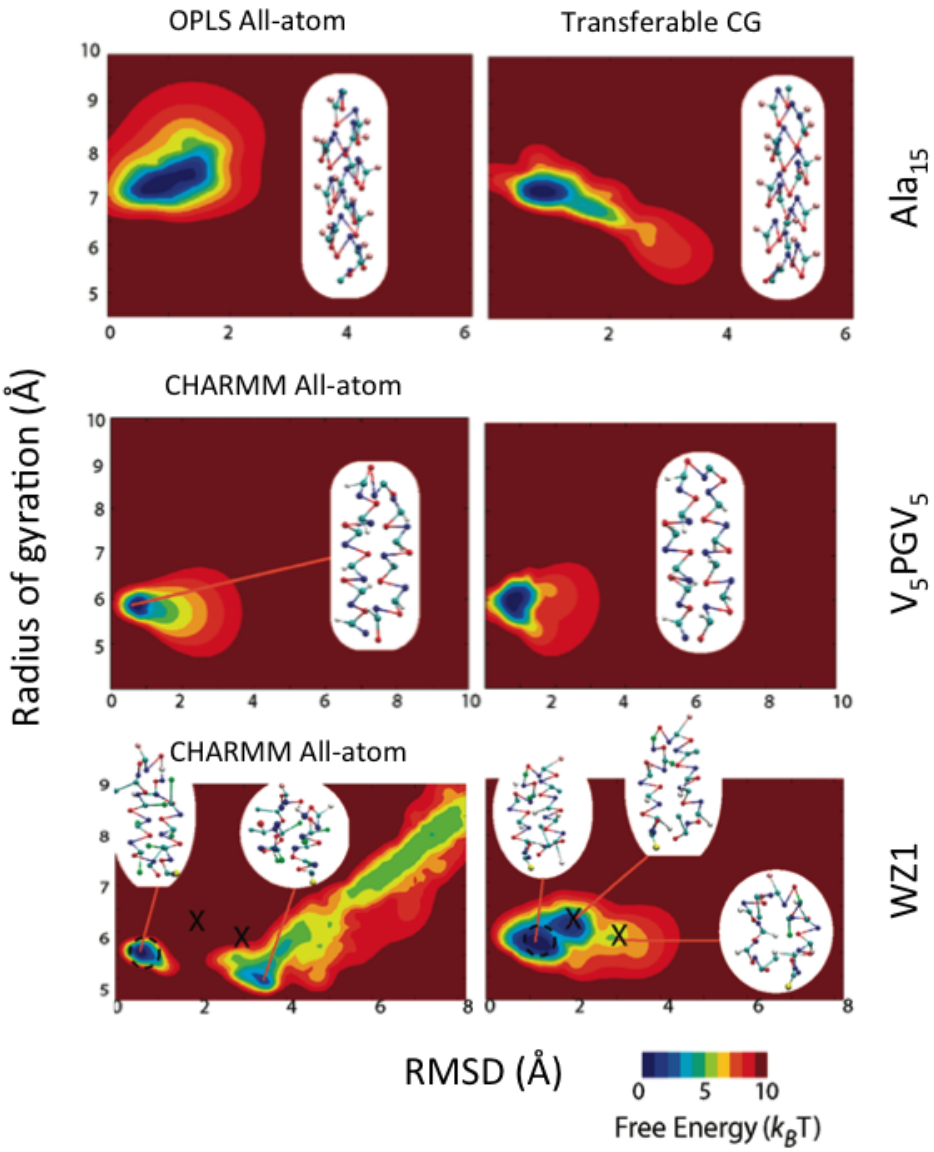

Figure 3.7: Comparison of the predicted free energy surfaces for three pairs of models of peptides. Each pair consists of an all-atom model of a peptide using the force field indicated and a coarse-grained model of the same peptide using the mapping scheme of Zhou et al.[371] and the transferable force field of Thorpe et al.[317] Inset ball-and-stick models are representative coarse-grained structures for each of the predicted basins; X's denote basins predicted by the transferable CG model but not present in the all-atom model. Adapted from Thorpe et al.[317]

stability of the native state, as illustrated in Figure 3.7.

The Ala15 force curve is used at longer distances, and the GB1 curve at shorter distances, and they are smoothly joined by interpolation in the middle. However, though the native

state basin is preserved, additional non-native basins that are not present in the fine-grained model appear. The work of Hill et al., though using a very different coarse-graining strategy, had similar results, reproducing native minima with a force field designed to be transferable but also introducing spurious non-native minima.[147] In all, these results seem to suggest that models for individual basins can be readily developed, but trying to make a force field which is transferable in sequence design space will result in a model that is not transferable in configuration space. Evidently, more subtle force fields and interpolation schemes will be needed to obtain truly transferable coarse-grained biomolecular force fields, but these studies nonetheless provide an interesting perspective on the physics of small peptides and how chemical design affects the conformational properties of complex molecules by taking advantage of force matching coarse-graining as a diagnostic tool.

Having discussed MS-CG studies of peptides, this section will conclude with a description of two applications of TAMD to the coarse-grained simulation of full-sized proteins and protein complexes. The first uses TAMD to explore configuration space for a large protein complex, GroEL, while the second uses it to gather data to calculate a potential of mean force for carbon monoxide diffusion within myoglobin.

E. coli GroEL is a tetradecameric complex consisting of two heptameric rings of 55-kDa protomers that is observed in several crystallographic conformations,[48, 45, 362] illustrated in Figure 3.8, that differ by major collective motions.[300, 222] Abrams and Vanden-Eijnden[4] studied a coarse-grained model of a GroEL subunit consisting of nine coarse-grained particles corresponding to the centers of mass of two 10-Å-radius subdomains for the protein's interdomain hinges, three apical subdomains, three equatorial subdomains, and a single domain for the remaining residues around the interdomain hinges, with residues assigned to each using a radius-of-gyration-based clustering algorithm similar to k-means.[125] They simulated the model using the CHARMM27 force field[208] at 310 K for the fine-grained coordinates in the extended Lagrangian, and using TAMD drag constant of 50 ps-1, spring constant of 100 kcal/mol/$Å^2$, and temperatures of approximately 1x, 2x, 3x, 10x, and

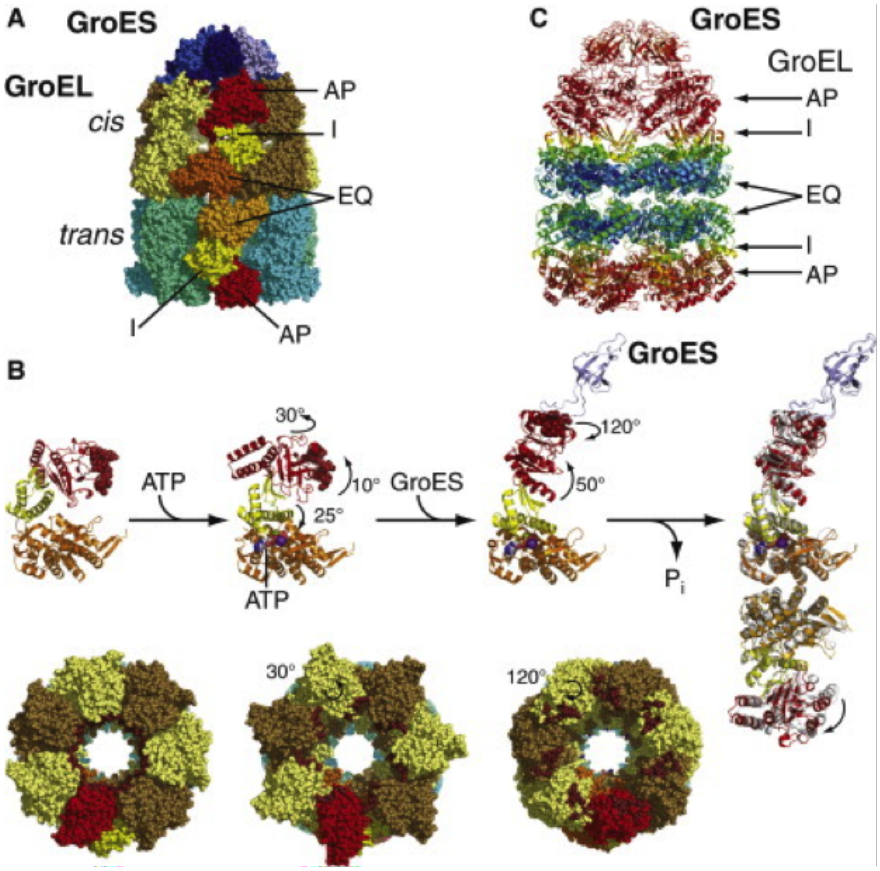

Figure 3.8: Conformational changes in E. coli GroEL. (A) and (C) depict the tetradecameric GroEL complex bound to the heptameric GroES complex, illustrating the quaternary structure of the complex by labeling the apical (AP), interdomain hinge (I), and equatorial (EQ) domains of two particular subunits (A) or all subunits (C). (B) depicts the major conformational changes of GroEL at the level of subunit domain motion (upper line) and motion of the entire complex (lower line). Reproduced from Mayer.[222]

20x the physical temperature.

The 10x and 20x temperature simulations showed deviations from the expected all-atom within-domain structure, but below these temperatures the domains remain folded, suggesting that the method samples the coarse-grained potential of mean force from a realistic atomistic ensemble. The 1x simulation shows no substantial acceleration over conventional MD, but the 2x and 3x explore the conformational space much more efficiently, the latter visiting both crystallographic structures and undergoing major domain-level conformational changes in just 40 ns of simulation. See the work of Skaerven et al.[305] for a comparable investigation using conventional MD.

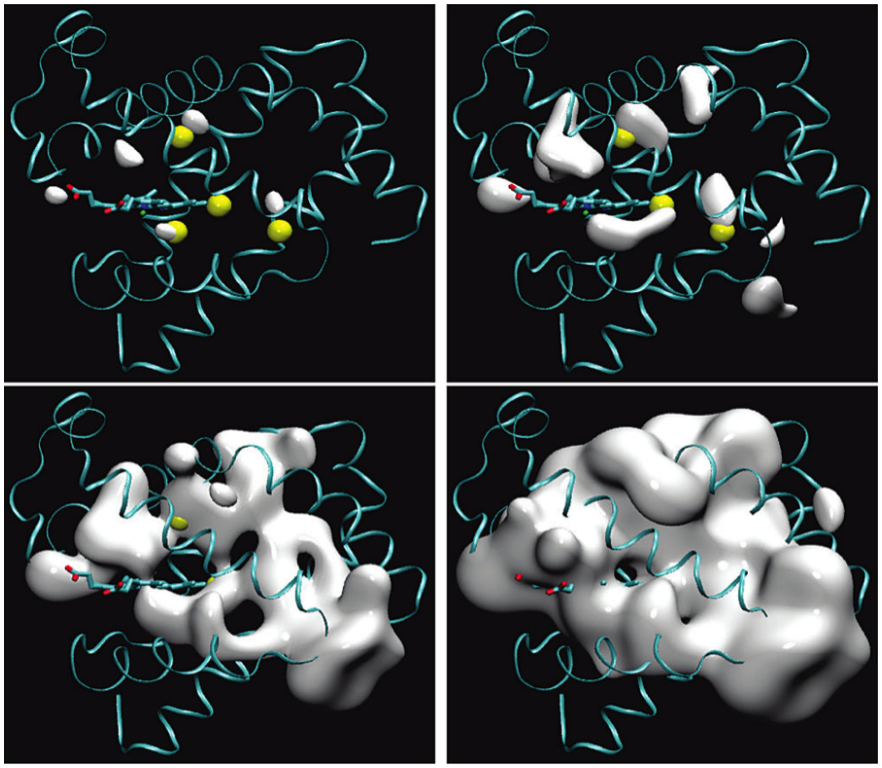

Figure 3.9: Isosurfaces of the potential of mean force for carbon monoxide diffusion in myoglobin in the frame of the protein predicted by Maragliano et al. using a variational force matching approach. The protein backbone is superimposed with the correct orientation to guide the eye, but does not necessarily represent the equilibrium structure given each possible location of carbon monoxide. Reproduced from Maragliano et al.[212]

Myoglobin is a small globular heme-binding protein involved in gas transport and is interesting for the way in which gas must diffuse through the protein to reach its binding site,[54, 127, 128] which is still not fully understood. In an effort to study the pathways available to a gas molecule, in this case carbon monoxide (CO), while diffusing through myoglobin, Maragliano et al.[212] used a combination of TAMD and variational force matching to construct a PMF for CO in the frame of reference of the protein,[30] isosurfaces of which are shown in Figure 3.9. This aggressive coarse-graining to a single particle can be thought of as an implicit solvent model in which the protein plays the role of a solvent. The system is simulated out of equilibrium, with CO molecules initialized from equilibrium conformations of a heme-associated state and then allowed to escape into solvent, but never returning. This would invalidate many of the assumptions in a standard PMF calculation, but here the

actual mechanism of sampling is irrelevant so long as the collected configurations come from some probability distribution and the actual distribution is irrelevant.

The researchers performed ten of these simulations lasting about 200 ps each using TAMD drag constant 250 ps^{-1} and spring constant 200 kcal/mol/$Å^2$ to sample the motion of CO through the protein at 300 K. Following these quick exploratory simulations, each of 239 sampled coarse-grained configurations was simulated using the same strength of harmonic restraint, 200 kcal/mol/$Å^2$, for 500 ps each, to obtain mean force values. These forces were then used as the basis for variational force matching using the Maragliano and Vanden-Eijnden residual shown earlier with trial force fields consisting of a linear combination of 239 Gaussian radial basis functions with uniform, isotropic variance centered at the sampled forces.

The resulting potential of mean force clearly shows minima where CO has been experimentally observed to rest within the protein,[253, 307, 293] and a number of possible pathways into and out of the protein. The PMF does not show if motion of CO is correlated with any internal motions of the protein, giving only a three-dimensional PMF, and it is hard to imagine this method scaling to produce accurate many-body force fields unless more physically motivated basis sets are used for the trial force fields. However, this PMF, derived from an exceptionally poorly defined ensemble and yet still showing the essential physics of gas diffusion in the complex, inhomogeneous protein environment through multiple previously unknown pathways, clearly shows the power of local PMF calculation methods.

3.5 Conclusion

Force matching is an important alternative to structure matching methods, with clear advantages in certain areas and disadvantages in others. Force matching methods are local and diagnostic, which makes them extremely useful for exploratory research, whether that exploration is of phase space, as with TAMD, or model space, as with MS-CG. At the same time, they are often less able to reproduce known behavior, such as the radial distribution

function for liquid water, without going through extra effort to build a more complex model. For some applications, such as parameterizing the intramolecular interactions of peptides, this has proven to be a high barrier to the adoption of force matching. However, where force matching is impractical it is often because it is acting as a diagnostic, and so the challenges to further application of force matching are equally challenges to our true physical insight into complex molecular systems.

This chapter has demonstrated the rigorous statistical mechanical basis for force matching in the canonical ensemble and its use to formulate two variational force matching methods currently being explored in the literature. It has presented how transferability has been treated in this framework, and a sketch of how to extend the force matching equations to other ensembles. These equations are old, but the recent rise of MS-CG after years of Blue Moon sampling[80] and *ab initio* force matching[112] and TAMD[214] after years of AFED[281] and Car-Parrinello molecular dynamics[62] show well how much unexplored potential this old theory still contains.

Numerical strategies based on physical insight have been key to the recent progress in force matching, but significant challenges remain. This chapter focused particularly on the current state of the art in MS-CG, highlighting the related challenges of developing models from poor sampling of a fine-grained model and designing appropriate basis sets for complex coarse-grained models. It is rare that an MS-CG model is developed without encountering one of these problems, so it is important to keep track of the common ways MS-CG is modified using other methods and extra assumptions and to look for the physical meaning behind them– to the extent that these can hide the flaws of a model, they obscure the diagnostic power of the method. However, there are clear strategies available for attacking these problems, including using advanced thermostatting methods[65, 66] or hybrid molecular dynamics/Monte Carlo schemes such as replica exchange[314] to sample the canonical ensemble more efficiently, using inhomogeneous regularization schemes to address overfitting, and trying orthogonal special functions, Fourier series, wavelets, and multiplicative combinations of

more standard basis functions to improve basis sets.

The problem of transferability, for any definition and type of transferability, will be linked to the problem of defining a good basis set. As shown in the examples of the NPT ensemble and temperature transferability, transferability can be treated as dependence of the force field on extra parameters. Unfortunately, these parameters may be coupled perniciously to nearly every other degree of freedom in the system, as would have been the case for volume transferability in the NPT ensemble if it had been examined without performing a change of coordinates to remove the dependence of particle positions on volume. In many cases, adding parameters to a force field that couple to all interactions will drastically increase the computational expense of the model for each configuration and state point. However, in the case that the new parameter is fixed during simulation then the interactions may still be efficient, since in a reasonably designed simulation program the extra work will be done at compile-time. Two prime examples would be designing sequence-dependent force fields for amino acids in proteins or force fields for a liquid model over a range of temperatures; for simple cases, the latter is already being explored in the literature.

Another promising approach to fixing the problems of coarse graining may be to add features of structure matching and other coarse-graining methods. The HAS approach has seen success in rectifying MS-CG force fields,[20, 19] as have hybrid structure and force matching methods[286, 98] and iterative force matching methods based on the YBG form of the MS-CG equations.[74] Similarly, TAMD could be combined with explicit coarse-grained force fields and umbrella sampling to cut down on the cost of exploring well-characterized and uninteresting regions of phase space. It may even be that a strict distinction between force matching and structure matching methods will not survive long if reliable methods combining the strengths of both can emerge from the rapidly growing overlap between the two.

CHAPTER 4

THE SENSITIVITY OF COARSE-GRAINED FORCE FIELDS

This chapter reproduces material from Wagner, J. W., Dama, J. F., & Voth, G. A. (2015). Predicting the Sensitivity of Multiscale Coarse-Grained Models to their Underlying Fine-Grained Model Parameters. Journal of Chemical Theory and Computation, 11(8), 35473560. Jacob Wagner and I shared first authorship; my contribution included the theory, project design, and most interpretation of the results.

This section describes how to calculate how the average forces between groups of fine-grained particles change upon changes in the fine-grained model parameters, without actually running the experiment of actually making those changes.

Returning to our neighborhoods of families, we would be attempting to infer how the family preferences change upon changes in individual family member preferences. The key complication here is that when the member preferences change, the responsibilities in the family may change; for instance the kids might trade chores in ways that bring them in contact with the neighbors differently, so even if their individual feelings for the neighbors don't change, the effects of those feelings change nonetheless.

Similarly, in the molecular models, changes in parameters of a model can have significant indirect effects on effective coarse-grained interactions because the parameters affect the distribution of substructure of the coarse-grained particles; the direct effect is simple to determine but the indirect effect requires more subtle theory.

We might consider income in our neighborhood model for special attention: as family income rises, how do neighborly interactions change? Does it matter if it's the little kids, the teenagers, a father, or a mother who brings in the new income? And how much change is required before gentrification sets in, making the neighborhood change so much that a whole new model is required?

4.1 Motivation

Coarse-grained (CG) models seek to capture the essential details of fine-grained (FG) models at reduced computational cost by eliminating degrees of freedom (DOF).[233, 337, 244, 51] To further increase computational savings with CG models, it is desirable to know when one can reuse the same CG model to describe FG models similar to the original FG model that was used to parameterize the CG model. Unfortunately, as discussed in the previous chapter, CG models tend to have limited transferability because eliminating DOFs leads to state point dependent effective interactions.[197, 165] The dual problem to designing transferability is the determination of model sensitivity, i.e., how a CG model would change if it were parameterized under a different set of FG interactions or at a different state point. CG sensitivities can be used simply to determine the transferability of CG models, but a more promising and enterprising approach is to correct CG models with the changes predicted from the calculated sensitivities. However, for this to be practical one must have a method of calculating sensitivities that is more computationally cost effective than running new FG simulations to directly calculate new CG interactions at each new state point. In this chapter, we address this need for a computationally efficient method to calculate model sensitivities by proposing novel formulae for computationally efficient, low noise estimates of sensitivities from single FG simulations.

The model sensitivity of molecular systems has been extensively investigated at a single, all-atom (AA) level of resolution. Wong and Rabitz[358, 359] calculated changes in free energy as a function of changes in the input Lennard-Jones (LJ) parameters in the simplest, linear response sense of sensitivity using finite differences (FD). Using *ad hoc* modifications to FDs, Rocklin et al.[280] calculated the sensitivities of binding free energies to changes in interaction parameters. More computationally efficient methods employing single state and multistate statistical reweighting[299, 298, 255] can also be used to get estimates of the sensitivity using FDs and have been used to find potential sensitivity to interaction cutoffs.[255] Fleishman and Brooks took a different approach,[124] using derivatives of the

partition function and thermodynamic integration to calculate sensitivity of entropy and enthalpy via perturbation theory. Later, Wong, Thacher, and Rabitz used careful statistical mechanical derivations to determine first and second order sensitivity coefficients.[360] These sensitivity coefficients have been used in materials design to optimize binding free energies by changing cation interactions[81] and to determine which input parameters are most influential in determining observables.[372] Recently, an expanded set of sensitivity equations were applied to improve a classical water model's agreement with *ab initio* and experimental measures.[345] However, sensitivities calculated at a single level of resolution do not address the problem of CG model transferability.

Sensitivity between models of two different resolution levels has been the subject of limited study. Krishna et al.[177] used the multiscale coarse-graining (MS-CG) methodology[154, 153, 246, 247] and statistical reweighting over temperatures to get different CG potentials, which is a complementary approach to the one taken in this paper. Lu et al.[204] used FDs with MS-CG to decompose free energies into entropic and enthalpic components using sensitivity to temperature. As an alternative, some researchers have tried to increase the transferability of CG potentials by including extra independent variables. A few of these approaches included three-body interactions,[185, 96] temperature dependent terms,[86, 273] density dependent terms,[9, 10, 151] or concentration dependent terms.[291] However, while much work has been done to improve the transferability of CG potentials, only reweighting and FD approaches have been used to probe sensitivity in multiscale calculations.

Several systematic approaches for developing CG models could potentially be used as a starting point for developing methods to calculate sensitivities, including relative entropy (RE),[296] inverse Monte Carlo (IMC),[239] iterative Boltzmann inversion (IBI),[207] force matching (FM),[154, 153, 246, 247, 202, 245] and the generalized Yvon-Born-Green (g-YBG) equations.[235, 236] RE minimization is a general approach in which one aims to minimize the loss of Shannon information from the FG model to in the CG model potential.[296] If RE is minimized using Newton's method,[239] one obtains IMC,[276] which inverts radial

distribution functions (RDF) iteratively to provide interaction potentials – though sampling noise must be taken into account.[44] Similarly, IBI inverts the RDF iteratively and is an approximate RE minimization just as IMC is, using a simpler fixed-point optimization than Newton's method.[69] Likewise, FM and g-YBG, implemented as MS-CG, converge to the same result as RE in the limit of a complete basis set since FM minimizes the average of the gradient squared of the relative entropy.[285] Since RE converges to the same results as IBI, IMC, FM, and g-YBG in the appropriate limits,[69] it is important to consider which method is most appropriate given the limitations of the problem at hand. In the case that sampling is incomplete at short interaction distances and there are potential issues with basis sets to describe both the CG potential force field and the sensitivity of the CG potential or force field, the local nature of MS-CG becomes appealing. Local nature here refers to the fact that in MS-CG, the fit in each portion of the force field is linked to fits in other portions linearly, through a g-YBG equation,[235, 236] rather than through a complex, nonlinear, and nonlocal dependence. This feature of MS-CG removes the nonlinearity found in the other global, distribution matching minimization methods, so it leads to a more direct, computationally straightforward method of calculating sensitivity better suited to rapid prototyping.

The present work develops reweighting-free, single simulation formulae that calculate the sensitivity of CG potentials and force fields to changes in the underlying FG interaction parameters and state points at the level of linear response. The calculated sensitivities are used to develop corrections to CG models that increase model accuracy when the CG potentials and force fields are transferred alchemically across interaction parameters or thermodynamically across state points. The accuracy of these predicted sensitivities are evaluated by comparison with reweighted FDs, and the accuracy of the corrected, transferred potentials are compared against potentials without any sensitivity correction. The remainder of the article is structured as follows: A Theory and Methods section describes the derivation and significance of the formulae developed in this work as well as the numerical and simulation methods used; the next shows the application of these formulae to single site methanol and

solvent free sodium chloride systems and the resulting accuracy of the predicted sensitivities and potentials, then provides a general discussion of those results including suggestions for future work; the final section provides conclusions.

4.2 Theory and Methods

4.2.1 Sensitivity Theory

The fundamental measure of sensitivity to small changes is the derivative

$$\frac{dU(\mathbf{R}^N;\lambda)}{d\lambda} = \lim_{\delta\lambda\to 0} \frac{U(\mathbf{R}^N;\lambda+\delta\lambda) - U(\mathbf{R}^N;\lambda-\delta\lambda)}{2\delta\lambda} \tag{4.1}$$

Here, the sensitivity of the CG potential $U(\mathbf{R}^N;\lambda)$ to a FG parameter λ is calculated by finding the difference between CG potentials obtained using modified FG parameters $\lambda \pm \delta\lambda$. In the above equation, $\mathbf{R}^N$ are the CG configurational variables. This method requires the calculation of at least two CG potentials to calculate the sensitivity. However, the range of $\delta\lambda$ in which this limit is approached, known as the linear regime, is not known *a priori*. This means that in fact more than two CG potentials must be calculated to verify the meaningfulness of a single calculated sensitivity. A second problem with this FD approach is that the random noise and fluctuations in estimates of the CG potentials are magnified dramatically when $\delta\lambda$ is small, exactly where the limit is approached. For the FD approach to be feasible, therefore, one must find a that is both in the linear regime and sufficiently large to make pulling the sensitivity signal out of sampling noise tractable, but there is no guarantee that such a $\delta\lambda$ exists in every case.

An alternative to the basic multi-trajectory FD (MTFD) is to use statistical reweighting to obtain the CG potentials at different $\delta\lambda$ values from a single FG simulation. Statistical reweighting reuses configurations generated using a given parameterization by applying a reweighting factor, the ratio between Boltzmann factors across parameterizations, to the

results of reanalyzing the configurations using a difference parameterization. For generating CG models using MS-CG or g-YBG, this amounts to weighting the FM residual from the reevaluated forces by the exponential of the difference in the CG potential calculation (or inverse temperature, if temperature is varied). The use of a single trajectory in reweighting should minimize the noise seen at small $\delta\lambda$, which makes this approach appear relatively promising compared to FD, but the range of linear regime is still not known *a priori*. Another problematic condition required for reweighting to be practical is that the original ensemble and the ensemble estimated by reweighting must have significant overlap so that the reweighted trajectory can give reliable averages. Furthermore, the averages are susceptible to bias when the original sample may not provide configurations that overlap with the reweighted ensemble *evenly*. An example of this is the application of reweighting to calculate averages at higher temperatures than the temperature of an initial simulation. Since the original trajectory explores only a subvolume of the phase space explored at higher temperatures, the reweighting procedure typically biases the resulting potentials to over-represent behavior characteristic of lower-energy conformations. For reweighting across interaction parameters, it is not always clear when this is a problem or how significant the bias may be—even after the calculation is complete.

Ideally, one would like a method of calculating sensitivities in the linear regime that does not depend on knowledge of the size of the linear regime, requires minimal computation, and is less susceptible to bias than reweighted finite differences (RFD). We can do so by analytically evaluating the limit in the FD above, then using the resulting formulas to make our calculations. Starting from the FD formula above, one arrives at the equation

$$\frac{dU(\mathbf{R}^N;\lambda)}{d\lambda} = \left\langle \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} \right\rangle_{\mathbf{R}^N;\lambda}, \tag{4.2}$$

where $u(\mathbf{r}^n;\lambda)$ is the FG potential given parameter λ in terms of the FG coordinates $\mathbf{r}^n$. However, this equation is remarkably data-inefficient: it only uses one scalar value of infor-

mation per sampled frame, the derivative of the potential with respect to λ for that frame. Fitting a many-body function with many free parameters requires a great deal of input training data, and using only one datum per frame of input data would require a huge number of frames to properly parameterize the many-body sensitivity. Therefore, we apply a trick with a long history. By analogy to FM,[154, 246, 247, 202, 245, 112] which uses more data per frame than potential matching by matching derivatives of the potential with respect to particle positions instead of matching per-frame potentials directly, one can also derive formulas for sensitivity matching that match the sensitivity of the derivatives of the potential with respect to all particle positions instead of matching the per-frame sensitivity directly. The gain in information per frame is proportional to the number of particles, which can be quite large. The remainder of this section describes the derivation of two such formulae, each of which results from a different approach to the problem of representing the many-body sensitivity in a reduced space of trial functions. *We propose that the first formula be used for practical calculation of sensitivity and the second formula be used as a theoretical diagnostic tool.*

In the first derivation, we find a formula by considering the sensitivity of approximations to the many-body potential given a fixed basis. After all, any practical CG potential will be an approximation, and we are therefore interested in the sensitivity of approximations when we talk about the sensitivity of CG models. A natural choice here is to look at the sensitivity of an approximate CG potential in the same set of trial functions used to construct the CG potential; because the basis functions for the CG potential and sensitivity are the same in this case, we call this a self-consistent basis (SCB) single-point formula. In order to construct this formula, one needs to start from the FM residual expression reweighted from λ to $\lambda+\delta\lambda$ with the framewise weight function

$$w_t(\mathbf{r}^n;\lambda,\delta\lambda) = \frac{\exp(-\beta u(\mathbf{r}^n;\lambda+\delta\lambda)+\beta u(\mathbf{r}^n;\lambda))}{\frac{1}{N_t}\sum_{t=1}^{N_t}\exp(-\beta u(\mathbf{r}^n;\lambda+\delta\lambda)+\beta u(\mathbf{r}^n;\lambda))}, \tag{4.3}$$

where N_t is the total number of simulation frames. Optimization of the residual with respect to basis functions in order to get the reweighted FM normal equations yields[247, 201]

$$\frac{1}{N_t}\sum_{t=1}^{N_t} w_t(\mathbf{r}^n;\lambda,\delta\lambda)\mathbf{F}^T\mathbf{F}\phi = \frac{1}{N_t}\sum_{t=1}^{N_t} w_t(\mathbf{r}^n;\lambda,\delta\lambda)\mathbf{F}^T\mathbf{f} \tag{4.4}$$

where $\mathbf{F}$ is a matrix of configurational information about the basis function values for each particle, $\mathbf{f}$ is the $3N$ vector of the target forces, and ϕ is a vector of the unknown linear basis function coefficients. Then, taking a derivative of both sides with respect to $\delta\lambda$ and taking the limit $\delta\lambda \to 0$, a set of normal equations for the sensitivity emerges. Taking the limit of a long trajectory and a complete basis set and then rearranging those normal equations, the expression for the approximate sensitivity matching in terms of thermodynamics averages is

$$\begin{aligned}\frac{d\nabla U(\mathbf{R}^N;\lambda)}{d\lambda} = \Bigg\langle M^+\left(\frac{d\nabla u(\mathbf{r}^n;\lambda)}{d\lambda}\right) - \frac{\beta}{N_{CG}}\left(\frac{du(\mathbf{r}^n;\lambda)}{d\lambda} - \left\langle\frac{du(\mathbf{r}^n;\lambda)}{d\lambda}\right\rangle_\lambda\right)\cdots \\ \cdots\times\Big(M^+(\nabla u(\mathbf{r}^n;\lambda)) - \nabla U(\mathbf{R}^N;\lambda)\Big)\Bigg\rangle_{\mathbf{R}^N;\lambda},\end{aligned} \tag{4.5}$$

where $\langle du(\mathbf{r}^n;\lambda)/d\lambda\rangle_\lambda$ is the Boltzmann weighted expectation value of $du(\mathbf{r}^n;\lambda)/d\lambda$ over the entire FG ensemble, and M^+ is the mapping operator that transforms the FG forces into CG forces (See Appendix B for derivation and details). Every term on the right can be estimated directly from simulation, so these estimates can be calculated in a single step without iteration.

An interesting alternative to the derivation in the previous section is to consider what the expression would be like if one treated the sensitivity of the many-body potential directly regardless of basis set limitations. A sensitivity estimator based on approximating the sensitivity of a full basis set potential using a finite basis set rather than on the sensitivity of approximate potentials using finite basis sets would provide a diagnostic for assessing the importance of renormalized many-body effects in CG sensitivities. As before, we look for a formula in terms of derivatives of forces instead of derivatives of potentials. Starting from the

FD of the forces with the averages in the ensembles written explicitly, then substituting in the mapped FG forces for the CG forces, the transformation of these ensembles to a common ensemble leads to

$$\frac{d\nabla U(\mathbf{R}^N;\lambda)}{d\lambda} = \lim_{\delta\lambda\to 0}\frac{1}{2\delta\lambda}\left\langle M^+\left(\nabla u(\mathbf{r}^n;\lambda+\delta\lambda)\right)\frac{e^{-\beta u(\mathbf{r}^n;\lambda)+\beta u(\mathbf{r}^n;\lambda+\delta\lambda)}}{\left\langle e^{-\beta u(\mathbf{r}^n;\lambda)+\beta u(\mathbf{r}^n;\lambda+\delta\lambda)}\right\rangle}\cdots\right.$$
$$\left.\cdots - M^+\left(\nabla u(\mathbf{r}^n;\lambda-\delta\lambda)\right)\frac{e^{-\beta u(\mathbf{r}^n;\lambda)+\beta u(\mathbf{r}^n;\lambda-\delta\lambda)}}{\left\langle e^{-\beta u(\mathbf{r}^n;\lambda)+\beta u(\mathbf{r}^n;\lambda-\delta\lambda)}\right\rangle}\right\rangle_{\mathbf{R}^N;\lambda} \tag{4.6}$$

After expanding the exponentials in terms of $\delta\lambda$ and discarding all terms higher than linear in $\delta\lambda$ (see Appendix A for details), one gets

$$\frac{d\nabla U(\mathbf{R}^N;\lambda)}{d\lambda} = \left\langle M^+\left(\frac{d\nabla u(\mathbf{r}^n;\lambda)}{d\lambda}\right)\cdots\right.$$
$$\left.\cdots -\beta\left(\frac{du(\mathbf{r}^n;\lambda)}{d\lambda} - \left\langle\frac{du(\mathbf{r}^n;\lambda)}{d\lambda}\right\rangle_\lambda\right)M^+(\nabla u(\mathbf{r}^n;\lambda))\right\rangle_{\mathbf{R}^N;\lambda}, \tag{4.7}$$

where N_{CG} is the number of CG sites in order to make the sensitivity intensive. This equation is a self-consistent iterative (SCI) single point formula because while the left hand side seems optimistically like it could be computed in a variational approximation by performing FM on the right hand side, this is actually not correct. The term $\langle du(\mathbf{r}^n;\lambda)/d\lambda\rangle_{\mathbf{R}^N;\lambda}$ is exactly the many-body function that the formula is meant to calculate, and therefore the equation must be solved iteratively: after each variational calculation step to find the left-hand side, $\langle du(\mathbf{r}^n;\lambda)/d\lambda\rangle_{\mathbf{R}^N;\lambda}$ must be reevaluated framewise using the integrated form of the new left hand side to generate new target derivatives, and a new variational calculation must be run. The process repeats until self-consistency. Note that FM calculates a potential up to an additive constant. Normally, this constant has no physical effect, but in this case the constant is important in the nonlinear term containing $\left(\frac{du(\mathbf{r}^n;\lambda)}{d\lambda} - \left\langle\frac{du(\mathbf{r}^n;\lambda)}{d\lambda}\right\rangle_\lambda\right)$. We therefore apply a configurationally independent constant correction to that difference so that its average over all frames is zero. This amounts to a single step of direct scalar matching

used to seed the iterative FM calculations; the scalar does not affect the distributions of configurations in sensitivity-corrected models.

Both the SCB and SCI single point formulae have the same first term on the right hand side, which can be considered the naive sensitivity since it neglects any non-pairwise effects on the CG potential and force field. Interestingly, this is what one would obtain if one reanalyzed a trajectory using a different parameter set as in the RFDs but neglected to apply the reweighting factor. This is in effect what was reported in Ref.[280] One can see that if the second set of terms on the right hand side of both single point formulae were to be zero, this naive sensitivity would in fact be the correct sensitivity. Thus, differences between the naive sensitivity and the single point formulae reflect the importance of the correlation correction to the naive sensitivity. Differences between the SCI and SCB equations reflect the importance of basis set effects in determining which correlations should be used to correct the naive sensitivity for approximate models.

Even though these formulae both capture correlated many-body effects of the sensitivity, they do so in different ways. In the SCI single point formula, the correction to the naive sensitivity is like a transport term since it is the product of the mapped forces and the deviations in $du/d\lambda$; it measures the amount the FG distributions corresponding to each CG distribution are pushed around "underneath" the CG configuration. In the SCB single point formula, this correction is the product of the deviation in $du/d\lambda$ and the deviation in the forces, making it a covariance term that closely echoes the covariance corrections to the naive sensitivity found in the literature for single resolution sensitivity.[360] A practical difference between the two single point formulae is that the SCB averages $\langle du(\mathbf{r}^n;\lambda)/d\lambda\rangle_{\mathbf{R}^N;\lambda}$ over the entire FG ensemble requiring no iteration, while the SCI averages $\langle du(\mathbf{r}^n;\lambda)/d\lambda\rangle_{\mathbf{R}^N;\lambda}$ are conditional on the CG ensemble that needs to be reevaluated based on the most recent estimate from the previous iteration. These differences in averaging are consistent with the differing applications of basis sets made in the derivation of each model. Both approaches become equivalent in the limit of a complete basis set, but for a finite basis set the SCB

formula describes a practically useful sensitivity and the SCI formula is better used as a diagnostic for understanding the physics of renormalized many-body effects.

4.2.2 Simulation and Fitting Details and Conditions

Molecular dynamics (MD) simulations were performed on AA methanol and 1M sodium chloride systems in LAMMPS.[265, 53, 52] All systems were run with a 1 fs timestep and used nonbonded Lennard-Jones (LJ) interactions with a radial cutoff of 1.0 nm as well as particle-particle particle-mesh (PPPM) electrostatic interactions. Both systems were equilibrated by simulating them for 5 ns at constant NPT at 1 atm and 300K, setting their volume to the average of the last 2 ns of NPT simulation, and then simulating them for at least 1 ns at constant NVT at 300K. Subsequent sampling for modified parameters or temperatures were started from this equilibrated configuration, but allowed to evolve for an additional 1 ns before sampling. The OPLS[168] methanol system of 1,000 molecules were sampled for 2 ns with configurations recorded every 250 fs, consistent with other studies.23 For the sodium chloride system, 20 sodium and 20 chloride ions were simulated using Joung and Cheatham's[169] parameterization solvated in 3330 SPC/E[167] water molecules for 20 ns with configurations recorded every 200 fs, consistent with other studies.[61] The methanol system was coarse-grained to one site using a center of mass mapping, as in previous work.[247] The sodium chloride system was coarse-grained by eliminating all water molecules to create a solvent free model. All CG forces, potentials, and sensitivities were calculated using the MS-CG FM code with a nonbonded cutoff of 1.0 nm, and sixth order spline basis functions with a resolution of 0.07 nm. CG simulations were started from the mapped version of the final configuration for the sampling run. A total of 1,000 CG timesteps were allowed for equilibration and randomization. Configurations were sampled every 1,000 CG timesteps for both systems with sampling runs of 2×10^6 CG timesteps for CG methanol and 2×10^7 CG timesteps for CG solvent free sodium chloride.

Independent samples and reweighted potentials were calculated for changes to all LJ

epsilon, LJ sigma, and partial charge parameters. In units of kcal/mol for LJ epsilon, Åfor LJ sigma, and e, the fundamental charge, for charge, CG potentials were calculated with positive or negative changes in one parameter of 0.001, 0.002, 0.005, 0.010, and 0.020. Changes to the charge of one atom type were offset by changes to the charge on an adjacent atom type in order to keep each molecule charge neutral. For methanol, this meant moving charges on the carbon and the neighboring methyl hydrogens or the oxygen and the neighboring hydroxyl hydrogen for methanol. For sodium chloride, this meant moving charges on the ions or within the water molecule. For the graphs comparing the single point formulae to the MTFDs and RFDs, the confidence ranges for MTFD and RFD curves were determined by integrating the 95% confidence interval calculated for all pairs of FDs within 0.005 parameter units. Confidence ranges for the single point formulae and the RFDs in these curves were likewise calculated by integrating the 95% confidence interval from 5 replica simulations. The confidence range for all of the independent trajectories corresponds to the integrated 95% confidence interval of 6 replicas using the original parameterization.

To compare the effectiveness of these sensitivity formulae for transferring potentials, predicted potentials, i.e., original potentials plus the sensitivity with respect to a parameter times the change in the parameter, are compared to the CG potentials obtained from both 1) independent trajectories using an actually modified parameter set and 2) Boltzmann reweighting the original trajectory to the modified parameter set. The difference in these modified CG potentials via sensitivity, via reweighting, and via independent trajectories from the CG potential with the original parameters is quantified by integrating the absolute difference multiplied by the RDF and divided by the range of integration. This gives a single number summary (in energy units) of how different the variously transferred potentials are from the original potential for a given change in parameters. In this section, the confidence ranges for each point were calculated by propagating the uncertainty from the potentials through each of the operations in equation 9. The uncertainty of each point of the potentials was calculated as the root mean square (RMS) fluctuations of six independent trajectories.

This uncertainty in the potential was used to calculate the uncertainty in the difference of the potentials at each point, combining the uncertainties via an RMS calculation (also referred to as error propagation in quadrature). Then, this uncertainty in the difference was scaled by the magnitude of the RDF at that point and the normalization before combining via an RMS calculation to give the uncertainty used to calculate the confidence ranges shown for each point.

4.3 Results and Discussion

4.3.1 Numerical finite differences

Before the performance of the single point formulae developed in the work is evaluated, it is worth evaluating the noise and performance of the existing numerical FD calculations. Figure 4.1 compares the MTFD and the RFD with confidence ranges for the sensitivity of the single site methanol CG potential to changes in the charge on the hydroxyl's hydrogen. As expected, both estimates agree within the confidence ranges for sufficiently small changes to the charge, but the RFD has significantly smaller confidence ranges than the MTFD, as expected because small differences in the MTFD denominator magnify sampling noise. In fact, the RFD confidence ranges are more than 100 times smaller than for MTFD. For the purposes of initially verifying the precision of our single point formula, only RFD with confidence ranges will be shown since it is expected that any predicted sensitivity that agrees with the RFD within the confidence ranges will also agree with the MTFD. However, this is not always the case, especially when the RFD calculations are strongly biased, so to demonstrate the accuracy of the single point formulae, comparisons will be made to MTFD or independent trajectories (IT) later in the paper.

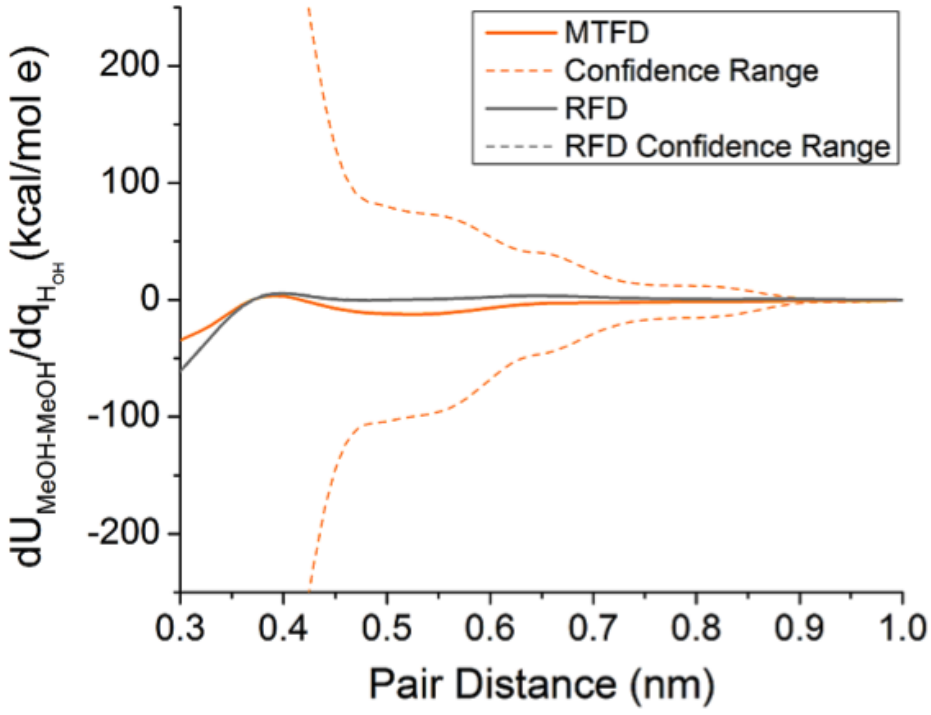

Figure 4.1: Comparison of multi-trajectory FD (MTFD) and reweighted FD (RFD) for the sensitivity of the MeOH CG potential to changes in the charge on the hydroxyl hydrogen (HOH). Confidence ranges show the relative noise of each estimated sensitivity, as defined in the main text. The RFD confidence range is so small relative to the MTFD confidence range that it is not distinguishable from the RFD curve on this scale.

4.3.2 Single Site Methanol

The sensitivities calculated using the SCB single point formula are compared to the SCI single point formula as well as the RFD sensitivity estimates with confidence ranges in Figure 4.2. For LJ epsilon (Figure 4.2a and Figure 4.2b) and sigma (Figure 4.2c), the SCB and SCI estimates superimpose, indicating that non-pair-representable many-body effects play little role in the pair-representable part of these sensitivities. The SCB and SCI estimates for these graphs are generally within the shown confidence range and only slightly overestimate the magnitude of the sensitivity at short interaction pair distances. For sensitivity to charge (Figure 4.2d), the SCI estimate is significantly different from either the SCB estimate or the actual RFD sensitivity. This difference between SCB and SCI estimates indicates that

significant multibody correlations are important for charge interactions and correlations. These observations agree with and clarify prior work that indicated that CG potentials are significantly less transferable, in the naive sense, for charge interactions than epsilon and sigma interactions because of the significant multibody correlations present.[317]

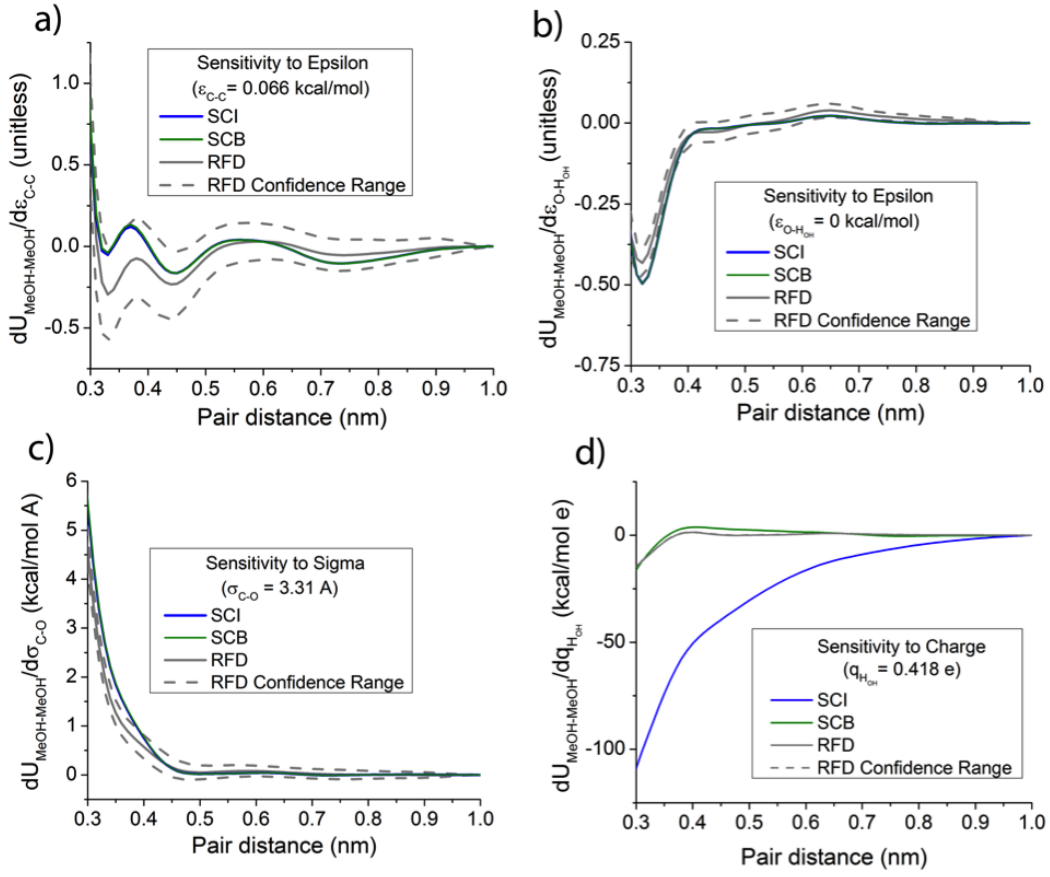

Figure 4.2: Comparison of methanol sensitivity estimates for different interaction parameters between RFD, self-consistent iterative (SCI) single point, and self-consistent basis (SCB) single point calculations. Sensitivities are taken with respect to a) Carbon-Carbon (C-C) LJ epsilon, b) Oxygen-Hydroxyl-Hydrogen (O-HOH) LJ epsilon, c) Carbon-Oxygen (C-O) LJ sigma, and d) Hydroxyl Hydrogen (HOH) charge interaction parameters. RFD confidence ranges are calculated as defined in the main text. The RFD confidence range for d) is so small that it is not visible on this scale.

As mentioned in the previous section, potentials can be predicted using sensitivities from

either of the single point formulae using

$$U(\mathbf{R}^N;\lambda+\delta\lambda) = U(\mathbf{R}^N;\lambda) + \delta\lambda\left(\frac{U(\mathbf{R}^N;\lambda)}{d\lambda}\right), \tag{4.8}$$

which is a simple correction to the original potential that is linear in $\delta\lambda$. The magnitude of change in CG interaction potential from the reference (REF) parameterization as plotted in Figure 4.3 was calculated as

$$|\Delta U| = \frac{1}{R_H - R_L}\int_{R_L}^{R_H} dR|U_{PRED}(\mathbf{R}^N;\lambda+\delta\lambda) - U_{REF}(\mathbf{R}^N;\lambda)|g_{REF}(R). \tag{4.9}$$

In this definition, $g_{REF}(R)$ is the radial distribution function of the reference parameterization, $U_{PRED}(\mathbf{R}^N;\lambda+\delta\lambda)$ is the interaction potential at a non-reference parameterization either from FMing independent FG NVT trajectories with the modified parameterization or using Equation 4.8 with the sensitivity calculated using RFD, SCI, or SCB formulae. Figure 4.3 shows the difference in potentials for different $\delta\lambda$'s from the original ($\delta\lambda = 0$) potential as described in the previous section. For the epsilon graph (Figure 4.3a), both the SCB and SCI curves agree with the reweighted curve for small differences–i.e., in the linear regime. It is remarkable that both the sensitivities have the same average slope as the curve for the CG potential determined for the new parameters with independent trajectories for changes of only 0.01 kcal/mol. For the sigma graph (Figure 4.3b), the reweighted curves are nonlinear, but the SCB sensitivity appears to have the same average slope as the SCI sensitivity and the reweighted curve. The independent trajectories curve has a similar initial slope to the single point sensitivities, but is below the single point sensitivities for larger changes. This is somewhat expected since as perturbations increase, systems will typically make compensating changes that result in a concave response. For the charge graph (Figure 4.3c), the reweighted curve shows nonlinearity, but the SCB curve is nonetheless reasonably consistent with the reweighted curve. As expected from the sensitivity comparisons, the SCI curve drastically overestimates the change in potential. While neither of the single point

sensitivities matches the independent trajectories for charge, neither does the reweighted curve beyond 0.005 e, indicating significant sampling changes in response to charge modification that may be the result of changes in complex many-body and long range effective interactions.

Another way to assess the accuracy of these predicted potentials is to compare the RDFs generated from CG simulations using both the actual CG potential and the potentials predicted from both sensitivity formulae. Figure 4.4 shows the RDF for a selected set of $\delta\lambda$. It is clear that any slight errors shown in Figure 4.3 for predictions across epsilon (Figure 4.3a) and sigma (Figure 4.3b) value do not lead to noticeable errors in the RDFs. For predictions across charge (Figure 4.3c) values, the RDF from the SCB predicted potential has only minor deviations in the height of the first peak and valley from the actual RDF. The agreement of the RDF from the SCB predicted potential bodes well for the application of sensitivity for generating predicted potentials. However, there is a limit to how far one can use these predicted potentials, which corresponds to the breakdown of the first order approximation of the sensitivity outside the linear regime. Figs. 4.3d, 4.3e, and 4.3f show that for sufficiently large changes in the FG interaction parameters, the magnitude of the RDF peaks and valleys differ significantly from those of the RDF obtained using the actual CG potential with the modified interaction parameters. Nonetheless, the RDFs for larger interaction parameter changes continue to show good agreement with the location of the peaks and valleys for the first two solvation shells.

4.3.3 Solvent Free Sodium Chloride

The sensitivities calculated using the SCB single point formula are compared to SCI single point formula and the RFD sensitivity estimates with confidence ranges in Figure 4.5. For epsilon (Figure 4.5a and Figure 4.5b) and sigma (Figure 4.5c), the SCB and SCI sensitivities are entirely within the shown confidence ranges. It is interesting to note that the magnitude of the SCI sensitivity is less than the magnitude of the SCB sensitivity when they deviate in

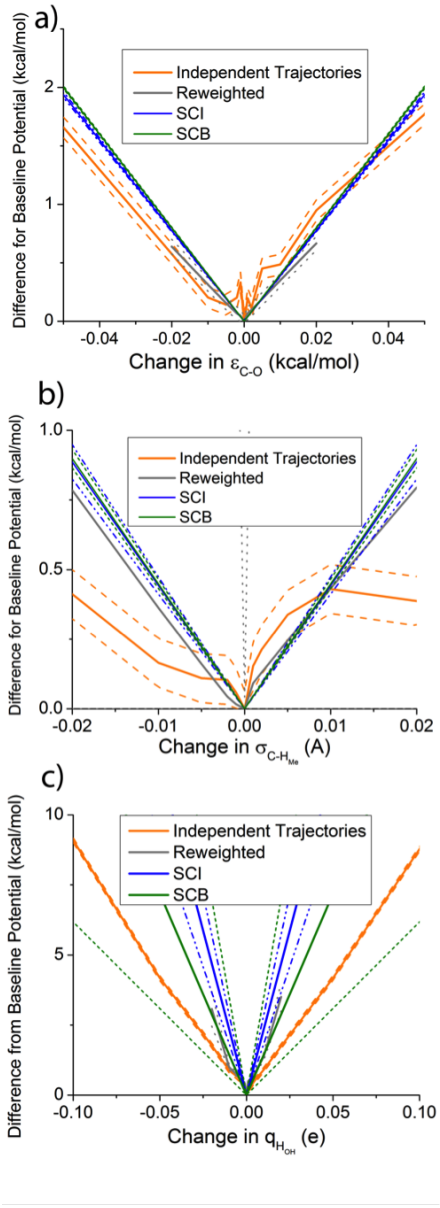

Figure 4.3: Magnitude of change in methanol CG interaction potential from OPLS parameterization, calculated (see Equation 4.9) as a weighted average absolute difference in predicted potential from a reference potential weighted by the reference RDF, for predictions via independent trajectories, reweighting, and the two single point sensitivities SCI and SCB. Predictions are compared for changes in a) Carbon-Oxygen (C-O) LJ epsilon, b) Carbon-Methyl-Hydrogen (C-HMe) LJ sigma, and c) Hydroxyl Hydrogen (HOH) charge interactions.

Figure 4.5a-c at short interaction distances, indicating that the SCB covariance correction is greater than the SCI transport correction because of differences in the amount of sensitivity captured due to the different basis set considerations between the two formulae. For

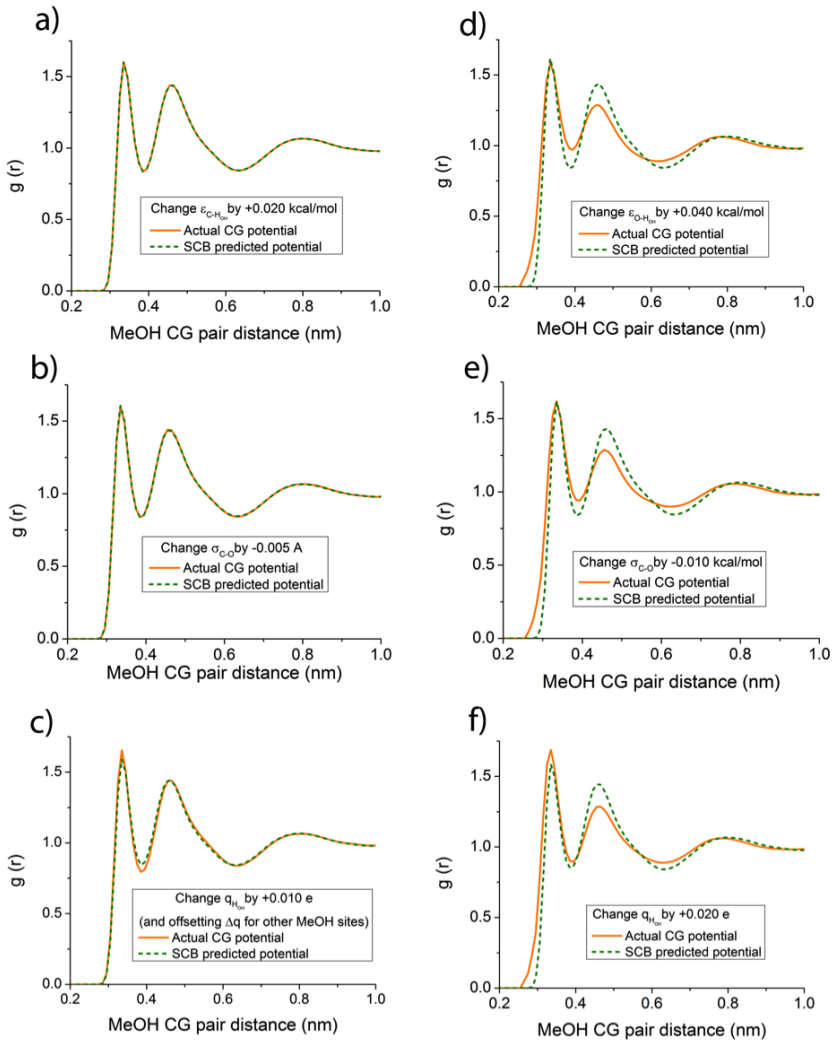

Figure 4.4: Radial distribution functions (RDFs) from CG methanol simulations. a) Changing Carbon-Hydroxyl-Hydrogen (C-HOH) LJ epsilon interaction parameter by 0.020 kcal/mol. b) Changing Carbon-Oxygen (C-O) LJ sigma interaction parameter by -0.005 Å. c) Changing the Hydroxyl Hydrogen's charge by +0.010 e and applying neutralizing charges on the other methanol FG sites. d) Changing Carbon-Hydroxyl-Hydrogen (C-HOH) LJ epsilon interaction parameter by 0.040 kcal/mol. e) Changing Carbon-Oxygen (C-O) LJ sigma interaction parameter by -0.010 Å. f) Changing the Hydroxyl Hydrogen's charge by +0.020 e and applying neutralizing charges on the other methanol FG sites.

sensitivity to charge (Figure 4.5d), the SCI estimate is significantly below the actual RFD sensitivity and the SCB sensitivity, which is above the RFD confidence range for large interaction distances and below it for intermediate interaction distances. However, both the SCB and SCI sensitivities show much better qualitative agreement for the sensitivity to charge

in the sodium chloride system than in the methanol system, suggesting that the effects of electrostatics are more pair-representable in this system.

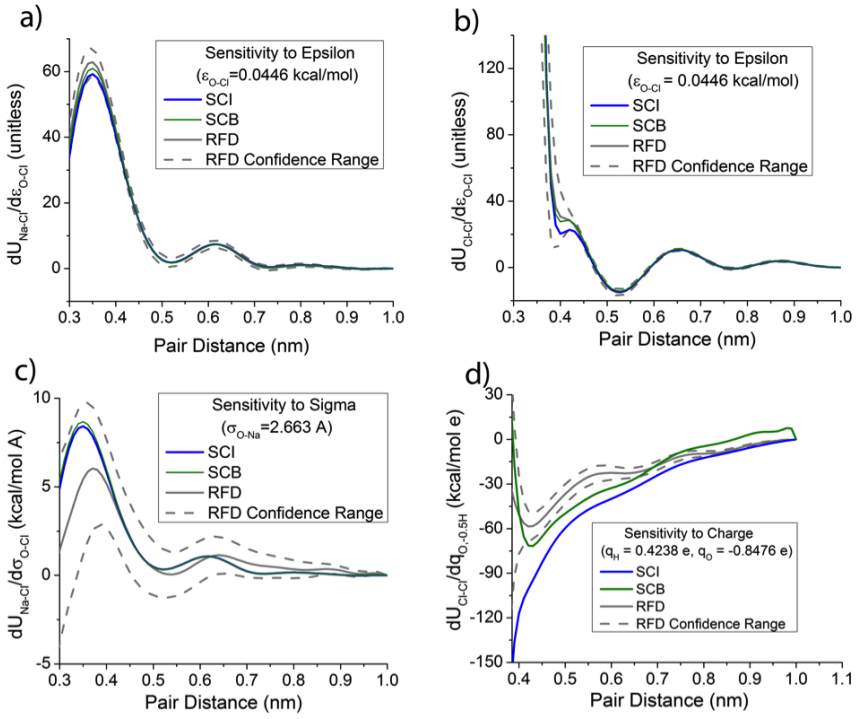

Figure 4.5: Comparison of solvent free sodium chloride Na-Na, Na-Cl, and Cl-Cl interaction potential sensitivities estimated for different interaction parameters between RFD, SCI single point, and SCB single point calculations. Sensitivities of the a) Na-Cl CG potential to the FG Oxygen-Chloride (O-Cl) LJ epsilon, b) Cl-Cl CG potential to the FG Oxygen-Chloride (O-Cl) LJ epsilon, c) Na-Cl CG potential to the Oxygen-Chloride (O-Cl) LJ sigma, and d) Cl-Cl CG potential to the water Oxygen and Hydrogen charge interactions.

Figure 4.6 shows the difference in potentials (Equation 4.9) for different $\delta\lambda$'s from the original ($\delta\lambda = 0$) potential as described in Section II. For the epsilon graph (Figure 4.6a), the SCB curve agrees with the reweighted curve for small changes until the nonlinearities appear in the reweighted curve, where bias is more of a problem. The SCB curve is also in the same range as the independent trajectories for large changes. The SCI curve deviates from both reference curves for sizable changes, but appears to have the same initial slope as the independent trajectories, which is likely within the linear regime. For the sigma graph (Figure 4.6b), the SCB curve shows even better agreement with the reweighted curve

and the average slope of the independent trajectories curves than in Figure 4.6a. The large difference between the SCI and SCB curves indicates the importance of CG basis set effects in capturing the correlations important for larger changes in parameters. For the charge graph (Figure 4.6c), the reweighted curve looks quite linear and shows agreement with the independent trajectories only for small changes. The difference between the SCB and reweighted curves from the independent trajectories may be due to underestimation of many-body charge screening effects that the reweighted and SCB methods do not incorporate due to configurational sampling bias. The SCB curve agrees with the reweighted curve, but both overestimate changes in the potential. The SCI curve here seems to have the same average slope as the independent trajectories for positive and negative changes, indicating either less bias or a cancellation of errors.

Figure 4.7 shows the RDFs for a selected set of $\delta\lambda$. For the sensitivity to epsilon example the heights of the first peaks predicted from the single-point sensitivities for the Na-Na and Cl-Cl RDFs (not shown, see SI) are slightly overstructured, but the opposite is true for the Na-Cl RDF (Figure 4.7a). The opposing errors in the RDFs and potentials illustrate the additional problems of fitting the three nonbonded interactions simultaneously. For the sensitivity to sigma example (Figure 4.7b), the RDF from the SCB predicted potential seems to be in agreement with the actual CG RDF with only minor understructuring of the contact-ion pair. When it comes to the sensitivity to charge (Figure 4.7c), it is clear that the errors in the sensitivity of the potential carried through to the RDFs as they are all uniformly overstructured.

It is not all that surprising that the charge sensitivities from the single point formulae overstructure the ions since this amounts to an underestimate of a many-body screening effect from waters that were coarse-grained out of the simulation. This continues to get worse for larger changes to the charges on the water (Figure 4.7f). The corrections to the water screening and structure are likely manifested in many-body interactions that are beyond the range of these first order, pair-representable sensitivities. Fortunately, the agreement

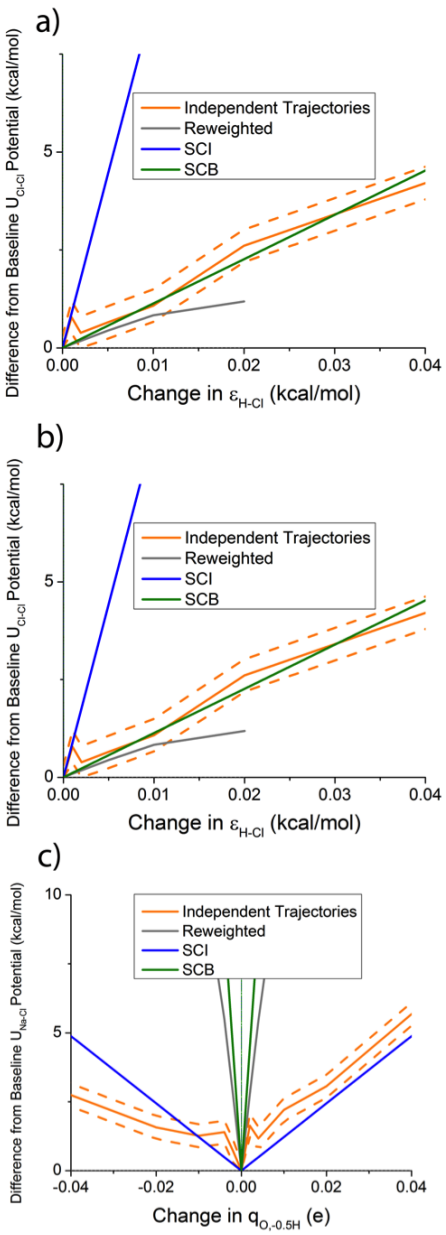

Figure 4.6: Magnitude of change in sodium chloride CG interaction potential from SPC/E water and Joung and Cheatham NaCl parameterization,[169] calculated as a weighted average absolute difference in predicted potential from a reference potential weighted by the reference RDF, for predictions via independent trajectories, reweighting, and the two single point sensitivities SCI and SCB. Predictions of a) $U_{\text{Cl-Cl}}$ to changes in Hydrogen-Chloride (H-Cl) LJ epsilon, b) $U_{\text{Na-Cl}}$ to changes in Oxygen-Chloride (O-Cl) LJ sigma, and c) $U_{\text{Na-Cl}}$ to changes in Water Hydrogen and Oxygen charges are compared.

between the RDFs to epsilon (Figure 4.7d) and sigma (Figure 4.7e) parameters is quite good over a larger range of parameter changes. As with the methanol system, the heights of the

RDF peaks and valley differ only slightly so for epsilon and sigma given the magnitude of the parameter change while the location of the peaks and valley agree quite well. Thus, using the sensitivities from less highly correlated interactions such as the LJ nonbonded interactions appears to lead to reasonable predicted CG potentials and CG RDFs.

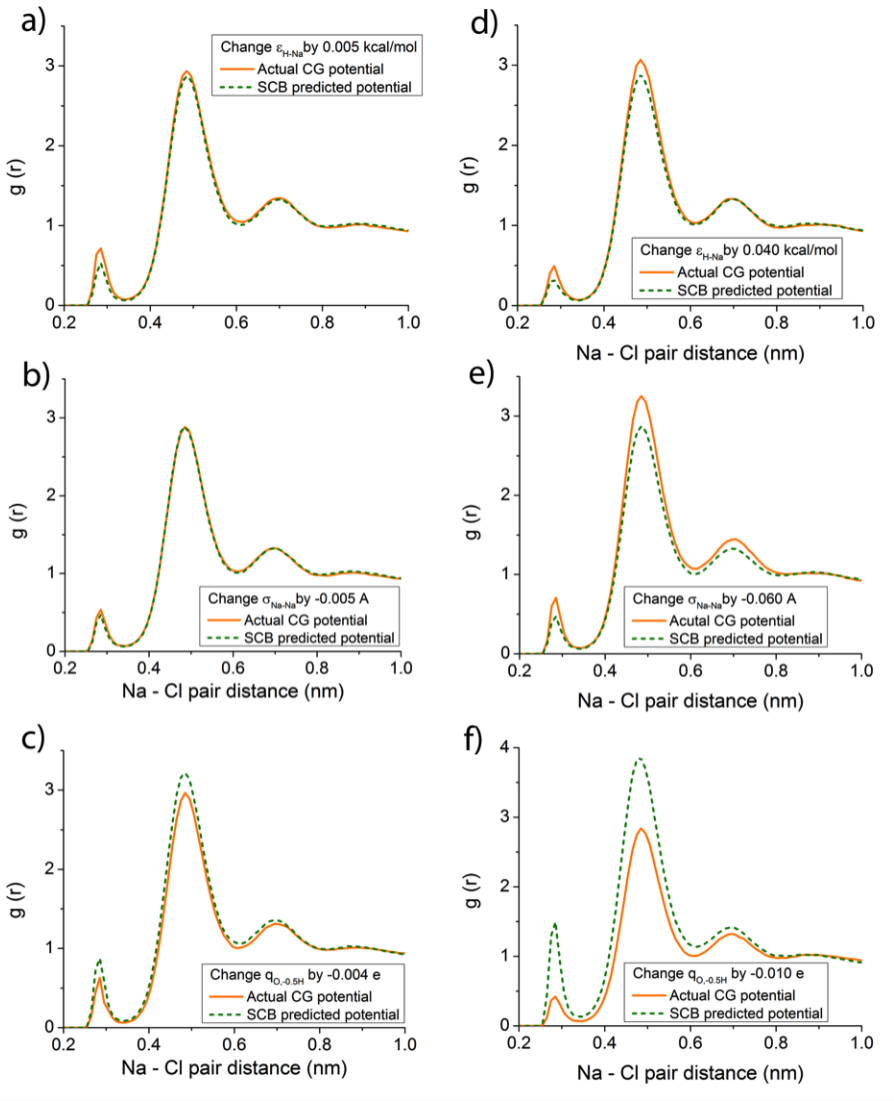

Figure 4.7: Radial distribution functions (RDFs) from CG sodium chloride simulations for the Na-Cl pair distance. a) Changing Oxygen-Sodium (O-Na) LJ epsilon interaction parameter by 0.005 kcal/mol. b) Changing Sodium-Sodium (Na-Na) LJ sigma interaction parameter by -0.005 Å. c) Changing the Water Oxygen charge by -0.004 e and the Water Hydrogen by +0.002 e. d) Changing Oxygen-Sodium (O-Na) LJ epsilon interaction parameter by 0.040 kcal/mol. e) Changing Sodium-Sodium (Na-Na) LJ sigma interaction parameter by -0.060 Å. f) Changing the Water Oxygen charge by -0.010 e and the Water Hydrogen by +0.005 e.

4.3.4 Discussion

In general, the sensitivities calculated with respect to LJ epsilon parameters show excellent agreement with the RFD sensitivities and the CG RDFs for both the CG methanol and solvent free sodium chloride systems. The agreement in the difference from the baseline potential in Figure 4.3a even well outside the linear regime is particularly noteworthy. The slightly more correlated sensitivities with respect to sigma parameters showed good agreement for the CG methanol system, better agreement than in the sodium chloride system that is more highly coarse-grained. However, the charge sensitivities are only qualitatively correct, reflecting the highly complex many-body correlations due to these long-range interactions. In the sodium chloride system, this was reflected in the overstructuring of the CG ion pair RDFs, but the fact that sensitivities to the charge on the atoms in the implicit water molecules are reasonable is promising. Unfortunately, agreement with RFD implies that while these formulae significantly improve on the signal to noise ratio of reweighted finite differences, in many cases they do not address the problem of bias. This work reveals that bias in these estimators, not noise, is the next truly difficult problem to overcome, and it will not be overcome simply with more data as we attempted with this per-particle derivative-matching approach; bias is better dealt with by acquiring better data[77] or more sophisticated estimators.[111, 166, 59]

One can expect that the SCB sensitivities will be the most accurate and predictive when the relevant interaction is short range and does not involve many-body correlations as is the case for any short range pair interaction modeling van der Waals interactions such as the Lennard-Jones potential. In contrast, interactions that are long range and involve many-body correlations such as charge interactions are expected to be less accurate and less predictive. While the sensitivity of intramolecular interactions such as bonds, angles, and dihedrals were not investigated, physical considerations suggest that these sensitivities will not be as well-behaved as those for Lennard-Jones parameters because intra-CG-site interactions affect the CG interactions only indirectly through multi-atom correlations, but

on the other hand they should be better than the sensitivities for charge interactions because intramolecular interactions are short ranged. Likewise, sensitivity of state parameters such as temperature and volume as well as their conjugate observables entropy and pressure were not investigated here, but we expect the sensitivities to these parameters to be inaccurate because the relevant interactions are long range and involve many-body correlations that may not be pair-representable.

One way to check these heuristics is to compare the SCB estimate to the SCI estimate for each parameter of interest in a given system. If the SCB and SCI estimates agree as in Figure 4.2a-c and Figure 4.5a-c, then the many-body correlations that lead to the correction of the naive sensitivity which the two formulae estimate differently are well-represented in the chosen basis set. This means that the SCB sensitivity is more likely to be accurate and predictive. This feature implies that these SCI and SCB formulae can be used diagnostically to evaluate hypotheses about transferability even when they do not provide accurate linear sensitivity measurements. However, when they do, they also provide previously infeasible checks on the precise size of the linear regime: in that case, this regime will be precisely the range over which the RFD and SCB estimates agree.

The times when the SCB and SCI single point formulae differ in their sensitivity estimates are significant because the two capture correlated many-body effects in different ways. In particular, any significant correction to the naive sensitivity from either single point formula indicates the presence of significant multi-body effects and correlations in the sensitivity to that parameter. Differences between them, moreover, specifically indicate that the non-representable many-body effects are folded into the corresponding representable force field for a given basis, which are sensitive to the parameter under study in ways that cannot be represented within that basis set. The difference in the estimates for the sensitivity to epsilon in the solvent free NaCl system, but not the CG methanol system, reflects that the NaCl system is more highly coarse-grained as expected. Also, the difference between the SCB and SCI sensitivity estimates for charge interactions quantifies and confirms the dependence of

the effective pair potential on the significant and complex many-body correlations among long range interactions that had been hypothesized in previous literature.[317]

Noise and modeling error can still remain problematic even when these formulae perform without significant bias. Approaches previously used to improve force matching's ability to deal with these problems could also potentially be brought to bear to improve these sensitivity calculations. For instance, regularization of these sensitivity estimators similar to Lu et al.'s approach could improve the performance of these estimates with noisy data.[201, 344] Alternatively, it may be that structural differences between the CG and FG models due to the use of a finite basis set lead to prediction errors that could be ameliorated via recalculation of the input sensitivity derivatives, $M^{+}(d\nabla u(\mathbf{r}^n;\lambda)/d\lambda)$ and $du(\mathbf{r}^n;\lambda)/d\lambda$, using statistics from CG sampling as well as FG sampling, as in iterative FM and iterative g-YBG approaches.[74, 200, 284]

More systematic study of the differences between calculations by these two formulae may reveal more of the character of important many-body effects in various systems. While we focused on well-known interaction parameters here, it is possible to use these formulae to investigate the addition of arbitrary biases to FG models to examine the effects of arbitrary correlations on the representable correlations in a system. This could reveal interesting experimental control parameters for fine-grained systems, e.g., in discovering manipulable fields conjugate to CG tetrahedrality correlations in water.[230, 14, 301] Furthermore, while we studied only pair-representable force fields here and considered only fixed CG basis sets, we can also use it as a criterion for basis set quality, indicating that it could also be useful for basis set design. One can use these formulae with arbitrarily complex CG basis sets and the comparisons between SCI and SCB formulae in various basis sets to choose the ones that will result in the most transferable models across a given parameter space.

Finally, the results indicate that the calculation of sensitivities to nonbonded interaction parameters is good for generating predicted potentials, especially for sensitivities to LJ epsilon and sigma interaction parameters. A direct extension of this work would be to cal-

culate thermodynamic derivatives by repeating the derivations in Section II with $\lambda = \beta$, the thermodynamic temperature. Since the formulae can only be used to calculate sensitivities to continuous parameters, sensitivities to volume $\lambda = V$ could be calculated in the NPT and PT ensembles while sensitivities to concentration could be calculated via sensitivities to chemical potential $\lambda = \mu$ in the grand-canonical μPT and μVT ensembles. Another extension of the work would be to calculate the sensitivity of other CG properties or observables such as the RDF $g(r)$ by applying a chain rule, where the sensitivity of the CG property or observable to the CG potential would be multiplied by the sensitivity of the CG potential to FG interaction parameters as presented in this work.

4.4 Conclusions

In this article, new reweighting-free formulae for the calculation of the sensitivity of CG potentials and force fields to changes in the underlying FG models' interaction parameters and state point were presented that require only a single trajectory for calculation. In the results, the SCB estimates predicted sensitivities to LJ epsilon and sigma parameters that were quantitatively correct to within the confidence ranges from RFDs, representative of the practical state of the art. The single point formula does not require *a priori* knowledge of the linear sensitivity regime for a given parameter and can be useful in generating predicted CG potentials for other interaction parameters, as demonstrated by the agreement of the CG RDFs from independent trajectories with predicted potentials using this sensitivity. Of the predictive sensitivity measures examined here, the single point methods provide the lowest noise estimates of all, providing the same sensitivities as RFDs at reduced computational cost, with comparable bias, and without ambiguity concerning the size of the linear regime.

Finally, beyond their purely computational significance, these results also serve to shed light onto relatively unexplored subtleties of CG representability. Consideration of both the SCB and SCI formulae offers a new window onto the fundamental theoretical problems of representing transfer of CG models between state points, providing a vivid example of how

the change from FG model to FG model in the CG-representable part of the correlations may not always be the same as the CG-representable part of the change in correlations from FG model to FG modeleven at the level of linear response. While this is sometimes mentioned in discussions of the foundations of coarse-graining, that subtlety is rarely investigated as a practical effect with fundamental physical importance in its own right. However, establishing one's foundations is also an eminently practical thing to do. We hope this work will spur deeper investigations into the theoretical interplays between representability and transferability, two of the most important challenge areas in state of the art of coarse-graining and CG modeling, in addition to providing new computational tools.

4.5 Appendix A: Derivation of the SCB single-point formula

The self-consistent basis single-point sensitivity formula describes the derivative with respect to system parameters of variationally force matched finite-basis approximations to the true many-body FES. The usual force-matching normal equations for a PMF approximated as a linear combination of a set of basis functions ψ_d with coefficients ϕ_d and λ are

$$\frac{1}{N_t N}\sum_{t=1}^{N_t}\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\phi_{d';\lambda}=\frac{1}{N_t N}\sum_{t=1}^{N_t}\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\mathbf{f}_I(\mathbf{r}_t^n;\lambda), \tag{4.10}$$

where $\mathbf{f}_I$ are CGed forces from sampled atomistic configurations. This is an equation valid for sampling from a system with fixed parameter λ. In order to take the derivative of this with respect to λ, one must find a way to express the sampling density as a differentiable function of λ in this expression. One option is to assume all sampling is run at a reference λ_0 and then perform a weighted least-squares optimization using the reweighting factors (see Equation 4.4 in the main text) rather than a uniformly-weighted least-squares optimization.

Using the usual equations for weighted least squares, one gets the normal equations

$$\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\phi_{d';\lambda} = \frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\mathbf{f}_I(\mathbf{r}_t^n;\lambda), \tag{4.11}$$

which are in principle valid for any λ with samples taken with respect to any λ_0, though of course only practical when the two are close. First define

$$G_{d,d'}(\lambda,\lambda_0) = \frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\frac{d\psi_{d'}}{d\mathbf{R}_I} \tag{4.12}$$

$$b_d(\lambda,\lambda_0) = \frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\mathbf{f}_I(\mathbf{r}_t^n;\lambda). \tag{4.13}$$

Now, taking the derivative of both sides, one gets

$$\sum_{d'} G_{d,d'}(\lambda,\lambda_0)\frac{d\phi_{d',\lambda}}{d\lambda} = \frac{db_d(\lambda,\lambda_0)}{d\lambda} - \sum_{d'}\frac{dG_{d,d'}(\lambda,\lambda_0)}{d\lambda}\phi_{d',\lambda} \tag{4.14}$$

This is a force-matching-like equation for the change in the expansion coefficients, which give the change in the PMF when multiplied by the basis functions. The equation matches the change from the true target forces, b_d, from the predicted forces with fixed PMF, $\sum_{d'} G_{d,d'}\phi_{d'}$, and adjusted sampling.

Evaluating the derivatives is easiest after re-expanding the new notation

$$\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\frac{\phi_{d';\lambda}}{d\lambda} = \dots$$
$$\dots\frac{d}{d\lambda}\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\mathbf{f}_I(\mathbf{r}_t^n;\lambda)\dots$$
$$\dots-\frac{d}{d\lambda}\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\phi_{d';\lambda^*}, \tag{4.15}$$

$$\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\frac{\phi_{d';\lambda}}{d\lambda} = \dots$$
$$\dots\frac{d}{d\lambda}\frac{1}{N_t N}\sum_{t=1}^{N_t} w(\mathbf{r}_t^n;\lambda,\lambda_0)\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\frac{d\mathbf{f}_I(\mathbf{r}_t^n;\lambda)}{d\lambda}\dots$$
$$\dots-\frac{1}{N_t N}\sum_{t=1}^{N_t}\frac{dw(\mathbf{r}_t^n;\lambda,\lambda_0)}{d\lambda}\sum_{I=1}^{N}\frac{d\psi_d}{d\mathbf{R}_I}\cdot\left(\sum_{d'}\frac{d\psi_{d'}}{d\mathbf{R}_I}\phi_{d';\lambda^*}-\mathbf{f}_I(\mathbf{r}_t^n;\lambda)\right), \tag{4.16}$$

which is just a weighted force matching for the newly-apparent framewise sensitivities of forces in the parentheses, which are straightforward to calculate after force-matching first to find $\phi_{d,\lambda}$. Using a finite sum with some large number of samples provides a practical calculation scheme. Replacing the sums with ergodic averages, however, in the complete basis set limit these normal equations correspond to Equation 4.5, the covariance-like SCB formula described in the main text. In a finite basis set and in the long time limit, it corresponds to the λ-derivative of the g-YBG equations.

4.6 Appendix B: Derivation of the SCI single-point formula

The self-consistent iterative single-point sensitivity formula is based on using a finite basis set to represent the per-particle-position derivatives of the full many-body sensitivity. To

derive this, start with the definition of the many-body CG sensitivity

$$\frac{dU(\mathbf{R}^N;\lambda)}{d\lambda} = \left\langle \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} \right\rangle_{\mathbf{R}^N;\lambda}, \tag{4.17}$$

take the derivative with respect to all CG particle positions

$$\frac{d\nabla_R U(\mathbf{R}^N;\lambda)}{d\lambda} = \nabla_R \left(\frac{\int d\mathbf{r}^n \delta(\mathbf{R}^N - M(\mathbf{r}^n)) \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} e^{-\beta u(\mathbf{r}^n;\lambda)}}{\int d\mathbf{r}^n \delta(\mathbf{R}^N - M(\mathbf{r}^n)) e^{-\beta u(\mathbf{r}^n;\lambda)}} \right) \tag{4.18}$$

apply the product rule to see

$$\begin{aligned} \frac{d\nabla_R U(\mathbf{R}^N;\lambda)}{d\lambda} = & \left(\frac{\int d\mathbf{r}^n \nabla_R \delta(\mathbf{R}^N - M(\mathbf{r}^n)) \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} e^{-\beta u(\mathbf{r}^n;\lambda)}}{\int d\mathbf{r}^n \delta(\mathbf{R}^N - M(\mathbf{r}^n)) e^{-\beta u(\mathbf{r}^n;\lambda)}} \right) \cdots \\ & \cdots - \left\langle \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} \right\rangle_{\mathbf{R}^N,\lambda} \left(\frac{\int d\mathbf{r}^n \nabla_R \delta(\mathbf{R}^N - M(\mathbf{r}^n)) e^{-\beta u(\mathbf{r}^n;\lambda)}}{\int d\mathbf{r}^n \delta(\mathbf{R}^N - M(\mathbf{r}^n)) e^{-\beta u(\mathbf{r}^n;\lambda)}} \right) \end{aligned} \tag{4.19}$$

and simplify using the integration by parts formulas used in Ref.[90] to get

$$\begin{aligned} \frac{d\nabla_R U(\mathbf{R}^N;\lambda)}{d\lambda} = & \left\langle M^+ \left(\frac{d\nabla u(\mathbf{r}^n;\lambda)}{d\lambda} \right) \right\rangle_{\mathbf{R}^N;\lambda} \cdots \\ & \cdots - \left\langle \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} M^+ \left(\nabla \beta u(\mathbf{r}^n;\lambda) \right) \right\rangle_{\mathbf{R}^N;\lambda} \cdots \\ & \cdots + \left\langle \frac{du(\mathbf{r}^n;\lambda)}{d\lambda} \right\rangle_{\mathbf{R}^N;\lambda} \left\langle M^+ \left(\nabla \beta u(\mathbf{r}^n;\lambda) \right) \right\rangle_{\mathbf{R}^N;\lambda} \end{aligned} \tag{4.20}$$

Finally, rearrangement and grouping leads to Equation 4.7, the transport-like SCI equation in the main text. This corresponds to the λ-derivative of the g-YBG equations with a complete basis set.

CHAPTER 5

THE REPRESENTATION OF OBSERVABLES IN COARSE-GRAINED MODELS

This chapter reproduces material from a paper Wagner, J. W., Dama, J. F., Durumeric, A. E. P., and & Voth, G. A. currently in preparation; Jacob Wagner and I again will share first authorship; we independently derived all of the theory in the paper and collaborated extensively on its structure and writing. Alek Durumeric provided valuable feedback and editing aid throughout.

This section describes how to define observables at the coarse-grained level that are consistent with the fine-grained level, which is most important when the form of the observable seems clear by analogy to the fine-grained system but actually requires more care. For instance, coarse-grained potentials are commonly interpreted as internal energies though they include entropic contributions.

In our social system of families with various preferences for neighbors, the relevant observables might include things like sensitivity to housing availability (pressure), tendency to settle (temperature), association with like neighbors (radial distribution function), and so on. Even if individuals have preferences independent of their preferences to settle for any one of them, the family's overall preferences might shift as overall preferences to settle shift, since it may affect the order in which the family members are willing to settle and thus the aggregation function of their settling-independent preferences.

Returning to income, we might ask: if we're just looking at where families choose to live, could we infer their income? Could we infer the type of income they're earning and who in the family is earning it? Obviously there will be limits to these inferences, but there are many cases where knowing a person's neighborhood reveals a great deal about them.

5.1 Introduction

Whatever its resolution, a computational model achieves its physical significance from comparison to experiment and this comparison is between experimental observables and corresponding model observables. Therefore, the relationship between each model's observables and experimental observables must first be firmly established in order for these models to be as meaningful as possible.

Often, however, the relationships between models and experiment can be unclear. Common statistical mechanical results and intuitive structural relationships establish connections between atomistic models and experiment,[12, 131, 84, 168, 208] but the model's connection to experiment ultimately depends on the resolution of the model, as some authors have attempted to make clear.[309, 197, 165] This concept was applied, for example, by D'Adamo et al.,[86] Das and Andersen,[95] and Dunn and Noid,[106] who studied how observable representations may change based on resolution and thermodynamic ensemble. While this literature focuses on thermodynamic observables,[16, 309, 197, 58, 165, 315, 95, 86, 106] the issue of CG observable representation is more fundamental: it concerns every aspect of CG model interpretation involving comparison with experiments or FG physics. Thus, the recognition that CG observables are not simply analogs of their FG counterparts is fundamental to understanding and ultimately addressing the issue of representability in coarse-graining. This recognition is commonly overlooked or ignored in the CG modeling and simulation literature. A more nuanced understanding is therefore needed to interpret CG models so that they have more meaningful connections to experiment.

Often, models are parameterized using observable constraints so that they reproduce a given experimental observable using a chosen observable expression.[233, 154, 239] Any one experimental observable can always be reproduced this way. However, models need to reproduce several experimental observables simultaneously, and a model cannot reproduce several observables simultaneously if their corresponding constraints conflict. For example, the Henderson uniqueness theorem guarantees a unique radial distribution given a pair potential.[141]

To reproduce any additional experimental observable such as the pressure with this fixed pair potential, the model observable must be able to reproduce the experimental values using the previously determined pair potential;[165] otherwise, the observables are incompatible. While adding three-body interactions[185, 97, 230] or density dependence[151, 9, 10, 11] can improve observable compatibility tradeoffs, it is not computationally tractable to add additional interactions indefinitely. Thus, one needs to choose the set of observable expressions carefully if the model is to successfully reproduce all of the corresponding experimental observables simultaneously.

Systematic coarse-graining (e.g., the multiscale coarse-graining methodology[154, 153, 248, 247, 94, 177, 95, 185, 204, 97, 96, 98, 60] as one such case) provides an illuminating case in which to investigate CG observable compatibility. In systematic coarse-graining, a CG model of arbitrary resolution is defined in terms of a given FG model using a mapping and an effective potential. The mapping M defines CG configuration variables $\mathbf{R}^N$ in terms of the FG configuration variables $\mathbf{r}^n$ as the product of the individual CG particle maps M_I for each CG particle $\mathbf{R}_I$, such that $\delta(M(\mathbf{r}^n) - \mathbf{R}^N) = \prod_{I=1}^{N} \delta(M_I(\mathbf{r}^n) - \mathbf{R}_I)$.[246] Then, the effective CG potential $U(\mathbf{R}^N)$ for a specific CG configuration $\mathbf{R}^N$ in terms of the FG potential $u(\mathbf{r}^n)$ is given by[204]

$$e^{-\beta U(\mathbf{R}^N)} \sim \int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)} \tag{5.1}$$

where $\beta = (k_B T)^{-1}$, k_B is Boltzmann's constant, and T is the temperature. The CG observable A corresponding to an FG observable a should ideally ensure that the FG ensemble average of an observable property $a(\mathbf{r}^n)$ given the FG potential $u(\mathbf{r}^n)$ matches the CG Boltzmann ensemble average of $A(\mathbf{R}^N)$ in terms of the CG potential $U(\mathbf{R}^N)$. One natural choice that satisfies this requirement is

$$A(\mathbf{R}^N) = \langle a(\mathbf{r}^n) \rangle_{\mathbf{R}^N} \equiv \frac{\int d\mathbf{r}^n a(\mathbf{r}^n) \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)}}{\int d\mathbf{r}^n \delta(M(\mathbf{r}^n) - \mathbf{R}^N) e^{-\beta u(\mathbf{r}^n)}} \tag{5.2}$$

There may be other choices for the CG observable that would ensure that the FG and CG ensemble averages are equal, but few are so immediately obvious. Moreover, this choice and no other guarantees the equality of the ensemble average between the CG and FG observables under any possible bias potential applied to the CG degrees of freedom. In addition to ensuring the equivalence of ensemble averages, Equation 5.2 also implies that the CG observable for a given CG configuration is equal to the ensemble average of all FG configurations that map to the given CG configuration. Thus, the relationship here between FG and CG observables creates a set of compatible CG observables that can be used to simultaneously reproduce experimental observables under a range of conditions imposed at the CG level.

As a result, all experimental observable relationships are present between CG observables that satisfy Equation 5.2 since Equation 5.2 establishes a way to identify an indefinite number of compatible CG observables. It should be noted that Equation 5.2 is greatly simplified if the target observable for both the FG and CG levels depends only on the CG coordinates $\mathbf{R}^N$. One such example is if one is interested in the structure of a large biomolecular system and then the carbon-alpha atoms, for instance, might be a good enough choice to understand that structure of both the FG and CG models. However, when one wishes to calculate thermodynamic and many other structural properties, it is very rarely the case that AFG and ACG depend on the same set of variables.

The relationships in Equation 5.2 also have direct implications for bottom-up CG models. In bottom-up coarse-graining, the FG model is usually parameterized to correspond with experiment (Figure 5.1a) or from "first principles" quantum calculations (which are presumed to also agree with experiment, even if this is rarely the case due to inaccuracy in the "first principles" quantum method). Then, the CG model is constructed using the mapping operator, which establishes a strict correspondence between FG and CG model configurations. This fully specifies the CG model given configurational expressions (e.g., radial distribution functions, RDFs). So, additional observable expressions incompatible with these configu-

rational observables cannot properly correspond with experiment (Figure 5.1b). However, Equation 5.2 provides a way to identify observables compatible with these configurational observables using this strict correspondence between FG and CG models. The problem is that it is very difficult to know an explicit form of $A(\mathbf{R}^N)$ in Equation 5.2 beyond its formal expression appearing in that equation.

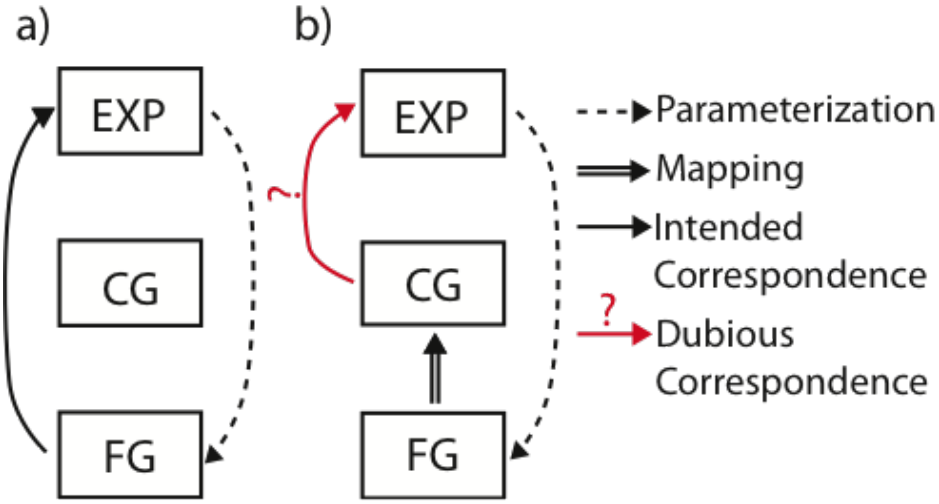

Figure 5.1: The relationships between experiment (EXP), fine-grained (FG), and coarse-grained (CG) models in bottom-up CG models: a) relationship between experiment and FG models, and b) the intended relationship between CG models and experiment. The dashed lines show parameterization. The solid lines show intended correspondences between the models, while the red line with a question mark indicates a dubious correspondence. The double line indicates the strict correspondence from FG configurations to CG configurations through the mapping operator.

On the other hand, top-down CG models do not have a strict correspondence with a specific FG model because they are parameterized using experimental data or "bulk" FG simulations directly. This loose correspondence between FG and top-down CG models means one might decide that one also has the flexibility to implement a choice of CG expressions for physical observables with a similarly loose connection to a FG model, but even then one still needs a compatible set of CG observables in order for the CG model to fully correspond with experiment (see Figure 5.2a), Unfortunately, the loose correspondence between FG and top-down CG models also means that the mapping needed to evaluate Equation 5.2 is not explicitly defined (Figure 5.2b). As a result, top-down models such as MARTINI[99, 217,

218, 216] and mW,[230] parameterized from top-down and interpreted using incompatible observables, do not obviously correspond to an underlying atomistic model for any real physical system. That is, they are purely models at the CG level defined to represent certain aspects of reality. However, the representation of observable calculations must then also be viewed a part of that CG model but having no real connection to an expression for the observable in the underlying FG system.

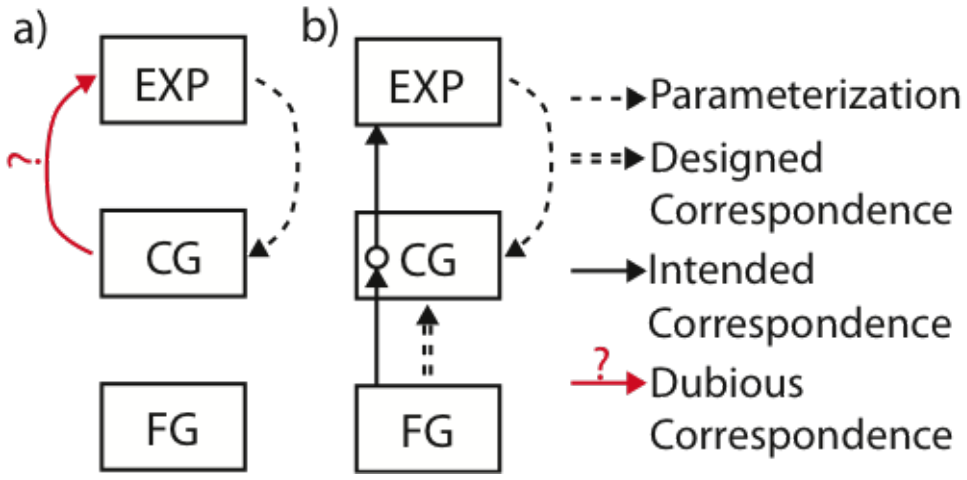

Figure 5.2: The relationships between experiment (EXP), fine-grained (FG), and coarse-grained (CG) models in top-down CG models: a) the relationship between a top-down CG model and experiment and b) the expected relationship between a FG model, a top-down CG model, and experiment. Dashed lines show parameterization. Solid lines show intended correspondences between the models, while the red line with a question mark indicates a dubious correspondence. The double dashed line indicates an intuitive, designed correspondence from the FG model to the CG model.

The present work discusses CG observable representation as an issue of model interpretation based on the correspondence between FG models, CG models, and experiment. The strict correspondence in systematic, bottom-up coarse-graining between FG and CG models creates a compatible set of CG observables given by Equation 5.2. In some cases, these may resemble FG observables; however, these expressions often still differ by the necessary introduction of a state-dependent CG potential term in the CG observable.[309, 197, 165, 106, 16, 315, 82, 8] On the other hand, there may be very little resemblance in general. For instance, Equation 5.2 can define useful CG observables for what may appear to be purely

FG quantities based on configuration variables integrated out of the CG model. Using results from analytical systems, we discuss the validity of previously used observable expressions, introduce new ones, and discuss implications for top-down CG models.

The remainder of this article is structured as follows: Section 2 discusses the correspondence in systematic coarse-graining between FG models, CG models, and experiment in depth and shows how the CG versions of FG observables often change by (and necessitate) the inclusion of new terms; Section 3 uses analytical ideal polymer chains to demonstrate the various issues with using incorrect but seemingly intuitive CG observables in top-down and bottom-up CG models. The final section provides conclusions.

5.2 Theory

CG observables that satisfy Equation 5.2 reproduce both experimental and FG model observables. These CG observables are the average of all FG observable values from FG configurations consistent with a given CG configuration. Thus the FG observable distribution projected onto the CG model is the CG observable distribution, and consequently a CG observable may have a very different functional form than its corresponding FG observable. FG observables are not the same as GG observables, just as FG force fields differ from CG force fields.[246, 247, 335] In order to maximize the benefits of Equation 5.2, one should consider that a CG observable's functional form may not be the same as the form of the atomistic observable expressions.

The similarity of FG and CG observable expressions is determined by the complexity of the projection of the FG observable onto the CG model. A naive CG observable expression treats the CG configuration variables as if they were FG configuration variables, as mentioned earlier (although this is rarely the case). However, the CG observable as a function of CG configuration variables must represent contributions from the eliminated configuration variables as well as the naive contributions from these CG configuration variables. The failure to represent the observable contributions from eliminated configuration variables pre-

vents naive CG observables from reproducing the corresponding FG observable distribution. Therefore, using naive CG observable expressions can prevent researchers from accessing the full predictive power of CG models.

There are three different possible relationships between FG observables and CG observables satisfying Equation 5.2. At one extreme, the FG and CG observable expressions may be the same. For example, the center of mass (COM) radial distribution function (RDF) from a COM CG model is directly identifiable: The COM is one of the CG configuration variables. In special cases like this, naive CG observables faithfully represent the entire projection of the FG observable. At the other extreme, some CG observables expressions have no clear similarity to the corresponding FG expression at all. For example, the magnitude of the molecular dipole of a CG model with a single site per molecule can appear undefined: All of the FG configuration variables traditionally associated with the FG observable have been eliminated. However, a bottom-up expression for the magnitude of the molecular dipole satisfying Equation 5.2 will capture the average behavior of these eliminated configuration variables. By construction it reproduces the projection of the magnitude of the molecular dipole even though this observable appears to be a purely FG quantity. In between these two extremes, there is a broad category of CG observables whose expressions resemble the FG observable expression in part, but not entirely.

For example, all FG configuration variables contribute to the FG pressure. Therefore, the contributions of the eliminated configuration variables must also enter into the CG observable somehow.[95, 106] For pressure, neglecting contributions from eliminated configuration variables treats the CG model as if it were a system of indivisible particles (i.e., a fully atomistically resolved system). While a naive expression for CG pressure could be used in the construction of a CG model, this choice is not compatible with the structural observables implied in the construction of bottom-up CG models. Determining if an observable representation is compatible with the system definition requires one to be aware of the model's resolution in order to understand what it actually represents. However, CG observables sat-

isfying Equation 5.2 correspond to the appropriate resolution and reproduce experimental observables simultaneously with the structural observables related to the system definition. In contrast, naive CG observables defined based solely on analogy with expressions for FG observables neglect essential physics of the CG correspondence by describing a system of possibly an even lower effective resolution than the CG model. Such observables are incompatible with the system definition and interfere with meaningful model interpretation, as its connections to an actual physical system are unclear. For example, if one uses a top-down CG model in a simulation using a CG MD software package, and then simply plugs it in to the well known virial expression for the pressure, but that virial expression is used in terms of only the CG variables as if they were FG variables, then when the simulation outputs "1 atm" as its system pressure this value has no clear relationship to the actual pressure of the real FG system that one is attempting to emulate with the top-down CG model. Indeed, that "pressure" is a part of the CG model as well, but this is purely a part of the overal model because the connection to the FG system such as in Equation 5.2 has not been followed. (The actual FG pressure could be 1000 atm, 1 atm, or 0.1 atm.) The connections between FG and CG models afforded by the statistical mechanics inherent in Equation 5.2 has been lost.

One can, however, in some instances write CG observable expressions as the sum of their FG counterparts plus additional terms that capture contributions from eliminated variables. In the case of thermodynamic observables, this extra part is usually directly related to the transferability of a CG model, and therefore to the thermodynamic state-dependence of the CG interactions a dependence which is often ignored but is clearly a fact based on any reasonable statistical mechanical formulation of the effective potential for a CG system. Though this approach of defining state-dependent CG potentials may not be as fundamental as Equation 5.2, it has strong precedent in the literature on CG thermodynamics,[309, 197, 165, 106, 16, 315, 82, 8] and when used correctly it is can often be shown to be equivalent to Equation 5.2. An especially revealing case is that of our previously

mentioned example, i.e., pressure.[197, 16] The pressure, P, of a CG configuration if taken straight from the FG virial expression would be

$$P_{naive}(\mathbf{R}^N) = \frac{\rho}{\beta} + \frac{\mathbf{R}^N}{3V} \cdot \frac{dU(\mathbf{R}^N)}{d\mathbf{R}^N} \tag{5.3}$$

where $\rho = N/V$ is the density and V is the system volume. In order to actually derive this expression from thermodynamics rather than simply using it in an *ad hoc* fashion, one must make the assumption that the effective CG interactions are volume-independent, so that

$$P_{naive}(\mathbf{R}^N) = -\left.\frac{dG(\mathbf{R}^N)}{dV}\right|_{\mathbf{X}^N} = \frac{\rho}{\beta} + \frac{\mathbf{R}^N}{3V} \cdot \left.\frac{-dU(\mathbf{R}^N)}{d\mathbf{R}^N}\right|_V \tag{5.4}$$

where $G(\mathbf{R}^N)$ is the coarse-grained configurational free energy defined as $G(\mathbf{R}^N) = U(\mathbf{R}^N) - Nk_BT\ln(V)$ and $\mathbf{X}^N \equiv V^{-1/3}\mathbf{R}^N$. This is the naive CG observable expression defined via direct analogy to the FG observable expression. On the other hand, one could derive an expression valid even if the effective CG interaction were volume-dependent. Then, the expression for pressure from thermodynamics has an extra term from the volume derivative, which acts on both the configuration variables and the volume-dependence:

$$P(\mathbf{R}^N) = -\left.\frac{dG(\mathbf{R}^N)}{dV}\right|_{\mathbf{X}^N} = \frac{\rho}{\beta} + \frac{\mathbf{R}^N}{3V} \cdot \left.\frac{-dU(\mathbf{R}^N;V)}{d\mathbf{R}^N}\right|_V - \left.\frac{dU(\mathbf{R}^N;V)}{dV}\right|_{\mathbf{R}^N} \tag{5.5}$$

where $dU(\mathbf{R}^N;V)/dV\Big|_{\mathbf{R}^N}$ may include constant terms such as $(n-N)k_BT/V$ due to the ideal gas contributions of eliminated degrees of freedom even if the interactions are otherwise perfectly transferable. This expression for CG pressure with a volume-dependent effective CG interaction corresponds exactly to expressions in the literature expressed in terms of density derivatives rather than volume derivatives.[197] However, it is important to note

that it satisfies Equation 5.2, and not by accident:

$$P(\mathbf{R}^N) = \langle p(\mathbf{r}^n)\rangle_{\mathbf{R}^N} = \langle \frac{n}{\beta V} + \frac{\mathbf{r}^n}{3V} \cdot \frac{du(\mathbf{r}^n)}{d\mathbf{r}^n}\rangle_{\mathbf{R}^N} \tag{5.6}$$

$$= \frac{\rho}{\beta} + \frac{\mathbf{R}^N}{3V} \cdot \frac{-dU(\mathbf{R}^N;V)}{d\mathbf{R}^N}\bigg|_V - \frac{dU(\mathbf{R}^N;V)}{dV}\bigg|_{\mathbf{R}^N} \tag{5.7}$$

This expression makes it clear that fitting virials is not the same as fitting pressures for each CG configuration. The virial is not the same as the pressure in a CG model in the general case.

Now, consider the configurational internal energy, E, starting from the viewpoint of Equation 5.2. The configurational internal energy, $E(\mathbf{R}^N)$, defined as an average over FG configurations of the system but for a fixed CG configuration is given by

$$E(\mathbf{R}^N) = -\frac{d\ln Z_{\mathbf{R}^N}}{d\beta} = \langle u(\mathbf{r}^n)\rangle_{\mathbf{R}^N} \tag{5.8}$$

where $Z_{\mathbf{R}^N}$ is the partition function of a FG system constrained to a particular $\mathbf{R}^N$. Using Equation 5.2, one can show that the true expression for the configurational internal energy in terms of only CG variables is

$$E(\mathbf{R}^N) = \frac{d\beta U(\mathbf{R}^N;\beta)}{d\beta} = U(\mathbf{R}^N;\beta) + \beta\frac{dU(\mathbf{R}^N;\beta)}{d\beta} \tag{5.9}$$

This is the result one would obtain if the effective CG interaction is temperature-dependent and one gives that fact the respect it requires. Conveniently, it is the sum of a temperature-dependence term and the naive CG configurational internal energy $U(\mathbf{R}^N;\beta)$.

The naive expression applied to the FG configuration variables instead of the CG configuration variables gives the correct FG configurational internal energy. However, it does not behave the same when applied to CG models.

Likewise, the configurational entropy, S, of a set of CG configurational states, Ω, is, according to Equation 5.2 the Gibbs entropy for the corresponding states of the FG model,

ω, averaged over the eliminated configuration variables:

$$S(\Omega) = \int_{\mathbf{r}^n \in \omega} d\mathbf{r}^n \delta(\mathbf{R}^N - M(\mathbf{r}^n)) p_\omega(\mathbf{r}^n) \ln p_\omega(\mathbf{r}^n). \tag{5.10}$$

where p_ω is the probability density of FG configurations among the configurations in ω. One can show that a general CG expression for the configurational entropy integrand in terms of only CG variables that satisfies Equation 5.2 is

$$S_\Omega(\mathbf{R}^N) = -k_B p_\Omega(\mathbf{R}^N) \ln p_\Omega(\mathbf{R}^N) + \frac{dU(\mathbf{R}^N)}{dT} \tag{5.11}$$

Again, this is expressed as the integral of the naive CG expression

$$S_\Omega(\mathbf{R}^N) = -k_B p_\Omega(\mathbf{R}^N) \ln p_\Omega(\mathbf{R}^N) \tag{5.12}$$

and a temperature-dependence term. The temperature dependence in Equation 5.11 is the complement of the temperature-dependence seen for configurational internal energy.

In some cases Equation 5.2 will correspond to naive CG observable expressions; however, this is not generally the case. For pressure, all configuration variables contribute to the FG observable expression, as discussed above, and so observables satisfying Equation 5.2 must represent the contributions from all eliminated configuration variables through additional terms. For configurational internal energy, for instance, the naive CG configurational internal energy that neglects the β derivative term of Equation 5.9 does not indicate what portion is truly internal energy; entropic effects from eliminated configuration variables are also included in the effective interaction.

In order to separate these effects, contributions from the eliminated configuration variables must be included using an additional term as in Equation 5.9. For configurational entropy, the issues are similar to those for configurational internal energy. Since entropy is a measure of the distribution of states on phase space, properties of the distribution beyond

the average are needed to accurately describe this observable. Consequently, all eliminated FG configuration variables make contributions to CG entropy that are captured using an additional term such as in Equation 5.11. This is to be expected as configuration-dependent thermodynamics are projected onto the many-body potential of mean force (PMF) as a result of coarse-graining.[248, 246, 247, 204, 296, 69, 285]

This has implications for the interpretation of top-down CG models as well. In top-down coarse-graining, analogy between CG particles and groups of particles in an FG model establishes sets of approximate structural observables, as depicted in Figure 5.2b. Any additional CG observables such as energies and pressures must be compatible with these structural observables if they are to reproduce both sets of experimental observables simultaneously (to say nothing of crucial structure-pressure and structure-energy cross correlation observables). In order to be compatible with the system definition, we showed that bottom-up CG observables satisfying Equation 5.2 will include contributions from eliminated FG configurational variables. Including these contributions leads to the introduction of extra terms at the CG level that did not appear in the FG observable representation. Thus, top-down CG observable expressions built by analogy to FG observables, ones that do not contain such extra terms—which is very often the case—will generally misrepresent the underlying FG or atomistic physics for the system at hand.

Top-down CG models are fundamentally inconsistent in their relationship to the real systems they are designed to model if they do not also take into account these facts about observable representability, i.e., thermodynamic state-dependence of CG potentials and the other possible complications discussed above.

5.3 Examples and Discussion

In this section, a simple model is used to demonstrate different aspects of CG observable representation. The analytical ideal polymer highlights how coarse-graining even a single intramolecular bond can lead to problems with naive CG observable expressions that can

then be corrected using expressions that satisfy Equation 5.2. The discrepancies between naive CG observables and the corresponding FG observables defy intuition and make it clear that the interpretation of naive CG observables can be highly suspect. These problems do not go away when the number of segments is increased or even when the rigid nature of the freely jointed chain (FJC) model is relaxed. Additionally, CG observables satisfying Equation 5.2 include contributions from segments that are eliminated when constructing an end-to-end representation of the polymer. The representationally consistent CG observables presented in the previous section are compared with FG observables and naive CG observables as appropriate for these models.

The ideal polymer chain has been a cornerstone of polymer physics because it is a simple model that displays the entropic character typical of polymers without the added complexity of non-bonded interactions between monomers.[103, 283] The FJC is an ideal polymer model where the bond vectors between segments $\vec{r}_i$ are fixed to be length b, but rotation is unimpeded. In this model, all configurations that simultaneously satisfy the bond length constraints are allowed and have zero interaction energy, while all other configurations are forbidden. As a result, it is obvious that the configurational internal energy is always zero and that any configurational free energy differences must be controlled by configurational entropy. A common reduced representation for this model is the magnitude of the end-to-end distance $R = |\sum_i \vec{r}_i|$, where all configuration variables other than the position of the polymer endpoints are integrated out. For an N segment chain on a d-dimensional lattice with steps of $\pm b/\sqrt{d}$ allowed in each dimension for each segment, the distribution of the end-to-end distance is given by

$$P^d_{FJC}(R;N) = (\frac{d}{2\pi N b^2})^{d/2} \exp\left(\frac{-R^2 d}{2Nb^2}\right) \tag{5.13}$$

The properties of this model are shown in the left column of Tables 5.1-5.3. The effective "CG" potential $U(R) = -\beta^{-1} \ln P^d_{FJC}(R)$ for this model is not independent of β, differing

from the state-independent FG model. When the seemingly intuitive, naive identification of the CG force field with the CG internal energy is used to interpret the CG model (Table 5.1), the resulting internal energy is non-zero, which disagrees with intuition. Adding contributions from eliminated configuration variables to that internal energy by satisfying Equation 5.2 gives Equation 5.9, which agrees with intuition. Likewise, excess entropy measured using the naive expression gives zero, which violates intuition for this fundamentally entropic model. However, the true CG entropy expression in Equation 5.11 satisfies Equation 5.2 and properly corresponds to the entropy of the FG model. The partitioning of configurational internal energy and entropy by the naive observables incorrectly allocates all of the configurational free energy to internal energy instead of entropy. However, the free energies corresponding to the two sets of observables must be equal since they both describe the same CG model. As noted in the previous section, the fact that the entropic and energetic contributions to the configurational free energy cannot be correctly distinguished by the naive CG observables is common. This is why an awareness of the model resolution is vital to determining the CG observable expressions that will correspond with experimental observables.

Table 5.1: Properties of freely jointed chain (FJC) models for the configurational internal energy (E) in the FG model, as well as the CG model, using both expressions for a naive CG observable defined by direct analogy of the AA observable and the representationally consistent observable that satisfies Equation 5.2. The CG models are end-to-end representations of the polymer chain. $Z_b \equiv \sum_{l=-N}^{N} \exp(-\frac{l^2}{2Nb^2} - \frac{\beta k}{2Nb^2}(R-lb)^2)$.

Model/Observable	$\langle u(\mathbf{r}^n)\rangle_R$	$E_{naive}(R) = U(R)$
Normal FJC	0	$-\frac{d}{2\beta}\ln\left(\frac{d}{2\pi Nb^2}\right) + \frac{R^2 d}{2\beta Nb^2}$
2-segment Off-lattice	0	$-\beta^{-1}\ln\left(\frac{3R}{4b^2}\sqrt{1-\frac{R^2}{4b^2}}\right)$
Blurred FJC	$\frac{k}{2Nb^2}\langle(R-lb)^2\rangle - \frac{1}{2\beta}$	$-\frac{1}{2\beta}\ln\left(\frac{\beta k}{2\pi Nb^2}\right) - \beta^{-1}\ln Z_b$

The failure of naive CG observables to reproduce FG observables and the state-dependence acquired in the naive effective CG interaction are a fundamental feature of coarse-graining. To show that this is not just unique to the lattice model examined above, a 2-segment

Table 5.2: Properties of freely jointed chain (FJC) models for the entropy (S) of CG configurations in the FG model, using both expressions for a naive CG observable defined by direct analogy of the AA observable and the representationally consistent observable that satisfies Equation 5.2. The CG models are end-to-end representations of the polymer chain. $Z_b \equiv \sum_{l=-N}^{N} \exp(-\frac{l^2}{2Nb^2} - \frac{\beta k}{2Nb^2}(R-lb)^2)$.

Model/Observable	$S(\omega(R))$	$S_{naive}(R)$
Normal FJC	$\frac{k_B d}{2}\ln(\frac{d}{2\pi N b^2}) - \frac{k_B R^2 d}{2Nb^2}$	0
2-segment Off-lattice	$k_B \ln\left(\frac{3R}{4b^2}\sqrt{1-\frac{R^2}{4b^2}}\right)$	0
Blurred FJC	$-\frac{k_B}{2}\ln(\frac{\beta k}{2\pi N b^2} + k_B \ln Z_b - \frac{k_B}{2} + \frac{k_B \beta k}{2Nb^2}\langle(R-lb)^2\rangle$	0

off-lattice (OL) FJC model will be investigated in 3-D. The distribution of the end-to-end distance is

$$P_{FJC}^{OL} = \frac{3R}{4b^2}\sqrt{1-\frac{R^2}{4b^2}} \tag{5.14}$$

The naive effective CG potential has again acquired β-dependence that the FG model did not have. The properties of this model are shown in the middle column of Tables 5.1-5.3. This model also has zero configurational internal energy. The naive configurational internal energy expression does not correctly reproduce this FG value. However, the expression satisfying Equation 5.2 correctly reproduces the FG observable. The configurational entropy shows the same behavior as above. These patterns hold even for more complicated models with non-uniform FG energies and continuous end-to-end distance, which will be demonstrated below using an elastic FJC. In this model, there is a harmonic potential $u(r_i) = k(r_i - b)^2/2$ on the length of each segment i centered at distance b with spring constant k instead of the delta function in the FJC with fixed-length bonds. The distribution of the end-to-end distance of a 1-dimensional blurred FJC is given by the proportionality

$$P_{Blur}^{1}(R;N) \sim \sqrt{\frac{\beta k}{2\pi N b^2}} \sum_{l=-N}^{N} \exp\left(-\frac{l^2}{2Nb^2} - \frac{\beta k}{2Nb^2}(R-lb)^2\right) \tag{5.15}$$

where l is a summing index needed to consider the contribution to the probability from

each Gaussian centered at an end-to-end distance of lb. The second term in the exponential reflects the Gaussian chain aspect of this model, which allows all Gaussians centered at the FJC distance lb to have non-zero contributions to the probability of any given end-to-end distance. This model reduces back to the FJC model described in Equation 5.13 in the limit that k goes to infinity. The naive effective CG interaction for this model still has β-dependence that the FG model did not have, which comes from the FJC-like left term in the exponential. Thus, the thermodynamics are still not properly reproduced when the naive observable expression is used for internal energy and entropy. However, the CG observables satisfying Equation 5.2 have no problem reproducing the FG observable projected onto this CG representation. So, one should expect this behavior from any model that has some FJC-like character. The harmonic potential centered at distance b is representative of harmonic potentials used for bonded interactions in general, which suggests that this problem arises whenever bonded configuration variables are coarse-grained.

Until this point, the results have focused on thermodynamic observables. However, the pattern of creating CG observables that correspond to the FG model's observables is not just one of finding corrections to thermodynamic observables. Instead, it applies to any observable. Table 5.3 shows the magnitude of the average bond orientation $O = N^{-1} \sum_i \hat{r}_i$ for the different polymer models, where $\hat{r}_i$ is the unit vector for bond segment i. Given only the end-to-end distribution of the CG model, it would seem that the bond orientation is undefined in the CG model since configuration variables used in the FG observable were eliminated to reach the CG level of resolution. Using a naive orientation measure on the one "CG segment" gives a value of unity since the "CG segment" is always aligned with itself. For the normal FJC and the 2-segment off-lattice FJC, there is only one value for the magnitude of the average orientation for a given end-to-end distance, which is represented correctly using an expression satisfying Equation 5.2. In the case of the blurred FJC, there is a distribution of magnitudes for the average orientation for a given end-to-end distance. The projection onto the CG level of resolution, however, averages this distribution for each

end-to-end distance. In both cases, statements can be made about the magnitude of the average orientation. One could still improve the performance of the naive CG observable by changing the number of sites to be the implied number of FG particles instead of using the explicitly represented CG sites. This "resolution-aware" version of the naive CG observable expression reproduces the correct value of the observable by including information about the number of segments and the segment length. However, this observable is a special case like the COM RDF example discussed above. This works here because the observable itself involves an aggregation step that directly follows the elimination of configuration variables in mapping from the FG model to the CG model. Although the corrected naive observable expression satisfied Equation 5.2 here, there are cases where a multiplicative correction to the naive observable expression is insufficient to satisfy Equation 5.2.

The polymer radius of gyration, R_g^2, is an example of how the correct CG observable expression differs from the naive CG observable expression in more complicated ways. For the FG model, the radius of gyration is given by $R_g^2 = N^{-1}\sum_i(\mathbf{r}_i - \mathbf{r}_{cm})^2$, where $\mathbf{r}_i$ is the position of bead i and $\mathbf{r}_{cm}$ is the COM. Using the naive CG expression for the radius of gyration on the end-to-end CG representation of the 2-segment off-lattice FJC gives

$$R_{g,naive}^2 = R^2/4 \tag{5.16}$$

since one only knows about the effective end-to-end segment of the CG polymer. However, one can analytically evaluate the expectation of the radius of gyration conditional on the end-to-end distance. This is the projection of the radius of gyration from the FG model onto the CG model and satisfies Equation 5.2. In this case,

$$R_{g,true}^2(R) = \langle R_g^2(\mathbf{r}^n)\rangle_R = \frac{1}{9}(2b^2 + R^2) \tag{5.17}$$

As described by Equation 5.17, the radius of gyration behaves differently than Equation 5.16 would suggest. While both descriptions for $R_g^2(R)$ scale quadratically with end-to-end

distance, Equation 5.16 has the wrong prefactor. Using the same sort of multiplicative scaling as above would change the denominator from 4 to 6, but an updated version of Equation 5.16 would still not have the same prefactor as Equation 5.17. Also notably, Equation 5.17 shows that there is a non-zero $R_g^2(R)$ at an end-to-end distance of zero. This agrees with physical intuition for any FJC, but it is not captured by Equation 5.16. Using simple multiplicative corrections to the naive CG observable does not fix this violation of physical intuition. Here, one can only recover the correct behavior for $R_g^2(R)$ by satisfying Equation 5.2.

Table 5.3: Properties of freely jointed chain (FJC) models for the magnitude of the average orientation measured in the FG model, as well as the CG model, using both a naive CG observable $O_{naive}(R)$ defined by direct analogy of the FG observable and a resolution-aware observable ($O_{true}(R)$) satisfying Equation 5.2. The CG models are end-to-end representations of the polymer chain.

Observable/Model	Normal FJC	2-segment Off-lattice	Blurred FJC
$\langle O(\mathbf{r}^n)\rangle_R$	$\frac{R}{Nb}$	$\frac{R}{2b}$	$\frac{R}{Nb}$
$O_{naive}(R)$	1	1	1
$O_{true}(R)$	$\frac{R}{Nb}$	$\frac{R}{2b}$	$\frac{R}{Nb}$

Discussion

The example systems in the last section demonstrate problems that can arise when interpreting CG models using naive CG observable expression. In both systems, naive CG observables incorrectly attribute the entropic part of the configurational free energy to configurational internal energy. For the FJC, this misattribution completely changes the interpretation of the model from being driven by entropy, as intuition suggests, to being driven by internal energy. The fact that this problem does not go away for the blurred FJC suggests that it is a general phenomenon affecting any CG model where intramolecular configuration variables are eliminated. In these systems, the expressions satisfying Equation 5.2 were derivable because the CG interaction potential could be written analytically; however, one does not have this luxury when studying complex systems. Instead, one could numerically determine this additional term using a method such as the single-point CG sensitivity formula.[338] Alter-

natively, one could fit the entire expression numerically since CG observables that satisfy Equation 5.2 are guaranteed to reproduce the projection of corresponding FG observable.

More generally, these results emphasize the importance of resolution awareness in model interpretation. For the magnitude of the average orientation, neglecting the resolution of the CG model by using the naive observable expression for the model leads to a nonsensical answer, given the FG model that it is supposed to represent. The end-to-end representation could be constructed for a polymer of any length and segment-type, but the correct interpretation of the CG model depends on what FG model it represents. While this observable was simple enough that only minimal consideration of the FG model was needed to satisfy Equation 5.2 and reproduce the FG value, this will not work in general. However, CG observables that satisfy Equation 5.2 will always work regardless of the observable complexity. For example, the added complexity of the radius of gyration means that simple corrections to the naive CG observable fail to satisfy Equation 5.2; however, the resolution-aware CG observable that satisfies Equation 5.2 correctly describes the scaling and asymptotes. Satisfying Equation 5.2 also ensures that CG observables faithfully reproduce the complete projection of FG observables.

This observable projection approach avoids the observable incompatibility seen when using naive CG observables. Of note, pressure has been a particularly problematic observable in the literature.33, 35, 64 Depending on what observables are used to parameterize the CG model, the compatibility of CG observable expressions with different model definitions may change. As a result, Equation 5.2 alone cannot always resolve the disputes over which pressure expression is correct for the potential in any given model. Many different systems can map to a given CG representation; the Debye-Huckel model is a particularly notable example in colloid physics.[82, 70, 327, 331, 264] Instead, we assert that the change in resolution from the FG model to the CG model will determine CG observable expression compatible with the model definition. Generally, naive CG observables not part of the model definition will be incompatible with the model definition. However, CG observables satisfying Equation

5.2 will always be compatible with the model definition by construction. Additionally, these CG observables will adjust to be compatible with different system definitions.

The observables used to parameterize top-down CG models also need to be compatible with each other. Otherwise, the CG model will not reproduce these observables; consequently this model will not correspond to the intended FG system. In fact, many CG models may not correspond to any physically realizable system. Without a way to determine if top-down CG observables are compatible, one may not know that their CG model is unphysical.

One most natural way to avoid this complexity in top-down coarse-graining is to look for maximally transferable models, since the least state-dependent potentials imply the smallest deviation from atomistic observable forms. This is no doubt part of what makes transferability such an important consideration when building top-down models. Though naively it might seem that non-transferable models are powerful at least at the state point they describe, in fact the principle embodied in Equation 5.2 implies that non-transferable models cannot be interpreted according to the usual atomistic-model-based observable expressions. This no doubt accounts for some of the strong prejudices in the field against non-transferable models.

An inability to determine if top-down CG observables are compatible using Equation 5.2 means that other approaches are needed. One could apply bottom-up CG observable expressions to top-down CG models. In this case, one needs to be aware of the different model resolutions even if the distribution of CG configurations is in both top-down and bottom-up CG models. Here, the bottom-up CG observable could be the basis for parameterizing a top-down CG model. Alternatively, one could try to correct naive CG observables to be compatible with top-down CG model definitions. However, the radius of gyration example showed that there are limits to what can be done using only simple corrections to naive CG observables.

Any comparison of top-down and bottom-up CG model properties is nonsensical unless care is taken in choosing observables. Indeed, a top-down model fit using incompatible

observables may reproduce a given observable using a naive observable expression. However, that does not mean that it corresponds to the intended experimental system.[132] This problem is directly encountered when top-down CG model definitions are adjusted to reproduce additional observables such as pressure and interfacial tension without considering the compatibility of all model observables.[242]

Improvements in CG pressure can benefit constant NPT CG simulations. The integrated approach developed by Das and Andersen[95] and extended by Dunn and Noid's[106] iterative refinement procedure correctly describes volume fluctuations as the observable contributions from eliminated configuration variables are projected onto the system volume. Correct pressure fluctuations would also be captured if these contributions were projected onto the CG configuration variables instead of the system volume. However, most constant NPT CG simulations currently barostat using the CG virial instead of using an approach like the one developed by Das and Andersen.[95] From our earlier discussion, it is clear that the CG virial is not compatible with any bottom-up model that is coarser than the atomistic model. Currently, the behavior of constant NPT CG models is different from the behavior of supposedly underlying FG models. Constructing barostats using an expression for CG pressure that satisfies Equation 5.2 might help align the behavior of CG models under constant NPT to experiment.

Finally, mixed resolution modeling can also benefit from improved CG observables.[157, 269] Since observable expressions change with model resolution, the expression for the pressure of the FG part of the system should be different from the expression for the pressure of the CG part of the system. Failure to recognize this causes unphysical density profiles at the interface between resolutions in adaptive resolution simulations that requires the introduction of thermodynamic forces to counteract this apparently pressure-induced drift.[343] This problem is even worse for adaptive resolution schemes with top-down CG models for the reasons already discussed.[365, 364] Using resolution appropriate observable expressions is an important first step towards improving these models.

5.4 Conclusions

In this article, we have discussed the importance of CG observables being compatible with the resolution of the CG model. Using Equation 5.2 an indefinite number of compatible CG observables can be identified. If defined correctly, these observables may also be consistent with top-down CG model definitions. However, naive expressions for CG observables taken directly from the FG expressions often fail to satisfy Equation 5.2 because they neglect contributions from eliminated CG configuration variables. These neglected contributions are sometimes captured by terms that depend on a system's thermodynamic state dependence, which can be used to correct naive CG observable expressions; this was demonstrated analytically for simple analytically solvable polymer systems. One must treat CG models differently than FG models by considering the resolution of the CG model in its interpretation. These concepts cannot be ignored in either bottom-up or top-down CG modeling if multiscale modeling is to generate physically meaningful predictions for physical systems which have their origins at a FG level.

Exploring the aspects of CG observable representation discussed here can impact research efforts in several areas. Using observable expressions that satisfy Equation 5.2 can give new hope to work on basis set representability.[197, 165] Additionally, CG simulations in the constant NPT ensemble can more closely follow the evolution of the FG system using expressions for CG pressure that correspond to the CG model's resolution by using approaches such as the method developed by Das and Andersen[95] or Dunn and Noid's[106] extension. Further improvements to better describe pressure fluctuations are also possible. Likewise, mixed-resolution models will need to use a variety of properly formulated observable expressions to describe the properties of the different resolutions present in the simulation and their interfaces so that large thermodynamic artifacts are not generated. Perhaps most importantly, top-down CG models must be parameterized using compatible observable expressions that are consistent with the model's underlying FG resolution. These top-down models can benefit from bottom-up analysis to determine these observable expressions. In the end, the

interpretation of CG models must ultimately depend on understanding how they relate to the actual, physical FG experimental systems they intend to describe.

CHAPTER 6

COARSE-GRAINING FEW COMPLEX MOLECULES: METADYNAMICS

Metadynamics is an adaptive biasing enhanced sampling method that forces a system to explore all of a configuration space evenly in order to reveal its probability of occupying configurations it would not otherwise visit frequently enough to lead to a good occupation probability estimate.

Transposing it to the social sphere, we might imagine trying to torment the soul of some bachelor by sending an impish spirit down to inspire him to the Faustian passion never to be satisfied with any situation enough to ask it to stay. Our bachelor will quickly grow dissatisfied with any situation that continues too long and will quickly explore every accessible corner of the earth. His explorations and encounters will reveal to us his character, as Faust's do.

6.1 Purpose and Precedents

Although in principle observing nature carefully over infinite time scales could be sufficient to reveal all physics, it is more practical to design experiments that investigate specific questions by intentionally varying physical parameters in a controlled manner. Similarly in computational modeling of complex molecules, when direct simulation of natural processes by molecular dynamics[171, 7, 295] is infeasible, specially designed simulations can nonetheless reveal key physics in far less simulation time.[175, 78, 3]

The adaptive enhanced sampling method metadynamics[181, 182, 27] is one such approach, specifically designed to accelerate the determination of potentials of mean force (PMFs) by promoting transitions between long-lived metastable states. Metadynamics has been widely applied across chemistry from materials science to biochemistry, but it is still quite young theoretically in some ways, with a rigorous proof of convergence appearing only

one year ago by the author of this thesis,[89] to be discussed in the following chapter.

Metadynamics[181, 28, 27, 313] accelerates computer simulations of molecular dynamics[171] along a set of several otherwise slow collective variables (CVs) characterized by the presence of multiple metastable states, such as solid lattice order parameters,[221, 251, 312] molecular dihedral angles,[181, 159] or protein-ligand binding coordinates,[133, 259, 193] to enable the calculation of free energy surfaces which would be prohibitively expensive to calculate using unbiased simulations. It accelerates sampling by iteratively constructing and applying a bias potential to offset free energy barriers along a small set of chosen collective variables (CVs);[227] the iterative construction rule is to add smooth hills to the bias around each successively sampled point to reduce re-exploration of already-sampled CV-space and thereby enhance escape from local minima.[137, 149, 181] It is one of a diverse set of adaptive biasing methods for efficient calculation of free energy surfaces (see e.g., [227, 42, 149, 29, 92, 219, 93, 369, 101, 302]) that also includes the widely applied Wang-Landau scheme.[341, 2, 323, 33, 113] Though conceived for the study of soft matter, like the Wang-Landau scheme its strategy applies to the efficient sampling of any multimodal distribution in which the most pernicious multimodality can be characterized in terms of just a few variables.[192, 18, 76, 43]

It is one of a long progression of enhanced sampling methods designed to calculate PMFs of several collective variables using the least foreknowledge of the system necessary, and it is worth placing it in context by tracing the history of its antecedents before going further. One of the earliest and best known enhanced sampling methods is umbrella sampling,[321, 172] in which one or more bias potentials are added to an otherwise normal molecular dynamics simulation to encourage sampling of high-free-energy configurations that otherwise act as kinetic barriers. By ameliorating these barriers, umbrella sampling makes calculation of averages of quantities that require high quality sampling across the barrier more efficient. Two basic classes of umbrella sampling are typical. In the first, a single potential is added to flatten out the free energy surface around and across the barrier. To engineer such a

potential requires knowing the approximate location and free energy surface of the barrier region. In another, simulations are run under many different so-called window potentials to ensure that sampling of each so-called window—including the ones around the barrier—is of roughly even quality. To engineer these requires knowing the approximate free energy gradients around the barrier. Unfortunately, the free energies, locations, and shapes of the barriers are often exactly what one wants to discover.

When very little information about the barrier is known in advance an alternative is to bootstrap an umbrella potential by starting with unbiased molecular dynamics and iteratively biasing the dynamics as more becomes known.[227, 29] Metadynamics[181, 182, 27] is an archetypal method of this type meant to take only a few reaction coordinates and information about their short-time dynamic fluctuations as inputs. Similar alternatives include Adaptive Biasing Force[92, 93] and Wang-Landau[341] methods, and its most immediate precursor was the Local Elevation Method.[149] In metadynamics, as in the Local Elevation Method, the simulation begins unbiased and the system is sequentially biased away from the neighborhood of the points it has already visited in reaction coordinate space. This bias builds cumulatively until all reaction coordinate points are equally likely to be sampled—at least in concept. In reality, one must taper the incremental bias additions to zero to achieve true convergence, which was the inspiration for well-tempered metadynamics (WTMetaD).[28]

Unfortunately, setting the tapering rate for the WTMetaD hills requires guessing the free energy barrier height in advance, information that metadynamics was initially designed to do without. Furthermore, even though WTMetaD was argued to converge in principle, in practice one often sees instability in WTMetaD simulations that causes nonsensical results and other times one sees false convergence that looks reasonable but cannot be reproduced in independent simulations. The problems were severe enough that there was widespread doubt that WTMetaD actually did converge exactly; the strongest rigorous convergence results assumed unrealistic levels of dynamical time scale separation.

In the following four chapters I first prove that WTMetaD and a family of related methods

do converge exactly as intended, then, using insight from the proof, address the problems of tuning the tapering rate, ensuring simulation stability, and preventing false convergence. Before I do this, however, it is worth presenting the mathematical formulation of metadynamics and making some notes on its generality.

6.2 Definition

Since it was first introduced, many newer forms of metadynamics have been developed with various strengths and weaknesses. Metadynamics is performed with multiple walkers, with replica exchange, and with combinations of the two. Gaussian hills are used in fixed-width[181] and adaptive-width[49] forms, though sometimes in a field-coordinate space,[326] and non-Gaussian hills have been found necessary in some cases[23, 224, 87] to avoid boundary artifacts. The hill sizes can be left constant,[181] iteratively decreased according to the WTMetaD rule,[28] iteratively decreased according to a variety of new global tempering rules,[89, 88, 353] or some combination of the latter two.

In every case investigated theoretically in the course of this thesis, the method has had a common general form that this section will now introduce. Consider an instance of metadynamics applied to a system sampling states x from a state space X and applied through CVs taking values s in a compact collective variable space S that are observed using some map $M : X \to S$, sometimes also called $\sigma : X \to S$. As in previous sections, let $\mathcal{M}$ be the marginalization map from measures $\pi(x)$ on X to measures on S, $\mathcal{M}(\pi(x)) = \int \delta(s - M(x))d\pi(x)$. The time-dependent bias on these collective variables will be denoted by $V(s, t)$.

Then an instance of metadynamics has two ingredients, a reference hill deposition operator that is independent of the bias and a rule for decreasing the size of the hills based on the current bias. The hill deposition operator is a linear operator denoted $\hat{G}$, $\hat{G} \circ p = \int_S ds' G(s, s')p(s')$; linearity simply means that the hills for two samples should be the sum of the hills for each sample. $\hat{G}$ should also be positive semidefinite and bounded, i.e., it

should never cause greater sampling of a particular sampling pattern but should not totally prevent re-sampling of any pattern, for stability. Further detail on these conditions will be given in the next chapter.

A typical choice for $G(s, s')$ is a Gaussian function with some fixed width. The rule for the size of the hill, on the other hand, should be a positive, bounded, "tempering" functional $\omega(V(s,t))e^{-\bar{V}/\Delta T_{gt}}$ with $\Delta T_{gt} \in (0, \infty)$ times a well-tempering factor $e^{-V(s,t)/\Delta T_{wt}}$ with $\Delta T_{wt} \in (0, \infty)$. With these specified, the metadynamics update rule can be written as

$$V(s, t_{n+1}) = V(s, t_n) + \omega(\tilde{V}(s, t_n))e^{-\bar{V}(t_n)/\Delta T_{gt}} G(s, s_{n+1}) e^{-V(s_{n+1}, t_n)/\Delta T_{wt}} \tag{6.1}$$

in terms of samples x_n taken at evenly spaced times $t_n = n\Delta t$ obtained from simulation using the most recent bias on each time interval and s_n the corresponding CV state samples, $M(x_n) = s_n$. All of the analysis and method development described in the immediately following chapters will reference this form.

The remarkable generality of this form is rarely entirely clear on first glance. It may appear to be limited to single-walker metadynamics. However, it is in fact quite general once the proper perspective is taken, encompassing methods from the initial metadynamics to multiple walkers metadynamics to bias exchange metadynamics, to field-coordinate metadynamics. The remainder of this brief chapter will discuss how it encompasses bias-exchange metadynamics as an example.

Consider the simplest case of bias-exchange metadynamics,[263] the one used in practice.[23, 22, 134] In the usual formulation, one begins with a set of N replica systems i with states x_i from a state space X_i. Each state can be mapped to one of N different CVs; the individual CVs s_j (each with a corresponding CV space S_j) are found given the individual CV maps $M_j : X \to S_j$ from the x_i by choosing a state-to-CV assignment $A : \{i\} \to \{j\}$. One sets $s_{A(i)} = M_{A(i)}(x_i)$. The system-CV correspondence is one to one, so each A can be considered an invertible permutation. Given this assignment, each replica i is subject to an

adaptive bias $V_{A(i)}(M_{A(i)}(x_i))$, which I will also write as $V_j(M_j(x_{A^{-1}(j)}))$ when it is more natural to keep track of a particular CV than a particular replica.

Each individual replica will typically have time-independent Markovian dynamics given a fixed bias and bias-assignment, i.e. the probability of transitioning from a state x_i to state x_i' in a time t is given by some distribution $\pi_{V_{A(i)}(M_{A(i)}(x_i))}(x_i, x_i', t)$. These must have stationary distributions

$$\lim_{t\to\infty} \pi_{V_{A(i)}(M_{A(i)}(x_i))}(x_i, x_i', t) \equiv p_{V_{A(i)}}(x') \sim p_0(x') \exp(-\beta V_{A(i)}(M_{A(i)}.(x_i'))). \tag{6.2}$$

Because the systems are replicas, $p_0(x)$ does not depend on A. However, these single-subsystem dynamics do not specify the entire system dynamics because the biases are exchanged, i.e. A is a function of time. In fact, one typically has a single overall Markovian, time-translation-invariant distribution for the probability of transitioning from replicas-and-permutation-state $\{\{x_i\}, A\}$ to replicas-and-permutation-state $\{\{x_i'\}, A'\}$ in time t; call this $\pi_{\{V_j\}}((\{\{x_i\}, A\}, \{\{x_i'\}, A'\}, t)$. In practice, the exchanges are performed according to a rule that preserves the stationary distributions for each individual subsystem dynamics and does not weight any permutations differently, so that the stationary distribution of that transition rule exists and can be written

$$\lim_{t\to\infty} \pi_{\{V_j\}}((\{\{x_i\}, A\}, \{\{x_i'\}, A'\}, t) \equiv p_{\{V_j\}}(\{\{x_i\}, A\}) \sim \prod_i^N \exp(-\beta V_{A(i)}(M_{A(i)}(x_i)))p_0(x_i). \tag{6.3}$$

There is one term per permutation and each term corresponds to roughly the same idea: a set of replicas sampling with a certain choice of 1D biases. A second natural stationary distribution to calculate is the distribution averaged over fine-grained variables and permu-

tations,

$$P_{\{V_j\}}(\{s_j\}) \equiv \sum_A \int dx^n \delta(M_{A(i)}(x_i) - s_{A(i)}) p_{\{V_j\}}(\{\{x_i\}, A\}) \tag{6.4}$$

$$\sim \prod_j^N P_{0,j}(s_j) \exp(-\beta V_j(s_j)), \tag{6.5}$$

where the $P_{0,j}$ are the N 1D CV marginal distribution functions in the unbiased system. It is a product distribution, corresponding to a separable free energy.

Finally, each individual bias V_j is updated with some typical 1D metadynamics rule, say

$$V_j(s, t_{n+1}) - V_j(s, t_n) = h_j e^{-(s - M_j(x_{A^{-1}(j,t_{n+t})}(t_{n+1})))^2/2\delta s_j^2} e^{-V_j(M_j(x_{A^{-1}(j,t_{n+t})}(t_{n+1})),t_n)/\Delta T} \tag{6.6}$$

or more simply and generally

$$V_j(s, t_{n+1}) - V_j(s, t_n) = G_j(s, M_j(x_{A^{-1}(j,t_{n+1})}(t_{n+1}))) e^{-V_j(M_j(x_{A^{-1}(j,t_{n+t})}(t_{n+1})),t_n)/\Delta T}. \tag{6.7}$$

This is a single update for all the biases that can also be written

$$V_j(s, t_{n+1}) - V_j(s, t_n) = \int ds' G_j(s, s') e^{-V_j(s', t_n)/\Delta T} \delta(s' - M_j(x_{A^{-1}(j,t_{n+1})}(t_{n+1}))) \tag{6.8}$$

The only apparent difference from the single-CV case is that here, instead of having a single bias one has a set of biases. The set of biases could however be identified with a single bias by choosing that single bias to be such that the assignment between replicas and CVs is the identity permutation, $V(\{x_i\}) = \sum_i^N V_i(M_i(x_i))$, or by choosing it to be a function on a product of the label set and the N-CV product space, i.e., $V(\{\{x_i\}, A\}) =$

$\sum_i^N V_{A(i)}(M_{A(i)}(x_i))$. In the latter case, this is an update with a well-defined $G(s, s')$,

$$G(\{\{x_i\}, A\}, \{\{x'_i\}, A'\}) = \sum_j^N G_j(M_j(x_{A^{-1}(j)}), M_j(x'_{A'^{-1}(j)})) \tag{6.9}$$

With typical hills on finite domains, it looks like N right-angle-intersecting hyperplanes of dimension $N-1$ dividing an N-dimensional box into orthants given a particular permutation, with different permutations corresponding to different orderings of axes in that space, so that the overall object is like a stack of boxes cut into orthants that have been flipped around in a variety of ways. Therefore all of the work on convergence and method improvement that will follow applies just as well to bias-exchange metadynamics (and multiple walker metadynamics, and more) as to the original metadynamics—so long as one does not forget the strange look of the hill function.

However metadynamics acts as an adaptive umbrella sampling in 1D, then, bias-exchange does the same job of doing adaptive replica exchange umbrella sampling. The replica exchange is performed between several umbrella potentials defined along different dimensions that all have the same target end states as they would if there were no exchange. Just as the replica exchange can be formulated as a single dynamics in a product space plus a permutation space, bias exchange metadynamics can be formulated as a single metadynamics in a product space plus a permutation space.

Keeping this in mind, it should quickly become apparent how general the work in the next chapters is, a generality that has never been fully conveyed in the individual papers each chapter represents.

CHAPTER 7

PROOF THAT TEMPERED METADYNAMICS CONVERGES ASYMPTOTICALLY ALMOST SURELY

This chapter substantially reproduces the material from the paper Dama, J. F., Parrinello, M., & Voth, G. A. (2014). Well-Tempered Metadynamics Converges Asymptotically. Physical Review Letters, 112(24), 240602 in its first section. The latter sections consist of previously unpublished notes.

This section provides proof that the method converges asymptotically to any intended sampling profile, as determined by the hill function, but the convention is to choose that to be entirely flat.

We now imagine our little imp of restlessness tries to not just push the unfortunate bachelor around endlessly, but to also leave him wandering forever with no preferences at all left at the end of time, the ultimate dissatisfied hipster, disgust with monotony having exactly balanced his natural preferences. How does he manage it? There are many technicalities to the proof, so I'll just outline a few.

First, he whispers slightly less in the bachelor's ear every day, just enough that the bachelor can ignore random variations in the imp's whispers without forming new preferences, but just enough that the effect of the whispers never stops building to an infinite dissatisfaction with any respite, sure to overcome any natural preferences. (Sum of the step sizes condition.)

Second, he whispers against the bachelor's exact position and the neighborhood, but he is careful to ensure that the dissatisfaction with the neighborhood leaves no room for the bachelor to seek refuge at very ends of the earth, thinking that beyond the walls some hope can be found because the imp cannot whisper past the walls. (Boundary conditions on the hills.) In other words, the bachelor must be led to believe that whatever he experiences that looks closest to what he cannot experience will not lead him to any escape: the edges of

experience should seem just as dull as its heart.

Third, the bachelor must not be so strong-willed and faithful to his natural inclinations that he cannot be moved even when dissatisfied, and nor can he be so excitable and desperate that any dissatisfaction sends him to the ends of the earth no matter how small. (Adequate relaxation speed and not too large dynamic response to the bias.)

7.1 Motivation, Context, and Sketch

Though metadynamics is widely used, open questions about its convergence properties have cast doubt on the accuracy of its results and made innovations appear difficult to justify in the minds of some and have impeded its adoption in other fields.[183, 57, 219, 228, 190, 85, 325, 101, 100, 302, 76, 43, 224] The most powerful previously extant convergence results, due to Bussi et al.[57] and Dickson,[100] relied on idealizing assumptions that metadynamics updates continuously in time and that the CV dynamics is effectively Brownian. Bussi et al. showed that metadynamics as originally formulated does not lead to a single predictable bias under those assumptions—instead, the bias fluctuates indefinitely. Dickson showed that an analytically convenient approximation of the variant well-tempered metadynamics (WTMetaD) does converge under those assumptions, but for this approximation the bias converges to a mollified version of the desired end state—leaving roughness in the final sampling distribution. From another branch of physics, work on Wang-Landau-like methods[34, 32, 126] indirectly suggests that WTMetaD using discrete CVs, hills that bias one CV state at a time, and finite time between updates could converge as expected under fairly general CV dynamics—but again, this speaks little to practical metadynamics.

As yet there has been no rigorous demonstration that metadynamics, as actually implemented in practice,[47, 122, 324] with finite-width hills deposited at finite time intervals according to a sampling dynamics with imperfect separation of time scales, is a stochastic process with a single end state that is always reached given enough time—or, if it is, that the end state reached is the intended one, as Dickson most clearly suggested might not be

the case.[101, 100] This chapter provides such a demonstration and the essential technical constraints that arise along the way.

The proof does not rely on assumptions of effective Brownian dynamics, continuous time updating, or idealizations of the hill function. The demonstration treats the method as-implemented by applying reliable stochastic mathematics.[179] The remaining body of this section describes the physical argument of the proof unburdened by technical mathematics; the following sections describe the technical details. The argument requires understanding a key auxiliary time scale, how the evolution of the bias in that time scale approaches an ordinary differential equation (ODE), the stable state for that ODE, and why the ODE converges to that stable state. It is a precise argument under fairly general conditions that are specified later in this section and does not contradict practical experience: though formal convergence is robust, the actual time to convergence may remain sensitive to system details, especially the presence of slow variables orthogonal to the CVs under study.[369, 27]

Metadynamics iteratively constructs an umbrella potential as a sum of hill functions around successively sampled states, a process referred to as "filling the free energy surface (FES) with computational sand."[181] These hills are smooth functions of just a few CVs and come in a variety of shapes. As a guiding example consider the typical case of Gaussian hills on a 1D periodic domain as illustrated in Figure 7.1. In well-tempered metadynamics (WTMetaD), hill shape remains constant but the hill height decreases as the bias at the sampled point becomes larger. The bias, which increases monotonically, eventually changes very slowly with time.[28] The precise rule for WTMetaD is

$$V_{n+1}(s) = V_n(s) + G(s, s_{n+1})e^{-V_n(s_{n+1})/\Delta T}, \tag{7.1}$$

letting s denote a CV state point, $G(s, s')$ an energy-valued hill function of s to be deposited when point s' is sampled, ΔT an energy-valued positive scalar parameter, s_n a sequence of CV state samples from the biased trajectory at times t_n, and $V_n(s)$ the sequence of biases at

times t_n. s may be a vector,[181] a discrete index,[85] a function,[326] or any combined set of these. The t_n are spaced evenly, and the underlying system's dynamics must be Markovian and ergodic for any fixed bias applied during the simulation. A Gaussian hill in 1D on a periodic interval corresponds to $G(s, s') = h \exp(-(s - s')^2/\delta^2)$ where h is the hill height, δ is the hill width, and the $s - s'$ subtraction obeys the minimum image convention.

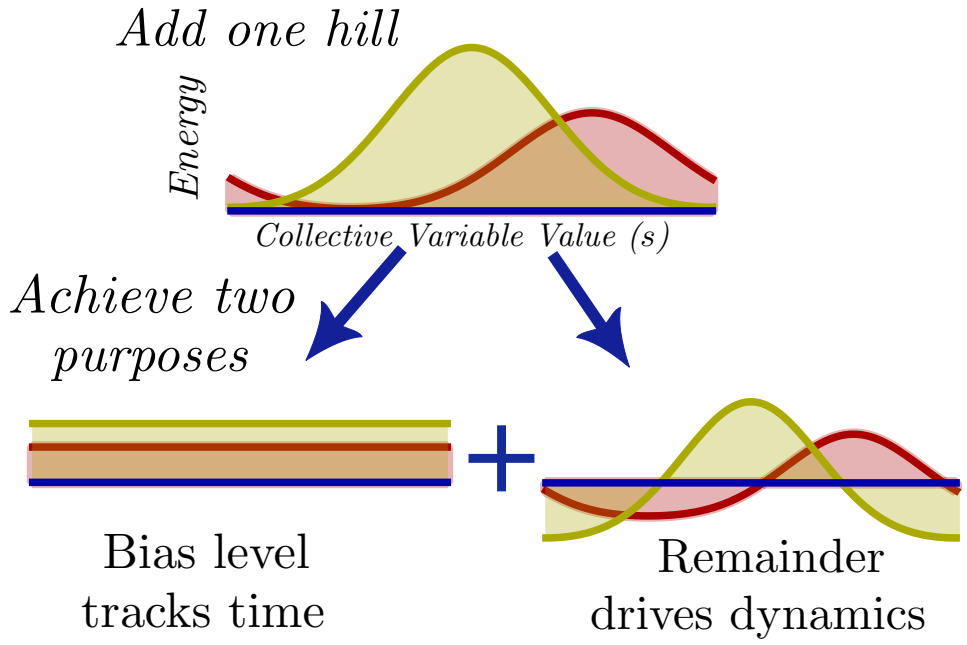

Figure 7.1: Two Gaussian hills on a 1D periodic domain (top) split into hill level components (lower left) and driving bias components (lower right). The hill level exerts no forces and serves only to mark time.

To see how metadynamics converges begin by noting that the bias has two parts with essentially different functions, illustrated in Figure 7.1. The first is the average level of the bias, $\bar{V}_n = \int V_n(s)ds / \int ds$. This level does not affect the system dynamics. It serves only to record the total amount of bias, or "computational sand," deposited so far, and it never stops increasing. It appears as nothing more than an intriguing way to measure the elapsed time in a simulation, as if the computational sand were emerging from an hourglass. The second is the driving bias, $\tilde{V}_n(s) = V_n(s) - \bar{V}_n$. This portion accounts for all effects of the bias on dynamics and we assume that its values (and derivatives if relevant) stay within a finite range by nature—in practice, metadynamics does not give rise to infinite energy differences

or forces except in pathological cases.[183, 57, 85, 224] In technical terms, we assume the driving biases remain in a bounded compact subset of bias-space.

To track the evolutions of these parts separately to reveal more about their functions, split the update into two parts,

$$\bar{V}_{n+1} = \bar{V}_n + \exp(-\bar{V}_n/\Delta T)\lambda(s_{n+1}, \tilde{V}_n) \tag{7.2}$$

$$\tilde{V}_{n+1}(s) = \tilde{V}_n(s) + \exp(-\bar{V}_n/\Delta T)\Gamma(s, s_{n+1}, \tilde{V}_n), \tag{7.3}$$

where the hill level $\lambda(s', \tilde{V}) = \int G(s, s')e^{-\tilde{V}(s')/\Delta T}ds / \int ds$ and a driving hill function $\Gamma(s, s', \tilde{V}) = G(s, s')e^{-\tilde{V}(s')/\Delta T} - \lambda(s', \tilde{V})$ account for how adding hills affects the average level of the bias and the driving bias, respectively. Written in this manner it is apparent that $\exp(-\bar{V}_n/\Delta T)$ functions as an ever-decreasing step size, and the sums of $\exp(-\bar{V}_n/\Delta T)$ makes a natural auxiliary time scale. Because $\tilde{V}_n$ is bounded by assumption and Gaussian hills are positive and bounded, $\lambda(s_{n+1}, \tilde{V}_n)$ must be bounded from above and below by some M_u and M_l,

$$M_l \leq \lambda(s_{n+1}, \tilde{V}_n) = (\bar{V}_{n+1} - \bar{V}_n)\exp(\bar{V}_n/\Delta T) \leq M_u. \tag{7.4}$$

Therefore, for the same reasons that $\dot{x}e^x \sim 1$ implies $x \sim \ln t$ and $e^{-x} \sim 1/t$,

$$b_l/n \leq \exp(-\bar{V}_n/\Delta T) \leq b_u/n \tag{7.5}$$

for all large n for some positive b_l and b_u. An inifinite sum of such terms always diverges, which implies there exists a time scale τ that is one-to-one with the original simulation time scale t and satisfies

$$\tau(t_n) \equiv \sum_{i=0}^{n-1} \exp(-\bar{V}_i/\Delta T). \tag{7.6}$$

Examining the updates in this time scale will reveal a simple structure to asymptotics of the overall method.

At early times, $\exp(-\bar{V}/\Delta T)$ is large, so few hills must be added to accumulate a given

increment of τ-time. As time goes on, $\exp(-\bar{V}/\Delta T)$ becomes smaller, so more hills must be added to accumulate the same increment. This is illustrated in Figure 7.2. Because each hill is proportional to $\exp(-\bar{V}/\Delta T)$, the bias changes overall by an amount roughly proportional to the length of the increment. As time goes on, even the shortest increments $d\tau$ will see the addition of many hills—corresponding to sampling for long t-times—yet because they are short the change in the bias will be small. Because the change in bias over each of these short increments is small, the update over each is approximately as if all of the sampling were done on the starting bias surface; because the sampling goes for a long time and many hills are added, the update over each is approximately an integral over hill functions chosen from the biased equilibrium distribution $p_b(s) \sim \exp(-(F(s) + V(s))/T)$. Together, the observations in the previous sentence imply that the cumulative update rule in the limit of large τ and small $d\tau$ is well-approximated by

$$\tilde{V}(s, \tau + d\tau) = \tilde{V}(s, \tau) + d\tau \int \Gamma(s, s', \tilde{V}(\tau)) p_b(s') ds', \tag{7.7}$$

suggesting the asymptotics follow the long-time differential equation

$$d\tilde{V}(s, \tau)/d\tau = \int \Gamma(s, s', \tilde{V}(\tau)) p_b(s') ds'. \tag{7.8}$$

A related equation can be found in the appendix of Branduardi et al.[49] Though this argument may seem heuristic, it is precise given conditions defined later in this text.[179]

Next, given this asymptotic ODE one must characterize its asymptotic behavior in turn. One viable strategy is to determine the fixed points of the equation and then characterize the stability of those fixed points. The fixed points are those biases for which $\int \Gamma(s, s', \tilde{V}) p_b(s') ds' = 0$, $\int G(s, s') \exp(-\tilde{V}(s')/\Delta T) p_b(s') ds' = C$ for some constant C, a very appealing result because it corresponds to the heuristic used to motivate metadynamics when it was first introduced: assuming perfect sampling.[181] For the 1D Gaussians being

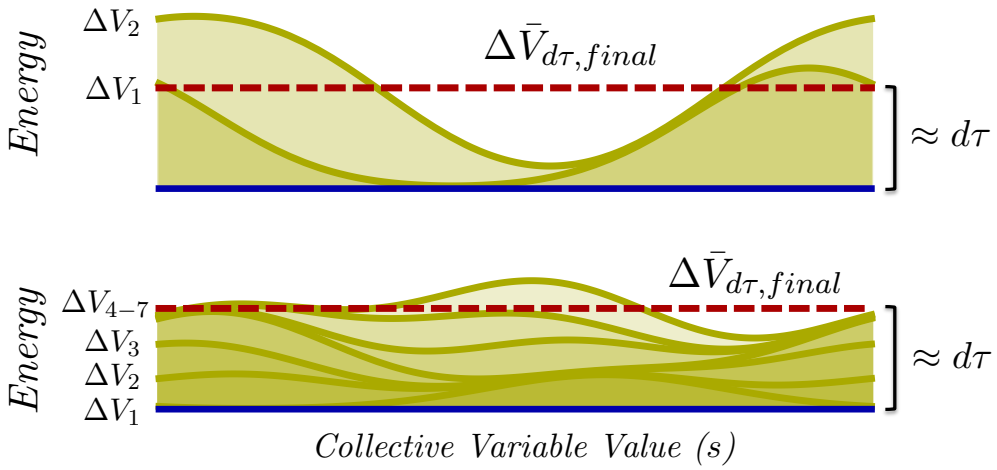

Figure 7.2: Typical cumulative bias updates from a flat sampling distribution over short intervals of equal τ-durations ($d\tau$) at early (top) and late (bottom) times. Yellow lines denote the bias after each successive update during the interval, with the top line in each figure the final updated bias. Red dashed lines denote the change in average bias level. As time goes on, the total bias deposited in each interval of fixed τ-duration stays roughly constant while the number of hills and the quality of sampling both increase.

considered, this amounts to

$$e^{-\tilde{V}(s)/\Delta T} e^{-(F(s)+\tilde{V}(s))/T} \sim C, \tag{7.9}$$

or

$$\tilde{V}(s) = -(1 + T/\Delta T)^{-1} F(s) + C' \tag{7.10}$$

for a different constant C'—exactly the end state originally suggested for WTMetaD.[28]

Already, the fact that there is only one unique fixed point in this case is a good sign. In fact, the fixed point is both unique and globally attracting, i.e., the ODE will converge to the desired end state whatever its physically reasonable initial state, as can readily be verified with an appropriate choice of Lyapunov function. One choice is the Kullback-Leibler divergence (relative entropy) of the current tempering-reweighted state, $p_w(s,\tau) \sim \exp(-\tilde{V}(s,\tau)/\Delta T) \exp(-(F(s)+\tilde{V}(s,\tau))/T)$, from the target sampling end state, $p_w(s,\tau) \sim$

1 named $p_c(s)$, given by

$$D(p_c||p_w) = \int p_c(s) \ln \frac{p_w(s,\tau)}{p_c(s)} ds. \tag{7.11}$$

After a series of algebraic manipulations, the τ-time deriviative is given by

$$dD(p_c||p_w)/d\tau = -r(\tau) \int \int (p_w(s,\tau) - p_c(s)) G(s,s') (p_w(s',\tau) - p_c(s')) ds ds' \tag{7.12}$$

where $r(\tau)$ is a positive parameter involving T, ΔT, and normalization factors for $p_w(s,\tau)$ (algebra in supporting material). When $G(s,s') = he^{-(s-s')^2/\delta^2}$, this derivative is always less than zero unless $p_w(s) = p_c(s)$, in which case it is exactly zero. Thus the relative entropy is a Lyapunov function for the ODE and the driving bias satisfying Equation 7.10 is a global attractor. The method converges asymptotically.

With that guiding example followed to its conclusion, now consider adapting the argument for more general cases. The most practically important generalization is across hill functions, e.g., geometrically adapted hills,[49] hills designed for metadynamics on finite domains,[85, 23, 224] and hills interpolated over discrete grids.[47] Considering the update Equation 7.1 once more but leaving the hill function unspecified, one may follow the same intuitions given design constraints on $G(s,s')$.

First, to find the time scale τ well-behaved, the quantity $\int G(s,s')ds / \int ds$ must exist and be positive for any s': hills must add bias, enough to eventually fill the region, so the region must be finite in size. Also, ΔT must be finite and positive. Second, for the ODE to have fixed points, the equation

$$\int G(s,s') p_c(s') ds' = C \tag{7.13}$$

must have at least one positive, normalized, physically possible solution $p_c(s)$ for some constant C: there must be a possible sampling distribution from which the bias update

will not change that distribution. Finally, for relative entropies to be Lyapunov functions, $G(s, s')$ must be positive semidefinite: the hills must be deposited so that they encourage escape from local traps, never tightening of traps.

Given all of these conditions on $G(s, s')$, the reasoning above can be extended to show that the instantaneous tempering-reweighted distribution for any such method asymptotically approaches the set of normalized solutions of Equation 7.13, possibly converging to one or possibly fluctuating in the set indefinitely: the Lyapunov function is the infimum of the relative entropies above over all physically realizable choices of p_c. Using smooth, boundary-consistent Gaussian hills satisfies all these conditions and guarantees a unique solution. The nonuniqueness and possible fluctuation are confined to the details that the hill function cannot resolve: the null space of $\int G(s, s')ds'$. For instance, if the hills are chosen to be step functions on certain regions of CV-space, the detail within each region will not be flattened but all coarser aspects of the FES will be—as in Wang-Landau sampling.[341]

The analysis above also suggests more general tempering schemes. A new type of metadynamics, introduced here and inspired by density of states methods[341, 33, 228] and self-healing umbrella sampling[219], is "globally tempered metadynamics," (GTMetaD), where hill size is independent of location and adaptively set based on global properties of the bias; the update rule here is

$$V_{n+1}(s) = V_n(s) + w[V_n]G(s, s_{n+1}), \tag{7.14}$$

where $w[V]$ is a unitless positive scalar functional of the bias and the other variables are as before. If $w[V]$ can be factored into bias-level and driving-bias components as $\omega[\tilde{V}]\exp(\bar{V}/\Delta T)$, then all of the previous convergence reasoning for WTMetaD also holds for GTMetaD. Simple GTMetaD schemes of this form appear to offer substantial advantages over WTMetaD[33, 228] and furthermore one could use both temperings together in hybrid schemes.

The part of the argument above requiring the most careful technical analysis is the derivation of the long-time ODE Equation 7.8. Precise, permissive sufficient conditions for this intuition to hold constitute an important result for the wider field of adaptive biasing

enhanced sampling methodology. It can be proven using the method of Theorem 6.6.1 from Kushner and Yin[179] given two additional conditions on the underlying system dynamics. Roughly put, this theorem proves that a stochastic iteration of the form $\theta_{n+1} = \theta_n + \epsilon_n Y_n(\theta_n, x_n)$, where θ_n are the desired iterates, ϵ_n are step sizes, and $Y_n(\theta_n, x_n)$ are updates that depends on the current iterate θ_n and the randomly sampled state x_n of a Markov model whose transitions may depend on θ_n, converges like solutions of the differential equation $d\theta/d\tau = \langle Y(\theta, x)\rangle$, where $\tau = \sum \epsilon_n$ and the expectation is taken over the equilibrium distribution of x given θ, with probability one. It holds if the sum of the ϵ_n diverges while the sum of ϵ_n^2 converges, the Y_n are bounded, and a few other technical conditions hold. In our case, θ is the driving bias and the theorem applies to Equation 7.3 with s_{n+1} a projection of a fine-grained system state x_n sampled from a fixed-bias-Markovian dynamics.

The first technical condition not already established matches the claim that one can approximate cumulative bias updates over fixed τ-increments as adding infinite hills chosen from the biased equilibrium distribution. A precise requirement is: however many hills are added, the expected difference between updating from the actual sampling distribution and from the quasi-equilibrium sampling distribution should always be at least as small as a single update step, i.e.,

$$\sum_{m=n+1}^{\infty} e^{-\bar{V}_m/\Delta T}(E_{ne,\tilde{V}_n,x_n} - E_{e,\tilde{V}_n})\Gamma(s, s_m, \tilde{V}_n) = O(e^{-\bar{V}_n/\Delta T}), \tag{7.15}$$

for every s, x_n visited, and V_n visited. Here $E_{ne,\tilde{V}_n,x_n}$ is the expectation over system states given non-equilibrium dynamics under bias $\tilde{V}_n$ given that the (fully-specified) system state is x_n at t_n, and $E_{e,\tilde{V}_n}$ is the equilibrium expectation under bias $\tilde{V}_n$. Since $\Gamma(s, s_m, \tilde{V}_n)$ is independent of t and the ϵ_n decay as $1/n$, if the system is Markovian and time-invariant this reduces to

$$\sum_{m=n+1}^{\infty} \frac{1}{m}(p_{V_n}(s, x_n, t_m - t_n) - p_b(s, V_n)) = O(1/n). \tag{7.16}$$

Here $p_V(s, x, t)$ is the probability of going from a system state x to any system state that

maps to CV state s in time t with the dynamics subject to fixed bias V, while $p_b(s, V)$ is the equilibrium CV distribution under fixed bias V. Note that it is each fixed-bias dynamics, not the CV dynamics or adaptively biased dynamics, which is Markovian here—a standard case for molecular dynamics and Monte Carlo simulations. This sum measures how quickly a system approaches equilibrium, so the condition is a more precise formulation of the intuition that metadynamics works only when the system equilibrates reasonably quickly. If nonequilibrium fluctuations in the system decay exponentially or streched-exponentially, the usual case for non-driven ergodic systems away from criticality, then the differences in the sum decay properly and the condition is satisfied.

The second matches the claim that one can approximate the cumulative update over a τ-time increment as if all of the sampling were done on the starting bias surface. This can only be true if the sum of many hills does not change too rapidly with small changes in the bias. In more precise terms, the sum in the previous paragraph must have derivatives with respect to V_n bounded by the step size,

$$\sum_{m=n}^{\infty} \frac{1}{m} \frac{\delta(p_{V_n}(s, x_n, t_m - t_n) - p_b(s, V_n))}{\delta V_n} = O(1/n) \tag{7.17}$$

for all s, x_n visited, and V_n visited. A sufficient condition for this is that $p_V(s, x, t)$ should have bounded functional derivatives with respect to V that relax to their long-time value with any algebraic dependence faster than $1/t$. This amounts to a precise formulation of the intuition that metadynamics only works if the system does not jam or become unstable in response to the bias, and it is also a permissive condition for non-driven ergodic systems away from a phase transition. It will typically be satisfied whenever the system is finite and the short-time generator of dynamics is a smooth function of the bias, for instance when the bias is implemented simply as an added force in Langevin molecular dynamics[181] or as an added energy in Metropolis Monte Carlo.[85]

These two conditions, together with the stated design constraints on the hill function and

the assumption that the driving bias stays properly bounded throughout the simulation, are enough to allow for a detailed proof that WTMetaD as actually implemented converges asymptotically and converges to simple, designable end states[179] provided in the next sections. However, it gives no guarantees that it will converge quickly, especially when used with poorly chosen CVs. The end states are exactly those suggested by the simple quasi-equilibrium heuristic that Laio and Parrinello used to motivate metadynamics when it was first introduced,[181] with no mollification artifacts,[100] and Bussi et al.'s gauge free energy concept[57] proves key. This analysis is flexible enough to cover metadynamics with or without adaptive Gaussians[49] or boundary corrections,[85, 23, 224] and with or without an approximating grid.[47]

Furthermore it provides design constraints for such methods and demonstrates a powerful yet little-known tool for rigorously characterizing non-equilibrium, non-Markovian systems. As sampling methods become ever more sophisticated and diverse, ensuring that sufficiently powerful analysis strategies are brought in to complement them becomes ever more important. The next two sections describe this tool in substantial detail with an effort to remain as accessible as possible by providing first, a "physicist's proof" that leaves out certain the most technical details, and second, a full mathematical proof stated as a theorem. The final section of the chapter provides the detailed algebra proving that the Lyapunov function cited in the text is indeed a Lyapunov function.

Most simply, our conclusion has been that metadynamics converges asymptotically to a desired target state given the conditions in the previous section. In particular, any bounded trajectory of physically reasonable biases converges to a target state. Furthermore, our conclusions have allowed us to describe the target states precisely and confirm that the heuristic arguments concerning the target states in the earliest metadynamics papers are well-justified. The target states are those for which $V(s,t) = -(\Delta T_{wt}/(T + \Delta T_{wt}))(F(s) + \ln p_c(s)) + C$ where $p_c(s)$ is any solution of the equation $\int_S ds' G(s,s') p_c(s') = C$. $-\ln p_c(s)$ is Bussi et al.'s "gauge free energy" and is not conceptually new though it has not been widely

emphasized in the literature to date. Note that we have not made any effort to investigate when the problem is or is not well-posed when $p_c(s)$ is not unique; we leave that to future investigation.

However straightforward that conclusion may be, the full power of this proof may not be immediately apparent on first glance. It has strong implications for a wide variety of metadynamics methods. For instance, taking X and S to be product spaces of N identical "single walker" state spaces, the proof applies to multiple-walker metadynamics and (adding a few dummy label variables) to bias-exchange metadynamics directly. Additionally, it allows for the case of discontinous, Wang-Landau-like step function hills; the proof never assumes continuity of $\int_S ds' G(s, s')$ or the uniqueness of the solution to $\int_S ds' G(s, s') p_c(s') = C$. Keeping fine-scale detail in the null space of $\int_S ds' G(s, s')$ makes a great deal of sense and the $p_c(s)$ are quite understandable in the Wang-Landau-like case. It also applies to field-coordinate metadynamics with the collective variables the same as the underlying system variables, though here the $p_c(s)$ are not so easily understandable. It does not apply to driven metadynamics, but the generalization appears conceptually straightforward: simply introduce a distinction between the state sampling sequence and the weighted state sampling density, as in the previous section, then follow the same strategy. This list could go on but it would be unsystematic and wide ranging.

7.2 Physicist's Proof of Convergence to a Differential Equation

Of all the sections of the convergence argument in the previous section, the one most dependent on ideas from Kushner and Yin[179] that are not widespread in the field of enhanced sampling is the transition from a sequence of discrete-time updates in possibly far-from-equilibrium system to an asymptotic differential equation at quasiequilibrium. It may therefore appear heuristic. However, the argument is in fact quite precise despite being written in natural language, and this section will summarize how it works for readers who would like an intermediate level of technical detail that does not require deep reading of Kushner and

Yin.

The argument in the previous section is, quoted in full:

> Consider now how the bias evolves in τ-time. At early times, $\exp(-\bar{V}/\Delta T)$ is large, so few hills must be added to accumulate a given increment of τ-time. As time goes on, $\exp(-\bar{V}/\Delta T)$ becomes smaller, so more hills must be added to accumulate the same increment. This is illustrated in Figure 2. Because the change in the bias with the addition of each hill is proportional to $\exp(-\bar{V}/\Delta T)$, the bias overall can only change by an amount roughly proportional to the length of the increment. As time goes on, even the shortest increments $d\tau$ will see the addition of many hills—corresponding to sampling for long t-times—yet because they are short the change in the bias will be small. Because the change in bias over each of these short increments is small, the update over each is approximately what it would be if all of the sampling were done on the starting bias surface; because the sampling goes for a long time and many hills are added, the update over each is approximately an integral over hill functions chosen from the instantaneous biased equilibrium distribution $p_b(s) \sim \exp(-(F(s) + V(s))/T)$. Together, the observations in the previous sentence imply that the cumulative update rule in the limit of large τ and small $d\tau$ is well-approximated by
>
> $$\tilde{V}(s, \tau + d\tau) = \tilde{V}(s, \tau) + d\tau \int \Gamma(s, s', \tilde{V}(\tau)) p_b(s') ds',$$
>
> suggesting the long-time differential equation
>
> $$d\tilde{V}(s, \tau)/d\tau = \int \Gamma(s, s', \tilde{V}(\tau)) p_b(s') ds'.$$

This argument reproduced from the previous section centers around the idea of "the update in a fixed increment of τ-time." This is a precise quantity. Using the time scale

introduced in the previous section, which satisfies $\tau(t_n) \equiv \sum_{i=0}^{n-1} \exp(-\bar{V}_i/\Delta T)$, there is also a natural reverse-mapping $m(\tau) = n$ such that $\tau_n < \tau \leq \tau_{n+1}$. In terms of this reverse map, "the update of the driving bias in a fixed increment of τ-time" can be written as

$$\tilde{V}(s, \tau + d\tau) = \tilde{V}(\tau, s) + \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \exp(-\bar{V}_{i-1}/\Delta T)\Gamma(s, s_i, \tilde{V}_{i-1}) \tag{7.18}$$

using the letter's variable definitions. The argument that small finite-increment updates at long times can be replaced by the quasiequilibrium expectation update is the argument that

$$\lim_{\tau\to\infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \exp(-\bar{V}_{i-1}/\Delta T)\Gamma(s, s_i, \tilde{V}_{i-1}) - \int_{\tau}^{\tau+d\tau} \int \Gamma(s, s', \tilde{V}(\tau'))p_b(s')ds'd\tau' = 0. \tag{7.19}$$

(Technically this must be convergence under the uniform norm. For details like this see the next section.) The conditions stated in the previous section guarantee that this limit holds, as one can see with a bit of reverse-engineering algebra. To be concise, use $\epsilon_i \equiv \exp(-\bar{V}_{i-1}/\Delta T)$ and $\bar{g}(s, \tilde{V}) = \int \Gamma(s, s', \tilde{V})p_b(s')ds'$ from here on.

First off, the two terms are clearly different in that the one on the left is continuous in time whereas the other jumps, so to reverse engineer the real update from the limit update, start by approximating the integral stepwise using a sum.

$$\int_{\tau}^{\tau+d\tau} \bar{g}(s, \tilde{V}(\tau'))d\tau' = \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i \bar{g}(s, \tilde{V}_{i-1}) + \rho(\tau), \tag{7.20}$$

$$\rho(\tau) = \int_{\tau}^{\tau+d\tau} \bar{g}(s, \tilde{V}(\tau'))d\tau' - \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i \bar{g}(s, \tilde{V}_{i-1}). \tag{7.21}$$

Now, because the $\tau(t_n)$ scale as $\log n$, the "discretization step" of the discrete sum is of order $1/n$, and

$$\lim_{\tau\to\infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i = \lim_{\tau\to\infty} d\tau + O(1/n) = \lim_{\tau\to\infty} d\tau + O(e^{-\tau}), \tag{7.22}$$

so, assuming that $\bar{g}(s, \tilde{V}(\tau'))$ is reasonably well-behaved (the proof that it is and a technical definition of well-behaved are in the more technical part of the supplement), this means

$$\lim_{\tau \to \infty} \rho(\tau) = \lim_{\tau \to \infty} O(e^{-\tau}) = 0. \tag{7.23}$$

Therefore our task will be accomplished if we can show that

$$\lim_{\tau \to \infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i (\Gamma(s, s_i, \tilde{V}_{i-1}) - \bar{g}(s, \tilde{V}_{i-1})) = 0. \tag{7.24}$$

The hardest part of simplifying this is accounting for the complex correlations between changes in the bias and changes in the state sampling sequence. The expectations of the individual differences in this sum depend on the entire bias history and fine-grained state history between τ and $\tau(t_i)$. First, consider the difference between the expectation of a non-quasiequilibrium update with a fixed driving bias $\tilde{V}$ and the expectation of an update from quasiequilbrium sampling,

$$E_n[\Gamma(s, s_{n+1}, \tilde{V}_n)] - E_{p_b(s_{n+1}, t_n)}[\Gamma(s, s_{n+1}, \tilde{V}_n)]. \tag{7.25}$$

Regarding the quasiequilibrium update as our ideal update, this should be regarded as a sampling artifact. The expectations of these and correlations between them are complex and our task can be formulated as showing that all of the artifacts will cancel out asymptotically. That would leave us with effective quasiequilibrium updating.

A good way to approach this is to introduce a term that accounts for all expected future non-quasiequilibrium biasing artifacts that would be seen if the bias were held constant indefinitely,

$$v_n(x_n, \tilde{V}_n) \equiv \sum_{i=n}^{\infty} \epsilon_i E_n[\Gamma(s, s_i^{\tilde{V}_n}, \tilde{V}_n)] - E_{p_b(s_i, t_n)}[\Gamma(s, s_i, \tilde{V}_n)], \tag{7.26}$$

where $s_i^{\tilde{V}_n}$ are samples taken from the system dynamics with the bias held fixed at $\tilde{V}_n$ rather than the actual metadynamics with a changing bias. Then, for each stepwise difference in the sum above, add that step's v to and subtract the previous step's v from the stepwise difference. These terms would cancel telescopically except for the first added term and the last added term, and those two that do not cancel go to zero asymptotically (under the conditions described in the previous section; see next section for the technical reasoning), so this leaves the limit unchanged. However, it simplifies the individual terms of the sum considerably.

The new condition we seek to prove with those artifacts introduced is

$$\lim_{\tau\to\infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i(\Gamma(s, s_i, \tilde{V}_{i-1}) - \bar{g}(s, \tilde{V}_{i-1})) + v_i(x_i, \tilde{V}_i) - v_{i-1}(x_{i-1}, \tilde{V}_{i-1}) = 0. \quad (7.27)$$

The convenient aspect of this is that $v_{i-1}(x_{i-1}, \tilde{V}_{i-1})) = \epsilon_i E_{i-1}[\Gamma(s, s_i, \tilde{V}_{i-1})] - \epsilon_i \bar{g}(s, \tilde{V}_{i-1}) + E_{i-1}[v_i(x_i, \tilde{V}_{i-1})]$; the first two terms help cancel the original part of the stepwise difference and that last expectation can help decouple the expected change in sampling and expected change in bias. Examine the update with this substitution,

$$\lim_{\tau\to\infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i(\Gamma(s, s_i, \tilde{V}_{i-1}) - E_{i-1}[\Gamma(s, s_i, \tilde{V}_{i-1})] + v_i(x_i, \tilde{V}_i) - E_{i-1}[v_i(x_i, \tilde{V}_{i-1})]) = 0. \quad (7.28)$$

The first two terms group nicely; save them for later. Next, use $E_{i-1}[v_i(x_i, \tilde{V}_{i-1})] = E_{i-1}[v_i(x_i, \tilde{V}_{i-1})] - v_i(x_i, \tilde{V}_{i-1}) + v_i(x_i, \tilde{V}_{i-1})$ and group terms to find a total of three con-

venient, physically meaningful correspondences, like so:

$$\sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i(\Gamma(s, s_i, \tilde{V}_{i-1}) - E_{i-1}[\Gamma(s, s_i, \tilde{V}_{i-1})]) + \dots \tag{7.29}$$

$$(v_i(x_i, \tilde{V}_i) - v_i(x_i, \tilde{V}_{i-1})) + \dots$$

$$(v_i(x_i, \tilde{V}_{i-1}) - E_{i-1}[v_i(x_i, \tilde{V}_{i-1})])$$

Now, whereas we originally had one difference between instantaneous updates and quasiequilibrium distributions, a hard thing to reason precisely about, we now have three simpler differences. The first is the difference of the actual instantaneous update from its expectation from the step before, and only involves a taking one actual sample of the short time dynamics using one bias surface. The second is related to the sensitivity of the projected future non-quasiequilibrium artifacts to changes in bias, and involves understanding the differences in expected sampling correlations between just two biases at a time and only on one actual sample, the current one. Finally, the third is the difference of the current projected future non-quasiequilibrium artifacts given the next sampled state and the current bias and our expected value for that projection on the next step at the current bias; that involves only a single actual sample using a single value of the bias. Each of these is much more straightforward to reason about than the previous differences, each of which depended on the samples of both the bias and the fine-grained state over the whole history from τ to $\tau(t_i)$.

For instance, take the first term. Assuming that the driving bias is bounded, this term is too (our hill deposition operator is by construction). Furthermore, at every step its expectation is zero no matter what the previous steps did; the expectation of the difference between a quantity and its expectation is always zero. That means it can be thought of as a white noise with variable but bounded intensity, and therefore the square of its sum without the epsilons would grow linearly in time. However, the squares of the epsilon terms multiplying the sum decay quadratically, so the variances of these partial sums go to zero

overall, and therefore we can guarantee that this random sum effectively always converges to zero in the limit. (The probability of it being any finite distance from zero becomes zero asymptotically.)

The third term is similar. By the linearity of the hill deposition operator,

$$v_n(x_n, \tilde{V}_n) \equiv \sum_{i=n}^{\infty} \epsilon_i (E_n[\Gamma(s, s_i^{\tilde{V}_n}, \tilde{V}_n)] - E_{p_b(s_i,t_n)}[\Gamma(s, s_i, \tilde{V}_n)]) \tag{7.30}$$

$$= \int \Gamma(s, s_i, \tilde{V}_n) \sum_{i=n}^{\infty} \epsilon_i (p_{\tilde{V}_n}(x_n, s_i, t_i - t_n) - p_b(s_i, t_n)) ds_i \tag{7.31}$$

and the first condition in the latter part of the previous section guarantees that this is $O(1/n)$. Therefore the third term is a sum of white noise terms with variable intensities that decay as $1/i$, and the squares of sums of those terms asymptotically decay to zero just like for the first term.

Finally, the third term is not random except in how it depends on the difference between two successive biases. That difference is $O(\epsilon_i) = O(1/i)$ because the updates are, and the terms in the difference are themselves of order $1/i$ as well by the argument in the previous paragraph. Therefore, if the projection of future non-quasiequilibrium artifacts were differentiable, this difference would be of order $1/i^2$. The second technical condition in the previous section guarantees that that is the case because the hill deposition function is differentiable. Sums of $O(1/i^2)$ terms are at most of order $1/i$, so these terms, too, go to zero, and the desired result is found.

$$\lim_{\tau \to \infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i ((\Gamma(s, s_i, \tilde{V}_{i-1}) - E_{i-1}[\Gamma(s, s_i, \tilde{V}_{i-1}))] +$$

$$(v_i(x_i, \tilde{V}_i) - v_i(x_i, \tilde{V}_{i-1})) +$$

$$(v_i(x_i, \tilde{V}_{i-1}) - E_{i-1}[v_i(x_i, \tilde{V}_{i-1})])) = 0,$$

$$\lim_{\tau \to \infty} \sum_{i=m(\tau)+1}^{m(\tau+d\tau)} \epsilon_i \Gamma(s, s_i, \tilde{V}_{i-1}) = \int_{\tau}^{\tau+d\tau} \int \Gamma(s, s', \tilde{V}(\tau')) p_b(s') ds' d\tau'. \tag{7.32}$$

This reverse-engineering required a few bits of circular logic and has been decidedly conceptual, but it gives a good idea of how the technical derivation works. It is a physicist's proof. The next section, which provides a mathematician's proof, is much like this one in reverse: it starts from the update, adds in these projections of future non-quasiequilibrium artifacts, splits the terms, shows partial sums of all terms but the quasiequilibrium update go to zero in the asymptotic limit, and then uses an existence theorem to show that that surviving term gives rise to an ODE.

7.3 Detailed Proof of the Convergence to a Differential Equation

7.3.1 Detailed Description of the Algorithm

Consider an instance of metadynamics applied to a system sampling states x from a state space X and applied through collective variables taking values s in a compact collective variable space S that are observed using some map $M : X \to S$. We assume S is compact because it almost always is in practice, typically being some product space of circles, line segments, and finite sets, so $\int_S ds$ is finite. When that is not the case, the argument in this SI can be recapitulated with integration over the natural Lebesgue measure, $\int_S ds$, replaced everywhere by integration over some almost-nowhere-zero finite reference measure, $\int_S d\mu_r(s)$, and the same conclusions will hold essentially unchanged. Let $\mathcal{M}$ be the marginalization map from measures on X to measures on S, $\mathcal{M}(\pi(x)) = \int \delta(s - M(x))d\pi(x)$. The time-dependent bias on these collective variables will be denoted by $V(s,t)$ and will be split into two parts for convenience: first, the average bias level $\bar{V}(t) = \int_S V(s,t)ds / \int_S ds$; second, the driving bias $\tilde{V} = V - \bar{V}$.

Assume that the dynamics of the system with a fixed bias are Markovian, time-independent, ergodic, independent of $\bar{V}(t)$, and have the transition probability measure $\pi_{\tilde{V}(s,t)}(x', x, t' - t)$ defining the expectation that the system transitions from state x at current time t to state x' at future time t' given that the bias is fixed and equal to $V(s,t)$ over the whole interval

$[t, t')$. As X may be an extended state space including dynamic memory variables, assuming Markovianity is not restrictive.

Typically it is the probability distributions over collective variables that concern us rather than distributions over underlying states. These distributions will be described using p-variables. In particular, define $p_{\tilde{V}(s,t)}(s, x, t' - t) = \mathcal{M}(\pi_{\tilde{V}(s,t)}(x', x, t' - t))$ to be the probability of observing collective variable state s at time t' given that the system is in state x at time t and the bias is held at $\tilde{V}(s, t)$ until t'. For any $\tilde{V}(s, t)$, insist that the stationary distribution of the dynamics satisfies $\lim_{t'-t\to\infty} p_{\tilde{V}(s,t)}(s, x, t' - t) \equiv p_{\tilde{V}(s,t)}(s) \equiv p_b(s, t) \sim p_a(s)e^{-\tilde{V}(s,t))/T}$, where $p_a(s)$ is the equilibrium distribution over collective variables in the unbiased system and T is a temperature. The subscript a stands for authentic and recalls the distribution's ordering with respect to other distributions used in this paper: p_a is biased into p_b until p_c is reached (up to reweighting). We will often leave the dependences of these quantities on s implicit where it seems cumbersome, and the objects always should be interpreted the same way regardless of whether or not the s-dependence is explicit or not. In particular, we will typically leave the s-dependence of function arguments implicit for functions of functions to prevent a profusion of dummy variables.

An instance of metadynamics is described by specifying an untempered hill deposition operator $\hat{G} \circ p = \int_S ds' G(s, s')p(s')$ that is linear and positive semidefinite with $G(s, s')$ bounded for each s' and $\int_S G(s, s')$ bounded from below by a positive number for each s', a well-tempering factor $\Delta T_{wt} \in (0, \infty)$, and a positive, bounded, Lipschitz global tempering functional $\omega(\tilde{V}(s, t))e^{-\bar{V}/\Delta T_{gt}}$ with $\Delta T_{gt} \in (0, \infty)$. Require that $\int_S ds' G(s, s')p_c(s') = C$ has at least one normalized, positive almost-everywhere solution p_c. With these specified, the metadynamics update rule can be written as

$$V(s, t_{n+1}) = V(s, t_n) + \omega(\tilde{V}(s, t_n))e^{-\bar{V}(t_n)/\Delta T_{gt}} G(s, s_{n+1})e^{-V(s_{n+1},t_n)/\Delta T_{wt}} \tag{7.33}$$

in terms of samples x_n taken at evenly spaced times $t_n = n\Delta t$ obtained from simulation

using the most recent bias on each time interval and s_n the corresponding collective variable state samples, $M(x_n) = s_n$. We consistently assume that all the biases sampled by a given metadynamics trajectory will lie within some bounded, compact region of the L_∞ bias-space, D, chosen by unspoken physical intuition.

Finally, realizations (individual runs with, say, particular random numbers generated) of the process are indexed by random variables ω taking values in some set Ω of possible realizations. Many of the functions written so far have implicit dependences on ω through their history dependence that we will sometimes make explicit by writing, for instance, $V(\omega, s, t)$ instead of $V(s, t)$ when it is important to draw the distinction between $V(s, t)$ obtained in different runs. We will make use of two basic types of expectation. First, we use conditional expectations E_n over ω that are conditional on all of the metadynamics-relevant simulation information up to time t_n, meaning all fine-grained states and biases: $\{x_i(\omega), V(\omega, s, t_i) : i \leq n\}$, so that, for instance, $E_n[\delta(s - s_{n+1})] = p_{\tilde{V}(s,t_n)}(s_{n+1}, x_n, \Delta t)$. Second, we use expectations over s with respect to distributions on s, for instance the expectation $E_{p_b(s,t)}[f(s)] = \int f(s)p_b(s, t)ds$.

7.3.2 Theorem Statement

For an instance of metadynamics as described above following the bias update rule (7.33),

$$V(s, t_{n+1}) = V(s, t_n) + \omega(\tilde{V}(s, t_n))e^{-\bar{V}(t_n)/\Delta T_{gt}} G(s, s_{n+1}) e^{-V(s_{n+1},t_n)/\Delta T_{wt}},$$

sequences of the driving bias in an auxiliary time scale, $\tilde{V}(s, \tau_n)$ with

$$\tau_n = \sum_{i=0}^{n} e^{-\bar{V}(t_i)(1/\Delta T_{wt} + 1/\Delta T_{gt})}, \tag{7.34}$$

asymptotically converge like solutions of the differential equation

$$\dot{\tilde{V}}(s,\tau) = \tilde{\omega}(\tilde{V}(\tau)) \int_S ds'(\delta(s-s') - (\int_S ds''')^{-1}) \int_S ds'' G(s',s'') e^{-\tilde{V}(s'',\tau)/\Delta T_{wt}} p_b(\tau).$$

with probability one given the following assumptions about the trajectories and the system:

1. The bias is physically reasonable, $\tilde{V}(s,\tau_n)$ remains in some bounded, compact region D of the space of functions on $\mathcal{S}$ for all τ.

2. The system relaxes to equilibrium reasonably quickly for all physically reasonable biases, $||\sum_{m=n+1}^{\infty} \frac{1}{m}(p_{V_n}(s,x_n,t_m-t_n) - p_b(s,V_n))||_{\infty(s,V_n\in D)} \leq K/n$ for some finite K for each visited x_n.

3. The system's dynamic response to biasing relaxes to its equilibrium value reasonably quickly for all physically reasonable biases, $||\sum_{m=n}^{\infty} \frac{1}{m}(p_V(s,x_n,t_m-t_n) - p_b(s,V) - p_{V'}(s,x_n,t_m-t_n) + p_b(s,V'))||_{\infty(s,V\in D,V'\in D)} \leq K/n||V-V'||_{\infty(s,V\in D,V'\in D)}$ for some finite K for each visited x_n.

And the following design constraints on the method:

1. G is bounded, $||G(s,s')||_{\infty(s,s')} < K$ for some constant K and $||\int G(s,s')f(s')||_{\infty(s)} < K'||f(s)||_{\infty(s)}$ for some constant K' and all $f(s)$.

2. G adds finite-positive-volume hills, $0 < K_l < \int_S dsG(s,s')/\int_S ds'' < K_h < \infty \; \forall \; s'$ for some constants K_l and K_h.

3. The tempering is finite and positive, $0 < (1/\Delta T_{wt} + 1/\Delta T_{gt}) < \infty$

All of these conditions are stated in the initial section of this chapter.

7.3.3 Conversion of the Iteration to Standard Form

This derivation will follow Theorem 6.6.1 of Kushner and Yin's (K&Y's) *Stochastic Approximation and Recursive Algorithms and Applications*[179] closely. The update (7.33),

$$V(s,t_{n+1}) = V(s,t_n) + \omega(\tilde{V}(s,t_n)) e^{-\bar{V}(t_n)/\Delta T_{gt}} G(s,s_{n+1}) e^{-V(s_{n+1},t_n)/\Delta T_{wt}},$$

can be rewritten into a recursive stochastic approximation algorithm in K&Y's standard form

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n \tag{7.35}$$

by defining

$$\theta_n(s) \equiv \tilde{V}(s, t_n) \tag{7.36}$$

$$\epsilon_n \equiv e^{-(1/\Delta T_{gt} + 1/\Delta T_{wt})\bar{V}(t_n)} \tag{7.37}$$

$$Y_n(s) \equiv \omega(\tilde{V}(s, t_n)) e^{-\tilde{V}(s_{n+1}, t_n)/\Delta T_{wt}} \left(G(s, s_{n+1}) - \int_S ds G(s, s_{n+1}) / \int_S ds \right). \tag{7.38}$$

For convenience, define an instantaneous tempered driving hill deposition kernel $\Gamma(s, \tilde{V}(s,t), s')$ by

$$\Gamma(s, \tilde{V}(s,t), s') \equiv \omega(\tilde{V}(s,t)) e^{-\tilde{V}(s',t)/\Delta T_{wt}} \left(G(s, s') - \omega(\tilde{V}(s,t)) \int_S ds G(s, s') / \int_S ds \right); \tag{7.39}$$

this is bounded given that $\tilde{V}$ is in D. Then the expectation of each next update just before it occurs, which we'll call $g(\theta_n, x_n)$, is

$$g(s, \theta_n, x_n) \equiv E_n[Y_n(s)] \tag{7.40}$$

$$= E_n[\Gamma(s, \theta_n, s_{n+1})] \tag{7.41}$$

$$= \int_S ds_{n+1} p_{\theta_n}(s_{n+1}, x_n, \Delta t) \Gamma(s, \theta_n, s_{n+1}); \tag{7.42}$$

the Markovianity of the system implies that this does not depend on any previous states or biases. Then this update can be rewritten

$$\theta_{n+1}(s) = \theta_n(s) + \epsilon_n [g(s, \theta_n, x_n) + \delta M_n(s)] \tag{7.43}$$

where

$$g(s, \theta_n, x_n) \equiv E_n[\Gamma(s, \theta_n, s_{n+1})] \tag{7.44}$$

and

$$\delta M_n(s) \equiv \Gamma(s, \theta_n, s_{n+1}) - E_n[\Gamma(s, \theta_n, s_{n+1})]. \tag{7.45}$$

This is now of the form treated by Theorem 6.6.1. The constraints in that theorem are unnecessary because we have assumed that the trajectory stays in a compact region given the extensive practical knowledge about metadynamics that has amassed since it was first tried. Now perform one last manipulation before beginning to verify the assumptions of that proof and then proving that metadynamics converges. Convert this into the recursion for a perturbed $\tilde{\theta}_n(s) = \theta_n(s) + v_n(s, \theta_n, x_n)$ where

$$v_n(s, \theta_n, x_n) \equiv \sum_{i=n}^{\infty} \epsilon_i E_n[g(s, \theta_n, x_i(\theta_n)) - \bar{g}(s, \theta_n)], \tag{7.46}$$

the $x_i(\theta_n)$ are states x_i from trajectories generated by seeding the dynamics at x_n and always evolving the state under the bias $\tilde{V}_n(s) \equiv \theta_n(s)$, and $\bar{g}(s, \theta_n) \equiv \int_S ds' p_{\theta_n}(s')\Gamma(s, \theta_n, s')$. The recursion for $\tilde{\theta}(s)$ is then

$$\tilde{\theta}_{n+1}(s) = \tilde{\theta}_n(s) + \epsilon_i[g(s, \theta_n, x_n) + \delta M_n(s)] + v_{n+1}(s, \theta_{n+1}, x_{n+1}) - v_n(s, \theta_n, x_n), \tag{7.47}$$

$$\tilde{\theta}_{n+1}(s) = \tilde{\theta}_n(s) + \epsilon_n[\bar{g}(s, \theta_n) + \delta M_n(s)] + \alpha_n(s) + \delta N_n(s), \tag{7.48}$$

where

$$\alpha_n(s) \equiv v_{n+1}(s, \theta_{n+1}, x_{n+1}) - v_{n+1}(s, \theta_n, x_{n+1}), \tag{7.49}$$

$$\delta N_n(s) \equiv v_{n+1}(s, \theta_n, x_{n+1}) - E_n[v_{n+1}(s, \theta_n, x_{n+1})]. \tag{7.50}$$

Of course, these manipulations can only be performed if the relevant sums $v_n(s, \theta_n, x_n)$ exist, which is discussed further in a later section. At this point, the equation (7.48) is in a form that we can prove converges to the quasiequilibrium update differential equation by verifying a number of assumptions about the step sizes, individual terms in the recursion, and the partial sums of the individual terms. The goal will be to show that all terms except for the $\bar{g}(s, \theta_n)$ are asymptotically irrelevant, to show that the survival of that term implies that the evolution of the bias converges like the solution of an ordinary differential equation. However, we start by characterizing the step size sequence and the $v_n(s, \theta, x_n)$.

7.3.4 Step Size Sequence Asymptotics

To prove that the iteration converges, our first step is to show that the ϵ_n satisfy K&Y's condition 5.1.1, that is, they satisfy

$$\sum_{n=0}^{\infty} \epsilon_n = \infty, \quad \epsilon_n \geq 0, \quad \lim_{n \to \infty} \epsilon_n = 0. \tag{7.51}$$

This can be done by bounding ϵ_n above and below using the fact that $\tilde{V}(s, t_n)$ is bounded. Begin with the definition of ϵ_n,

$$\epsilon_n = e^{-(1/\Delta T_{gt} + 1/\Delta T_{wt})\bar{V}(t_n)}, \tag{7.52}$$

and consider the change of $\bar{V}(t_n)$ from one timestep to the next,

$$\bar{V}(t_{n+1}) = \bar{V}(t_n) + \int_S ds \omega(\tilde{V}(s, t_n)) e^{-\bar{V}(t_n)/\Delta T_{gt}} G(s, s_{n+1}) e^{-V(s_{n+1}, t_n)/\Delta T_{wt}} / \int_S ds \tag{7.53}$$

$$\bar{V}(t_{n+1}) = \bar{V}(t_n) + e^{-(1/\Delta T_{gt} + 1/\Delta T_{wt})\bar{V}(t_n)} \int_S ds \omega(\tilde{V}(s, t_n)) G(s, s_{n+1}) e^{-\tilde{V}(s_{n+1}, t_n)/\Delta T_{wt}} / \int_S ds \tag{7.54}$$

Now, if $\tilde{V}(s,t_n)$ is in some bounded set D, $G(s,s')$ is bounded for each s', and $\int_S G(s,s')ds/\int_S ds$ is bounded from below by a positive quantity, then the quantity in the integrals is positive and bounded from above and below,

$$e^{-(1/\Delta T_{gt}+1/\Delta T_{wt})\bar{V}(t_n)}M_l \leq \bar{V}(t_{n+1}) - \bar{V}(t_n) \leq e^{-(1/\Delta T_{gt}+1/\Delta T_{wt})\bar{V}(t_n)}M_h \tag{7.55}$$

and, defining $\Delta T = (1/\Delta T_{gt} + 1/\Delta T_{wt})^{-1}$ and $\bar{V}^*$ to be a linear interpolation of $\bar{V}$ through its sampled values, then (neglecting the nonexistence of the derivative at the sampled points)

$$e^{-\bar{V}(t_n)/\Delta T}M_l \leq \bar{V}(t_{n+1}) - \bar{V}(t_n) \leq e^{-\bar{V}(t_n)/\Delta T}M_h \tag{7.56}$$

$$e^{-\bar{V}(t_n)/\Delta T}M_l/\Delta t \leq \frac{\bar{V}(t_{n+1}) - \bar{V}(t_n)}{\Delta t} \leq e^{-\bar{V}(t_n)/\Delta T}M_h/\Delta t \tag{7.57}$$

$$e^{-\bar{V}(t_n)/\Delta T}M_l/\Delta t \leq \frac{d\bar{V}^*(t)}{dt} \leq e^{-\bar{V}(t_n)/\Delta T}M_h/\Delta t \tag{7.58}$$

From which one can get a lower-bound differential equation from the monotonic increase of $\bar{V}^*$,

$$e^{-\bar{V}^*(t)/\Delta T}M_l/\Delta t \leq \frac{d\bar{V}^*(t)}{dt} \tag{7.59}$$

and rearrange it to show

$$d\bar{V}^*(t)e^{\bar{V}^*(t)/\Delta T} \geq M_l dt/\Delta t \tag{7.60}$$

$$\Delta T(e^{\bar{V}^*(t_n)/\Delta T} - e^{\bar{V}^*(t_0)/\Delta T}) \geq M_l(t_n - t_0)/\Delta t \tag{7.61}$$

$$(e^{\bar{V}^*(t_n)/\Delta T} - e^{\bar{V}^*(t_0)/\Delta T}) \geq M_l(t_n - t_0)/(\Delta T\Delta t) \tag{7.62}$$

$$e^{-\bar{V}^*(t_n)/\Delta T} \leq (e^{\bar{V}^*(t_0)/\Delta T} + (\Delta T\Delta t)^{-1}M_l(t_n - t_0))^{-1} \tag{7.63}$$

$$\epsilon_n \leq (a_h + b_h n)^{-1}. \tag{7.64}$$

To characterize the bound from below one must use a recursion relation instead of a differential equation. The conclusion is similar,

$$\epsilon_n \geq (a_l + b_l n)^{-1}. \tag{7.65}$$

The step sequence is bounded above and below by sequences that decay as $1/n$, proving that they satisfy the desired requirements (7.51).

The accumulation of steps sets an auxiliary time scale τ defined so that $\tau_n = \sum_{i=0}^{n} \epsilon_i$ and a reverse-mapping $m(\tau) = \sup\{n : \tau_n \leq \tau\}$; this is the auxiliary time scale discussed in the main manuscript. These will be essential in working towards the limit ODE.

7.3.5 Existence of a Useful Auxiliary Iteration

This section will prove the existence of the $v_n(s, \theta, x_n)$, required for the manipulations in the physicist's proof to make sense. First, let

$$v_n(s, \theta, x_n) = \sum_{i=n}^{\infty} \epsilon_n E_n(g(s, \theta, x_i(\theta_n)) - \bar{g}(s, \theta)). \tag{7.66}$$

Recall $x_i(\theta_n)$ are samples drawn from a dynamics with the bias held fixed at θ_n for all time, define $s_i(\theta_n)$ similarly as the corresponding collective variable value samples, and note that

$$E_n[g(s, \theta, x_i(\theta_n))] = E_n[E_i[\Gamma(\theta, s, s_{i+1}(\theta))]] \tag{7.67}$$

$$= \int_X dx_i \pi_\theta(x_i, x_n, (i-n)\Delta t) \int_S ds_{i+1}(\theta) p_\theta(x_i, s_{i+1}(\theta), \Delta t) \Gamma(s, \theta, s_{i+1}(\theta)) \tag{7.68}$$

$$= \int_S ds_{i+1}(\theta) p_\theta(x_n, s_{i+1}(\theta), (i-n+1)\Delta t) \Gamma(s, \theta, s_{i+1}(\theta)) \tag{7.69}$$

$$= E_n[\Gamma(s, \theta, s_{i+1}(\theta))] \tag{7.70}$$

after using Markovianity and temporal invariance to recognize that the previous equation was a Chapman-Kolmogorov equation. (Any such chained expectations would collapse like this.)

This makes it clear that a natural choice for $\bar{g}(s, \theta)$ is

$$\lim_{i-n\to\infty} E_n[\Gamma(s, \theta, s_{i+1}(\theta)) = \int ds_{i+1} p_\theta(s_{i+1})\Gamma(s, \theta, s_{i+1}) \tag{7.71}$$

$$\equiv E_{p_\theta(s)}[\Gamma(s, \theta, s)], \tag{7.72}$$

the hills deposited from a perfect sampling of an equilibrium distribution given the driving bias $\tilde{V}(s) = \theta(s)$. Then the entire sum is

$$v_n(s, \theta, x_n) = \sum_{i=n}^{\infty} \epsilon_i (E_n[\Gamma(s, \theta, s_{i+1}(\theta))] - E_{p_\theta(s)}[\Gamma(s, \theta, s)]) \tag{7.73}$$

$$= \int ds_{i+1} \Gamma(s, \theta, s_{i+1}) \sum_{i=n}^{\infty} \epsilon_i (p_\theta(s_{i+1}, x_n, (i + 1 - n)\Delta t) - p_\theta(s_{i+1})) \tag{7.74}$$

Clearly, if the system is geometrically ergodic for all $\theta(s)$ then this sum exists and in fact much weaker conditions can ensure that that is the case: if the differences between the transition probabilities and the equilibrium converge to zero with an algebraic decay under any norm, then the sum will converge to zero in that norm with at least the same algebraic decay because the differences are multiplied by the decaying sequence of step sizes. Therefore even if all time correlations in the system decay at slowest as power laws of any kind—the worst case of critical slowing down—this sum still exists.

This would suffice to show that v_n exists with $\bar{g}(s, \theta) = \int_S \Gamma(s, \theta, s')p_\theta(s')ds'$ under fairly general conditions. However, for the sake of later steps in the proof, we insist that the rate of convergence is fast enough that

$$v_n(\theta, x_n) = O(\epsilon_n). \tag{7.75}$$

Specifically, there must exist a K such that for all n, θ, and visited x_n,

$$||v_n(\theta, x_n)||_{\infty(s)} \leq \epsilon_n K \tag{7.76}$$

almost surely. This perfectly matches the second of our major assumptions on the trajectory and system in the theorem statement.

This will be the case if the system is geometrically ergodic (plus the χ_δ condition below) or even if the variation decays only algebraically at least as fast as

$$||p_\theta(s, x_n, t) - p_\theta||_{\infty(s,\theta)} \leq C_p(x_n)/t^{1+\epsilon} \tag{7.77}$$

at long times for some bounded set of constants $C_p(x_n)$ and any $\epsilon > 0$ for each visited x_n and so long as the initial transients where this may not hold are bounded in length and value.

A few comments on this condition may help make its meaning clearer. In general it is not possible to guarantee that $C_p(x_n)$ is bounded over all possible x_n, because there are some points (e.g., an infinite-velocity starting configuration) that take arbitrarily long to relax to equilibrium, but in practice trajectories do not sample such pathological configurations when the systems are prepared with a careful minimization and equilibration procedure before metadynamics is started. Intuitively, we need only worry that $C_p(x_n)$ is bounded over all visited x_n and then require that this is effectively always the case in our simulations. A good-enough precise requirement is this: that our trajectories sample a compact subdomain of X on which $C_p(x_n)$ is bounded with probability one, and this is hardly restrictive since it will be true if the trajectory can be represented with floating-point numbers. A simulation will crash before this condition is violated... which one cannot deny does happen with metadynamics not infrequently when CVs are chosen poorly. See the chapter on metabasin metadynamics for further discussion.

Next, also insist that $v_n(s, \theta, x_n)$ is Lipschitz continuous in $\theta(s)$ uniformly in n, x, and s

and that the derivatives also decay as ϵ_n; this is physically like saying that the mechanism of relaxation to equilibrium in the system depends smoothly on the applied bias. It matches the third of our major assumptions on the trajectory and system in the theorem statement. A more-than-sufficient condition for that third assumption to be satisfied is if

$$||(p_\theta(s, x_n, t) - p_\theta(s)) - (p_{\theta'}(s, x_n, t) - p_{\theta'}(s))||_{\infty(s,\theta,\theta')} \leq C_d(x_n)||\theta(s) - \theta'(s)||_{\infty(s)}/t^{1+\epsilon} \tag{7.78}$$

at all long times for some bounded set of constants $C_d(x_n)$ and any $\epsilon > 0$ for each visited x_n and the initial transients where this may not hold are bounded in length and value. Again, it will not always be the case that $C_d(x_n)$ is bounded over all possible x_n, but it is reasonable to assume that $C_d(x_n)$ is bounded over all visited x_n.

7.3.6 Asymptotic Irrelevance of Nuisance Terms

To prove that the term $\delta M_n(s)$ is irrelevant to the final converged solution, i.e. that the single-iterate noise averages itself out, begin by noting that each individual term

$$\delta M_n(s) = \Gamma(s, \theta_n, s_{n+1}) - E_n[\Gamma(s, \theta_n, s_{n+1})] \tag{7.79}$$

is a martingale difference with uniformly bounded variation. Therefore, following the proof of K&Y Theorem 5.3.3 (p. 138), the asymptotic rate of change of sums of this term is zero with probability one because, as shown above, the step size sequence is bounded from above by a sequence satisfying K&Y's A5.3.1 and A5.3.2. Use the discussion on p. 142 to prove that A5.3.3 holds in this case. The specific methods used in K&Y would use that the variation of $\delta M_n(s)$ at each point is bounded to prove that their sums converge pointwise, but starting a uniform bound on the variation at each point gives a uniform convergence

result. Therefore the asymptotic rate of change of

$$M^0(s,\tau) = \sum_{i=0}^{m(\tau)-1} \epsilon_i \delta M_i(s) \tag{7.80}$$

almost surely converges uniformly to zero: the update over τ intervals of constant size goes to zero almost surely, for each $T > 0$ and $\mu > 0$,

$$\lim_{n\to\infty} P\{\sup_{j\geq n} \max_{0\leq t\leq T} || \sum_{i=m(jT)}^{m(jT+t)+1} \epsilon_i \delta M_i(s)||_\infty(s) \geq \mu\} = 0. \tag{7.81}$$

To prove that the term $\delta N_n(s)$ is irrelevant to the final converged solution, i.e. that the single-perturbation-iterate noise averages itself out, begin by noting that each individual term

$$\delta N_n(s) = v_{n+1}(s, \theta_n, x_{n+1}) - E_n[v_{n+1}(s, \theta_n, x_{n+1})] \tag{7.82}$$

is a martingale difference. As described earlier, $v_n(\theta, x_n, s)$ for all $n, \theta(s), x_n$ are bounded by $\epsilon_n K_h(s), \epsilon_n K_l(s)$ at each point from above and below respectively, so these differences are also bounded by a term proportional to ϵ_n almost surely, and this term, like $\delta M_n(s)$, is a martingale difference of uniformly bounded variation with step sizes ϵ_n,

$$\epsilon_n(K_l - K_h) \leq \delta N_n(s) \leq \epsilon_n(K_h - K_l), \tag{7.83}$$

and therefore the asymptotic rate of change of

$$N^0(s,\tau) = \sum_{i=0}^{m(\tau)-1} \delta N_i(s) \tag{7.84}$$

goes to zero with probability one by the same logic. Therefore, for each $T > 0$, and $\mu > 0$,

$$\lim_{n\to\infty} P\{\sup_{j\geq n} \max_{0\leq t\leq T} || \sum_{i=m(jT)}^{m(jT+t)+1} \delta N_i(s)||_\infty(s) \geq \mu\} = 0. \tag{7.85}$$

To prove that the term $\alpha_n(s)$ is irrelevant to the final converged solution, i.e. that the bias due to the perturbation becomes negligible, begin by noting that each individual term

$$\alpha_n(s) = v_{n+1}(s, \theta_{n+1}, x_{n+1}) - v_{n+1}(s, \theta_n, x_{n+1}) \tag{7.86}$$

is a difference between two $O(\epsilon_n)$ terms. We have thus shown that every term in the update is almost surely of order ϵ_n, and if so $\theta_{n+1}(s) - \theta_n(s)$ is of order ϵ_n. Given that the $v_n(\theta, x, s)$ are Lipschitz continous in θ uniformly in n, x, and s and the Lipschitz coefficient decays as ϵ_n, then one may bound this update using

$$\alpha_n(s) \leq \epsilon_n K ||\theta_{n+1}(s) - \theta_{n+1}(s)||_\infty(s) \tag{7.87}$$

$$= \epsilon_n K O(\epsilon_n) \tag{7.88}$$

$$= K O(\epsilon_n^2) \tag{7.89}$$

almost surely for K from Lipschitz continuity (section 3D). Then the asymptotic bounds on ϵ_n determined in Section 3C imply that the sequence of $\alpha_n(s)$ are almost surely summable and therefore that the asymptotic rate of change of

$$A^0(\tau, s) = \sum_{i=0}^{m(\tau)-1} \alpha_i(s) \tag{7.90}$$

uniformly converges to zero almost surely: for each $T > 0$ and $\mu > 0$

$$\lim_{n\to\infty} P\{\sup_{j\geq n} \max_{0\leq t\leq T} || \sum_{i=m(jT)}^{m(jT+t)+1} \alpha_i(s)||_\infty(s) \geq \mu\} = 0. \tag{7.91}$$

7.3.7 Convergence Like a Differential Equation

With the asymptotic convergence of those terms established, it is now time to examine the asymptotic convergence of the overall process. As usual, we follow the proof of Theorem 6.6.1 from K&Y (p.187). Recalling the definition of the time scale τ, $\tau_n = \sum_{i=0}^{n} \epsilon_i$, and of $m(\tau)$, $m(\tau) = n \ : \tau_n \leq \tau < \tau_{n+1}$, drop the s-dependence of all functions for the moment and define the function-valued shifted processes $G^n(\tau)$, $M^n(\tau)$, $N^n(\tau)$, $A^n(\tau)$, $\theta^n(\tau)$ on $-\infty < \tau < \infty$ by

$$G^n(\tau) = \sum_{i=n}^{m(\tau_n+\tau)-1} \epsilon_i \bar{g}(\theta_i) \text{ for } \tau \geq 0, \qquad - \sum_{i=m(\tau_n+\tau)}^{n-1} \epsilon_i \bar{g}(\theta_i) \text{ for } \tau < 0 \tag{7.92}$$

$$M^n(\tau) = \sum_{i=n}^{m(\tau_n+\tau)-1} \epsilon_i \delta M_i \text{ for } \tau \geq 0, \qquad - \sum_{i=m(\tau_n+\tau)}^{n-1} \epsilon_i \delta M_i \text{ for } \tau < 0 \tag{7.93}$$

$$N^n(\tau) = \sum_{i=n}^{m(\tau_n+\tau)-1} \delta N_i \text{ for } \tau \geq 0, \qquad - \sum_{i=m(\tau_n+\tau)}^{n-1} \delta N_i \text{ for } \tau < 0 \tag{7.94}$$

$$A^n(\tau) = \sum_{i=n}^{m(\tau_n+\tau)-1} \alpha_i \text{ for } \tau \geq 0, \qquad - \sum_{i=m(\tau_n+\tau)}^{n-1} \alpha_i \text{ for } \tau < 0 \tag{7.95}$$

$$v^n(\tau) = v_n(\theta_n, x_n) - v(\theta^n(\tau), x_{m(\tau_n+\tau)-1}) \tag{7.96}$$

$$\theta^n(\tau) = \theta_n + G^n(\tau) + M^n(\tau) + N^n(\tau) + A^n(\tau) + v^n(\tau). \tag{7.97}$$

Also introduce

$$\rho^n(\tau) = G^n(\tau) - \int_0^\tau \bar{g}(\theta^n(\tau'))d\tau' \tag{7.98}$$

so that

$$\theta^n(\tau) = \theta_n + \int_0^\tau \bar{g}(\theta^n(\tau'))d\tau' + M^n(\tau) + N^n(\tau) + A^n(\tau) + v^n(\tau) + \rho^n(\tau). \tag{7.99}$$

As proven in the previous sections, processes $M^n(\tau)$, $N^n(\tau)$, and $A^n(\tau)$ converge to the zero process in the limit of large n almost surely, that is, except for ω in some null set O. Additionally, $\rho^n(\tau) = 0$ whenever $\tau = \tau_k - \tau_n$ for $k > n$, and $\rho^n(\tau) \to 0$ uniformly in s and τ as $n \to \infty$ by definition given the asymptotic decay rate of ϵ_n. $v^n(\tau)$ goes to zero by construction, as well.

Then for any $\omega \notin O$, by definition of O all of the families of functions $\int_0^\tau \bar{g}(s, \theta^n(\tau))ds$, $M^n(\tau)$, $N^n(\tau)$, $A^n(\tau)$, $v^n(\tau)$, and $\rho^n(\tau)$ on the right hand side satisfy the following condition: for each $T > 0$ and $\epsilon > 0$, there exists a δ such that

$$\limsup_n \sup_{0 \leq \tau - \tau' \leq \delta, |\tau| < T} ||f^n(\tau) - f^n(\tau')||_\infty(s) \leq \epsilon. \tag{7.100}$$

To use K&Y's terminology, each of the families is equicontinuous in an extended sense. The limits of $M^n(\tau)$, $N^n(\tau)$, $A^n(\tau)$, $v^n(\tau)$, and ρ^n are zero, which are smooth, while the integral is also smooth by definition and θ^n has jumps that decay as $1/n$ in the limit superior. Therefore we may apply a Banach space analog of K&Y Theorem 4.2.1 to show that the evolution converges to the limit ODE of the letter.

Consider each $\omega \notin O$ and for each bounded interval $[0, T]$. Choose a countable dense subset $\tau_k : \tau_k \in [0, T]$. Then because the τ-values of the sequence (the biases) come from a compact, bounded space of functions D, for each τ_k we can find we can find a subsequence n_k such that $\{\theta^{n_k}(\omega, \tau_k)\}$ converges. Moreover, from each subsequence that converges for τ_1 we can pick a further subsequence that converges for τ_2 as well, and so on for all τ_k. Call the resulting diagonal subsequences for any number p of distinct times $\{\theta^{n_{p,p}}(\omega)\}$.

Now for every δ, we can find some finite set of p τ_j from the countable dense subset chosen above such that $[0, T]$ is covered by union of the intervals $B(\tau_j, \delta)$ for $1 \leq j \leq p$ and a corresponding subsequence of the iteration that converges at all those times. Furthermore, by the equicontinuity above, for each $\epsilon > 0$ we can find a δ such that $||f^n(\tau) - f^n(\tau')||_\infty \leq \epsilon$ for all $|\tau - \tau'| < \delta$, and by the convergence of the subsequence above, for each ϵ there exists

an integer N such that $||\theta^{r_{p,p}}(\omega,\tau_j) - \theta^{s_{p,p}}(\omega,\tau_j)||_\infty(s) < \epsilon$ for all $r > N$ and $s > N$ and any j. Then for any τ and τ' such that $|\tau - \tau'| < \delta$, there exists a τ_j such that

$$\begin{aligned}||\theta^{r_{p,p}}(\omega,\tau) - \theta^{s_{p,p}}(\omega,\tau')||_\infty(s) \leq ||\theta^{r_{p,p}}(\omega,\tau) - \theta^{r_{p,p}}(\omega,\tau_j)||_\infty(s) \ \ldots \\ \ldots \ + ||\theta^{r_{p,p}}(\omega,\tau_j) - \theta^{s_{p,p}}(\omega,\tau_j)||_\infty(s) + ||\theta^{s_{p,p}}(\omega,\tau) - \theta^{s_{p,p}}(\omega,\tau_j)||_\infty(s)\end{aligned} \tag{7.101}$$

That first and third of these terms are less than ϵ follows from the uniform equicontiuity and initial choice of δ. The second is less also than ϵ for all $r > N$ and $s > N$ for the choice of N above, so we conclude that

$$||\theta^{r_{p,p}}(\omega,\tau) - \theta^{s_{p,p}}(\omega,\tau')||_\infty(s) \leq 3\epsilon \tag{7.102}$$

for all $r > N$, $s > N$, and δ chosen above. Therefore the diagonal subsequences $\{\theta^{n_{p,p}}(\omega)\}$ uniformly converge to a uniformly continuous limit.

Now consider the form of any these subsequential limits. In particular, observe that because we have proven that $M^n(\tau)$, $N^n(\tau)$, $A^n(\tau)$, $v^n(\tau)$, and ρ^n converge uniformly to zero, any such limits must satisfy the integral equation

$$\theta(\omega,\tau) = \theta(\omega,0) + \int_0^\tau \bar{g}(\theta(\omega,\tau'))d\tau', \tag{7.103}$$

given some initial condition $\theta(\omega,0)$. If this ODE has a certain set of stable limit points that all trajectories are drawn to, then this implies that all convergent subsequences of the original sequence $\{\theta_n(\omega)\}$ must converge to the asymptotically stable limit points of that ODE. It is in this sense that the evolution of the bias converges to the solution of the ODE: the tail behavior of any convergent subsequence of biases converges to a solution of the ODE with some initial condition, and the biases converge to the points that the ODE converges to initiated from any of the possible initial conditions for these convergent subsequences. Note

that there is no requirement that the method converge to any particular, single solution of the ODE—a subsequence of biases may jump back and forth between solutions of the ODE with different initial conditions, for instance.

This suffices to show that for any $\omega \notin O$ the $\theta^n(\omega, \tau)$ that converge converge asymptotically like solutions of

$$\dot{\theta}(\tau) = \bar{g}(\theta(\tau)) \tag{7.104}$$

if bias remains bounded and physically reasonable. This suffices to show how the ODE is derived and prove the theorem above. The next section provides the algebra needed to show that it converges to the desired end state for any physically reasonable initial condition, proving that all convergent subsequences converge to the desired state. Given the boundedness and compactness of the set of allowed biases, that any convergent subsequence converges to a single bias implies that every stochastic sequence of metadynamics biases converges almost surely.

7.4 Convergence of the Differential Equation

In lieu of solving the infinite system of ordinary differential equations from the above proof explicitly, we instead used a Lyapunov function approach to prove that the all solutions away from the boundary converge to a unique stable fixed point. This section of provides the detailed algebra required to verify that the expression for the derivative of the relative entropy provided above is in fact correct and that, as asserted there, the same Lyapunov function can also be adapted to the case when there is not a unique fixed point for the ODE.

To begin, rewrite the limit equation for the driving bias,

$$\dot{\tilde{V}}(s, \tau) = \tilde{\omega}(\tilde{V}(\tau)) \int_S ds' (\delta(s - s') - (\int_S ds''')^{-1}) \int_S ds'' G(s', s'') e^{-\tilde{V}(s'', \tau)/\Delta T_{wt}} p_b(\tau). \tag{7.105}$$

This is zero when $\int_S ds' G(s,s') e^{-\tilde{V}(s',\tau)/\Delta T_{wt}} p_b(\tau)$ is a constant.

Suppose for the moment that $\int_S ds' G(s,s') p_c(s') = C$ has a unique normalized solution p_c. $\int_S ds' G(s,s')$ is positive definite in this case and not simply positive semidefinite, or the solution would not be unique. Then some measure of the difference between $e^{-\tilde{V}(s',\tau)/\Delta T_{wt}} p_b(\tau)$ and p_c would appear to make an excellent Lyapunov function, and we will now show that in fact the Kullback-Leibler divergence (relative entropy)

$$D_{KL}(p_c||p_w) = -\int_S ds p_c \ln \frac{p_w}{p_c}, \tag{7.106}$$

where $p_w \sim e^{-\tilde{V}(s',\tau)/\Delta T_{wt}} p_b(\tau)$ is the normalized tempering-weighted instantaneous equilibrium sampling distribution, is a Lyapunov function for (7.105) in the special case that p_c is unique.

First, $D_{KL}(p_c||p_w)$ is a continuously differentiable real-valued functional of $\tilde{V}$ on each set D of physically reasonable bounded driving biases for any almost-nowhere-zero density p_c and almost-nowhere-zero density p_w. $D_{KL}(p_c||p_w) = 0$ for $p_w = p_c$ and $D_{KL}(p_c||p_w) > 0$ for $p_w \neq p_c$. In terms of the bias, $D_{KL}(p_c||p_w) = 0$ for $\tilde{V} = (1+T/\Delta T_{wt})^{-1}(\ln p_a - \ln p_c) + C$ and $D_{KL}(p_c||p_w) > 0$ otherwise. The time derivative of $D_{KL}(p_c||p_w)$ is given by

$$\dot{D}_{KL}(p_c||p_w) = -\int_S ds \frac{p_c(s)}{p_w(s,\tau)} \dot{p}_w(s,\tau), \tag{7.107}$$

which, noting that $\int_S ds \dot{p}_w(s,\tau) = 0$ as a consequence of its normalization, is

$$\dot{D}_{KL}(p_c||p_w) = -\int_S ds (1 - \frac{p_c(s)}{p_w(s,\tau)}) \dot{p}_w(s,\tau), \tag{7.108}$$

and using

$$p_w(s,\tau) = p_a(s) e^{-(1/T+1/\Delta T_{wt})\tilde{V}(s,\tau)} / \int_S ds' p_a(s') e^{-(1/T+1/\Delta T_{wt})\tilde{V}(s',\tau)} \tag{7.109}$$

and

$$\dot{p}_w(s,\tau) = -(1/T + 1/\Delta T_{wt})p_w(s,\tau)(\dot{\tilde{V}}(s,\tau) - \int_S ds' \dot{\tilde{V}}(s',\tau)p_w(s',\tau)), \tag{7.110}$$

becomes

$$\begin{aligned}\dot{D}_{KL}(p_c||p_w) &= -(1/T + 1/\Delta T_{wt}) \int_S ds(1 - \frac{p_c(s)}{p_w(s,\tau)})p_w(s,\tau) \ \ldots \\ &\ldots \cdot \left(\dot{\tilde{V}}(s,\tau) - \int_S ds' \dot{\tilde{V}}(s',\tau)p_w(s',\tau)\right) \qquad (7.111) \\ &= -(1/T + 1/\Delta T_{wt}) \int_S ds(p_w(s,\tau) - p_c(s))(\dot{\tilde{V}}(s,\tau) - \int_S ds' \dot{\tilde{V}}(s',\tau)p_w(s',\tau)) \qquad (7.112) \\ &= -(1/T + 1/\Delta T_{wt}) \int_S ds(p_w(s,\tau) - p_c(s))\dot{\tilde{V}}(s,\tau) \qquad (7.113)\end{aligned}$$

because $\int_S ds(p_w(s,\tau) - p_c(s))C = 0$ for any constant C and $\int_S ds' \dot{\tilde{V}}(s',\tau)p_w(s',\tau)$ is a constant (does not depend on s). Similarly, plugging in (7.105), the $(\int_S ds''')^{-1} \int_S ds'$ term will disappear because it is a constant offset, and, normalizing $e^{-\tilde{V}(s',\tau)/\Delta T_{wt}} p_b(\tau)$ to get p_w times a normalization constant, then absorbing all the constants into $\tilde{\omega}$,

$$\dot{D}_{KL}(p_c||p_w) = -\tilde{\omega}(\tilde{V}(\tau)) \int_S ds \int_S ds'(p_w(s,\tau) - p_c(s))G(s,s')p_w(s,\tau) \tag{7.114}$$

And finally, using $\int_S ds' G(s,s')p_c(s) = C$ and $\int_S ds(p_w(s,\tau) - p_c(s))C = 0$,

$$\dot{D}_{KL}(p_c||p_w) = -\tilde{\omega}(\tilde{V}(\tau)) \int_S ds \int_S ds'(p_w(s,\tau) - p_c(s))G(s,s')(p_w(s,\tau) - p_c(s)) \leq 0 \tag{7.115}$$

taking advantage of the positive definiteness of $\int_S ds' G(s,s')$. It is zero precisely where $p_w(s,\tau) = p_c(s)$ and nowhere else, that is, where $\tilde{V} = (1 + T/\Delta T_{wt})^{-1}(\ln p_a - \ln p_c) + C$ and nowhere else. Therefore solutions of (7.105) converge to the point $\tilde{V} = (1 +$

$T/\Delta T_{wt})^{-1}(\ln p_a - \ln p_c) + C$ asymptotically. Therefore the driving bias converges almost surely to $\tilde{V} = (1 + T/\Delta T_{wt})^{-1}(\ln p_a - \ln p_c) + C$ if p_c is unique and the biases stay in the compact region D.

Now suppose that $\int_S ds' G(s, s')p(s) = C$ has many solutions with at least one positive everywhere; together these define a linear space $\mathcal{N}_g$. The algebra above still applies to Kullback-Leibler divergences $D_{KL}(p_c||p_w)$ for for any nonnegative normalized $p_c \in \mathcal{N}_g$,

$$\dot{D}_{KL}(p_c||p_w) = -\tilde{\omega}(\tilde{V}(\tau)) \int_S ds \int_S ds' (p_w(s,\tau) - p_c(s))G(s,s')(p_w(s,\tau) - p_c(s)) \leq 0, \tag{7.116}$$

except that now the zero equality holds for any point such that $p_w \in \mathcal{N}_g$ because $G(s, s')$ is positive semidefinite rather than positive definite and $p_w \in \mathcal{N}_g$ is precisely the condition for the above equation to be zero. By construction p_w will always be nonnegative, normalized and correspond to a driving bias in D, so one suspects that the set of nonnegative, normalized elements of $\mathcal{N}_g$ that correspond to some $\tilde{V} \in D$, $\tilde{\mathcal{N}}_g$, will be the set of possible end states of the ODE described in the previous section. The remainder of this section will show that the set of driving biases corresponding to $\tilde{\mathcal{N}}_g$ is indeed the set of possible asymptotic end points of metadynamics by considering the behavior of the function $L(p_w) = \inf_{p_c \in \tilde{\mathcal{N}}_g} D_{KL}(p_c||p_w)$.

First, note that $L(p_w) \geq 0$ and attains its minimum value if and only if $p_w \in \tilde{\mathcal{N}}_g$. Next note that $\dot{L}(p_w) = 0$ for $p_w \in \tilde{\mathcal{N}}_g$ because $\dot{D}_{KL}(p_c||p_w) = 0$ for all p_c if $p_w \in \tilde{\mathcal{N}}_g$. Next, consider the case that $p_w \notin \tilde{\mathcal{N}}_g$. In this case, for all $p_c \in \tilde{\mathcal{N}}_g$, $D_{KL}(p_c||p_w) > 0$ by definition and $\dot{D}_{KL}(p_c||p_w) < 0$ by (7.115) and the definition of $\tilde{\mathcal{N}}_g$. We can use this inormation to prove that $L(p_w) = \inf_{p_c \in \tilde{\mathcal{N}}_g} D_{KL}(p_c||p_w)$ decreases by considering any ϵ-minimizer $D^{\epsilon}_{KL}(p_c||p_w)$ of $D_{KL}(p_c||p_w)$ and showing that its value decreases in time. For all $\epsilon > 0$, there exists

$$D_{KL}(p^{\epsilon}_c||p_w(t)) - \epsilon \leq L(p_w(t)). \tag{7.117}$$

The first time derivative is known and the function $D_{KL}(p_c^\epsilon||p_w(t))$ has derivatives to all orders, so for some K and small $\delta > 0$ one has

$$D_{KL}(p_c^\epsilon||p_w(t+\delta)) < D_{KL}(p_c^\epsilon||p_w(t)) + \delta \dot{D}_{KL}(p_c^\epsilon||p_w(t)) + K\delta^2 \tag{7.118}$$

$$D_{KL}(p_c^\epsilon||p_w(t+\delta)) - \delta \dot{D}_{KL}(p_c^\epsilon||p_w(t)) - K\delta^2 < D_{KL}(p_c^\epsilon||p_w(t)) \tag{7.119}$$

which implies

$$D_{KL}(p_c^\epsilon||p_w(t+\delta)) - \delta \dot{D}_{KL}(p_c^\epsilon||p_w(t)) - K\delta^2 - \epsilon < L(p_w(t)). \tag{7.120}$$

Now, for any small $\epsilon > 0$ and negative $\dot{D}_{KL}(p_c^\epsilon||p_w(t))$, there exists a small $\delta > 0$ such that

$$-\delta \dot{D}_{KL}(p_c^\epsilon||p_w(t)) - K\delta^2 - \epsilon > 0. \tag{7.121}$$

Moreover, if $\dot{D}_{KL}(p_c^\epsilon||p_w(t))$ is bounded from above by a negative number as a function of p_c, as it is in our case whenever $p_w \notin \tilde{\mathcal{N}}_g$, then for all small $\delta > 0$ there exists an $\epsilon > 0$ such that all ϵ-minimizers of $\inf_{p_c \in \tilde{\mathcal{N}}_g} D_{KL}(p_c||p_w)$ decrease by an amount proportional to δ, and thus that $L(p_w(t+\delta)) \leq L(p_w(t)) - R\delta$ for some R which depends on $p_w(t)$ and for all small δ. Then L is minimized only on the target set and is always decreasing except on that target set, so by the usual Lyapunov function reasoning (7.105) always converges to some point such that p_w is in the target set provided the bias remains within the compact region of physically reasonable biases, D, on which this function is continuous and continuously differentiable.

We conclude that for any positive semidefinite $\int_S ds' G(s,s')$ with at least one positive normalized p_c such that $\int_S ds' G(s,s') p_c(s') = C$, the method will almost surely converge to one of the p_c if the driving bias stays in the physically reasonable region D.

CHAPTER 8

MAXIMIZING EXPLORATION: TRANSITION-TEMPERED METADYNAMICS

This chapter substantially reproduces the material from the paper Dama, J. F., Rotskoff, G., Parrinello, M., & Voth, G. A. (2014). Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. Journal of Chemical Theory and Computation, 10(9), 36263633. I am indebted to Michele for encouragement to complete this project and Grant Rotskoff for his enthusiasm for the method and help setting up and running the first of the simulations of myoglobin.

This section describes a way to ensure early exploration is efficient in metadynamics, to remove the tradeoff between convergence speed and exploration efficiency.

We now imagine our imp has some idea that the bachelor is most drawn to two cities, say Berlin and Freiburg, and that if he can ensure that he feels satisfied in neither then he will be satisfied nowhere. The imp whispers his hardest until the bachelor has traveled one way, and only lessens his whispering each time the bachelor moves again. In this way he drives the bachelor out of Freiburg as quickly as possible, then Berlin less quickly, then Freiburg again less quickly, always fast enough to drive him out but decreasing at exactly the rate required to ensure that eventually his preference between the two is completely nullified and he can find no satisfaction anywhere.

No longer does our imp need to modulate his whispers before driving the bachelor out of his home city; the end result is the same but the drama grows more rapidly.

8.1 Motivation

Well-tempered metadynamics (WTMetaD) adds increments of bias that smoothly taper to zero as the bias grows across the reaction coordinate space to ensure that the bias does eventually converge to the desired umbrella potential. It has close connections with the

Wang-Landau $1/t$ method[33] and the Robbins-Munro algorithm[278] in that the size of WTMetaD updates asymptotically decays as $1/t$.[49, 89] Unfortunately, the tapering rate is set using a tuning parameter that should match the barrier height to be overcome, introducing a requirement for information that metadynamics was initially designed to do without. Furthermore, bias updates, hereafter referred to as added hills, are tapered from the start to achieve long-time convergence, forcing researchers to make tradeoffs between the time until they see the first barrier crossing events and the asymptotic convergence rate. Higher tapering can accelerate convergence, but will slow escape from basins. Since barrier crossings are often an important preliminary sign of whether or not the method is working but the converged free energy is the final goal, this is not a pleasant tradeoff to accept.

This chapter presents a new metadynamics method, TTMetaD, that is smoothly convergent like WTMetaD and the $1/t$ method but does not require foreknowledge of the barrier height and avoids the tradeoff between fast barrier crossings and fast convergence. Instead, it asks only for approximate basin positions and decreases the height of added hills roughly according to the number of round trips between the basins. The basins can be chosen using a combination of experimental data and physical insight, for instance crystal structures of different conformers, FRET measurements, or intuition for peptide bond angles, and they will typically be a subset of only a few of the total number of basins in the system, chosen based on the goals of the metadynamics investigation. It is an asymptotically convergent method like WTMetaD, and in the examples we have examined it consistently and substantially improves upon WTMetaD in terms of both barrier crossing rates and convergence rates.

8.2 Method Definition

TTMetaD is inspired by an elegantly idealized picture of metadynamics developed by Laio et al.[183] and Bussi et al.[57] These works showed that metadynamics can profitably be thought of as a two-stage process. In the first stage, hills of bias fill the basins in the energy

landscape to the point that the biased landscape has a uniform roughness set by the size of the hills added. After this point is reached, in the second stage the biased landscape fluctuates indefinitely, keeping that characteristic roughness indefinitely as well. In this picture, tempering the hills—decreasing the size of added hills as time goes on—is a way of ensuring that the characteristic roughness of the fluctuating bias eventually falls to zero so that the biased surface becomes perfectly flat. Tempering shortens the second phase because it decreases the characteristic roughness, but lengthens the first phase because it decreases the basin filling rate. TTMetaD is designed to minimize the overall runtime of the method by ensuring that the hills are not tempered at all in the first phase, making that phase as short as possible, but also that they are still tempered properly in the second phase, making that phase also as short as possible.

Accomplishing this design requires some automatic way to recognize that the second phase has begun, and its definition provides a natural choice: the second phase occurs when all of the major basins of the PMF are filled by the bias and the biases filling them link up across the transition regions between them. Put differently, given a PMF with a few major basins, the second phase begins once there are paths between all of those basins that never cross a region in which the bias is zero. In fact, we can go further and measure the amount of overfilling quantitatively as the minimum bias on the maximally-biased path between the basins; call this V^*. As long as the biases filling the basins are not connected V^* remains zero, because any path between the basins must enter an unbiased region, and it grows steadily after the connections are established.

In the two-phase picture of metadynamics, the more overfilled the basins are the smaller the hills should be. This provides a recipe for TTMetaD. The recipe follows our work on the convergence of metadynamics and the example of WTMetaD, in which tempering is achieved by replacing the hill height h with $he^{-V(s(t),t)/\Delta T}$ everywhere it appears. $V(s(t),t)$ is the bias at the CV-space point $s(t)$ sampled at time t; s may be a discrete number, a continuous number, a vector, a function, or a combination of all of these. ΔT is the tuning parameter

discussed in the introduction that sets the speed of convergence. In TTMetaD we replace h with $he^{-V^*(t)/\Delta T}$ everywhere it appears. This is a special case of globally tempered metadynamics as defined in our convergence work[89] because in the notation of that paper $e^{-V^*(t)/\Delta T} = e^{-\bar{V}(t)/\Delta T}e^{-\tilde{V}^*(t)/\Delta T}$ and $e^{-\tilde{V}^*(t)/\Delta T}$ is a positive-valued functional of the driving bias $\tilde{V}$ alone. Using an exponential form like this ensures that the updates for TTMetaD decay as $1/t$ asymptotically like WTMetaD, Wang-Landau $1/t$ methods,[33] and Robbins-Munro.[278] That guarantees asymptotic convergence.[89]

That suffices to describe the essence of the method. The remainder of this section discusses technical details, providing explicit equations for the update rule and describing our software implementation.

First, we define V^*. We will give two definitions, one which is more intuitive and the second of which is less clear but is completely unambiguous once understood. V^* as introduced above is the minimum bias on the maximally-biased path between all basins of interest. These basins of interest are specified in advance by defining a set of n CV-coordinates s_b, where b is a basin index. For instance, in the case of an A to B transition, n will be two, s_A and s_B, no matter how many intervening basins are involved in the transition mechanism. These s_b are points within each basin of interest and need not be the minimum-energy points in each basin. Given a bias surface $V(s)$, we can then look at paths $s(\lambda)$, with $\lambda \in [0, 1]$ a path progress variable, that link all of the points s_b, then pick out one that is maximally biased, $s^*(\lambda)$. Then V^* is the minimum bias on the path $s^*(\lambda)$.

Examine once more how this definition detects the transition between the initial filling phase and the second convergence phase of tempered metadynamics. In the inital phase, the bias will not entirely fill all of the basins. Therefore, the bias-surface looks like a number of hills, corresponding to partially-filled basins, separated by regions where the bias is zero. Any path between the hills must pass through a region where the bias is zero, so V^* must also be zero. As the filling stage continues, these initially fully-separated hills start to overlap, but V^* still remains zero until every one of the bias hills corresponding to basins chosen by s_b is

nonzero and overlaps with basins that connect up with the others. This behavior guarantees that there is no need to know in advance where the barrier is, which basins will fill first, or anything more than the rough locations of the basins of interest in order to detect the end of the initial filling phase. It also gives a clear picture of how roughly those basin locations can be defined: so long as the basin-defininig coordinate s_b is within a reasonable isosurface of free energy around the true basin minimum, say half the barrier-basin free energy difference, then it will work well for TTMetaD. This is very permissive.

Now we unambiguously define what is meant by maximally biased path. First define the uncountable set $\mathcal{P}$ of all continuous paths $s(\lambda)$ that pass through all of the selected basin points s_b in the landscape; $\lambda \in [0, 1]$ indexes the points along the path from start ($\lambda = 0$) to finish ($\lambda = 1$) as before. Then

$$V^*(t) \equiv \max_{s(\lambda)\in\mathcal{P}} \min_{\lambda\in[0,1]} V(s(\lambda), t). \tag{8.1}$$

In other words, maximally biased paths are those for which the minimum bias is maximal. Since this definition allows for many possible equally-maximally-biased paths, there is some flexibility in choosing $s^*(\lambda)$, e.g., $s^*(\lambda)$ may or may not have any number of loops within the individual already-filled basins. However, even though multiple $s^*(\lambda)$ might be possible, V^* is always defined unambiguously by this definition. It suffices to find any single $s^*(\lambda)$ to find V^*.

Because metadynamics is stochastic, it will sometimes be possible that bias connects across the transition regions before the basins are fully filled. To guard against that possibility, we can add a bias threshold that serves to delay tempering until multiple crossings have occurred. Thresholded TTMetaD (ThTTMetaD) replaces $V^*(t)$ with $\max(V^*(t) - V_{th}, 0)$ everywhere it appears, where V_{th} is an adjustable parameter. It works to increase the robustness of the method, because decreasing the hill height too early is generally more costly than waiting slightly later to enter the convergence phase. However, as we show later it

was inessential in our tests; we have included it here primarily for the sake of completeness. Note that ThTTMetaD could also be used as a method for exploring a single basin up to a roughly specified free energy level—one can choose two points in the same basin as the endpoints of the transition and then choose the threshold bias value to match the desired free energy level.

The optimal choices of ΔT and V_{th} are no longer directly tied to the magnitude of the energy barrier and are quite robust, as we will show below, but they may still be fine-tuned for each system. Because they control convergence rate, it is intuitive that they should be matched to the natural sampling time scales of the method after bias has filled the main basins. Each energy parameter can be converted to a time scale using the hill energy addition rate $h/\tau_{revisit}$, where $\tau_{revisit}$ is the characteristic time between adding hills to the highest barrier point between the basins, to convert from energy to time. One should choose $h\Delta T/\tau_{revisit}$ and optionally $hV_{th}/\tau_{revisit}$ so that they are comparable to the longest important relaxation time scales in the almost-converged biased system. If it is chosen to be longer, then the method will not converge as quickly as it could. If it is chosen shorter, then the method will appear to converge too early, but will in fact be biased by the initial conditions and the bias may take a long time to converge to zero. These time scales will typically be related to physics not included in the CVs, the so-called orthogonal variable relaxations. Though this may seem harder to discover than the barrier energy for WTMetaD, in fact fine-tuning WTMetaD must also take this into account, so there is no disadvantage here—we have simply made the difficulty of fine-tuning more apparent because the coarse-tuning is performed automatically.

Next, like WTMetaD, TTMetaD can also be used with sophisticated hill types; see adaptive Gaussian metadynamics,[49] boundary-consistent metadynamics,[224] or field-coordinate metadynamics[326] for examples. Whatever the case, the metadynamics rule without tempering is to deposit a bias hill with the functional form $G(s, s(t_n))$ when point $s(t_n)$ is sampled at time t_n. With these defined, the final iterative TTMetaD rule with generic hills

and thresholding is

$$V(s, t_{n+1}) = V(s, t_n) + e^{-\max(V^*(t_n)-V^{th},0)/\Delta T} G(s, s(t_{n+1})). \tag{8.2}$$

With Gaussian hills of width δ and height h, this is

$$V(s, t_{n+1}) = V(s, t_n) + he^{-\max(V^*(t_n)-V^{th},0)/\Delta T} e^{-||s-s(t_{n+1})||^2/2\delta^2}. \tag{8.3}$$

Properly tempered metadynamics methods converge to biases V_c for which the expected bias update is flat. In TTMetaD, those biases satisfy

$$\int_S e^{-\max(V^*-V^{th},0)/\Delta T} G(s, s(t_n)) e^{-(F(s(t_n))+V_c(s(t_n))/T} ds(t_n) = C \tag{8.4}$$

$$\int_S G(s, s(t_n)) e^{-(F(s(t_n))+V_c(s(t_n))/T} ds(t_n) = C' \tag{8.5}$$

where $F(s)$ is the true CV PMF. For standard hills, such as Gaussians on a periodic domain or boundary-corrected Gaussians on a finite domain, requiring $\int_S G(s, s_n) \exp(-(F(s_n) + V_c(s_n)/T) ds_n$ to be flat requires that $\exp(-(F(s_n) + V_c(s_n)/T)$ must also be flat. Therefore in these normal cases the final converged bias for the method is $V_c(s) = -F(s)$. This is exactly the target of the original metadynamics, making the bias easy to interpret and fully compatible with existing metadynamics postprocessing tools and workflows.

We implemented this method in PLUMED[47] and PLUMED2, cross-platform enhanced sampling plug-ins for running enhanced sampling calculations in a number of molecular dynamics packages including Gromacs,[41] LAMMPS,[265] and NAMD,[260] among others. Our PLUMED code is freely available on request, and the PLUMED2 code can be downloaded from a public github fork of the PLUMED2 project. The implementation required minimal changes mostly consisting of added algorithms for calculation of the V^* and added code to set hill height based on $e^{-V^*(t)/\Delta T}$ rather than $e^{-V(s(t))/\Delta T}$. We calculate V^* to a close approximation by using Dijkstra's algorithm[102] to find paths of minimum cost $s^*(\lambda)$

between every pair of basins on a finely-sampled gridding of the bias. The basins are specified as specific points picked to be within the basins based on foreknowledge of the system, e.g., an *apo* crystal structure and a *holo* crystal structure for a ligand-binding protein. The path cost function used in the Dijkstra's algorithm computation is the negative of the minimum bias along the path, and thus negative of the maximum final path cost can be used as an estimate of $V^*(t)$. More sophisticated path finding algorithms would offer superior performance, but in our test cases we found that even this path finding calculation is so fast compared to force-field evaluation that the performance cost is negligible in practice. Specific numbers are provided with the test results below where applicable.

8.3 Results and Discussion

The TTMetaD method is designed to improve upon well-tempered metadynamics in three areas. It is meant to be more robust to changes in parameters, to offer superior asymptotic convergence rates, and to explore phase space more efficiently at early times. We performed three sets of tests to verify that these goals were met. First, we studied the behavior of TTMetaD and WTMetaD on an ensemble of 1D systems to compare robustness, convergence, and fast barrier crossing in an idealized setting where many replicates can be performed with analytic target results. Second, we studied the two methods applied to alanine dipeptide to compare convergence and barrier crossing in a simple real setting where the PMF can nonetheless be calculated accurately. Finally, we compared barrier crossing in the two in a practical setting, histidine gating in myoglobin. We also compared to metadynamics without tempering (untempered metadynamics, UTMetaD) to verify that TTMetaD is as fast to drive barrier crossing as UTMetaD and confirm that it avoids the tradeoff between fast initial exploration and efficient convergence. We did not attempt to tune TTMetaD in these comparisons; our purpose in these initial tests is to show that untuned TTMetaD can perform as well or better than WTMetaD using rough heuristics, and the question of how far TTMetaD can be pushed is a subject for future work.

8.3.1 Random 1D surfaces

Approximating real system dynamics as overdamped Brownian motion along the CV coordinates makes an excellent heuristic for understanding metadynamics theoretically. Accordingly, Brownian dynamics in a few dimensions on an analytically defined potential surface makes a good first test case for new metadynamics methods before applying them to more realistic systems. We chose to begin our tests with one dimensional systems for ease of prototyping; finding a barrier in 1D is a matter of taking a minimum over a predefined range. These tests investigated the speed, accuracy, and robustness of TTMetaD as compared to WTMetaD.

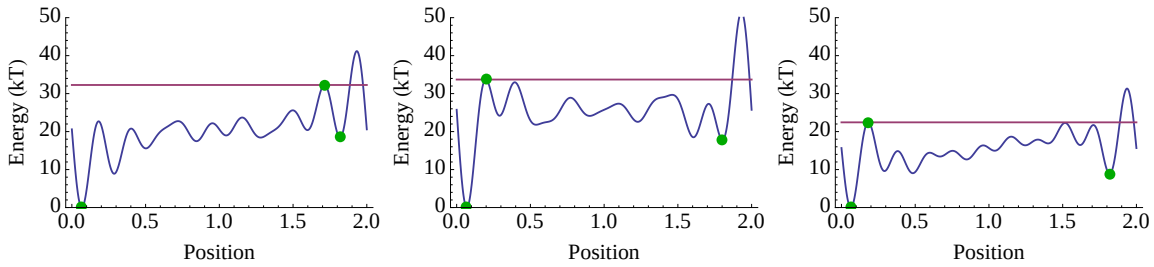

Figure 8.1: Three characteristic examples of potentials in the test set of 48 used to compare TTMetaD and WTMetaD in this section. Blue lines denote the potential surface, while green dots indicate leftmost basin from which dynamics was initialized, rightmost basin used as the second target basin, and the highest barrier between the two, and the red lines mark the barrier energy level. The test set includes potentials with early barriers, late barriers, and both early and late barriers.

To test the robustness, it is necessary to test accuracy and speed on many different potential surfaces with a variety of characteristic barrier heights and barrier structures. For the tests reported here, we chose 48 potentials of the form

$$F(x) = c_0 + \sum_{n=1}^{10} c_n \sin(n\pi x) \tag{8.6}$$

for x confined to the interval $[0, 2]$ and standardized so that the minimum over the accessible interval is zero using $c_0 = \min_{x\in[0,2]} \sum_{n=1}^{10} c_n \sin(n\pi x)$, with the c_n for $n \geq 1$ each chosen uniformly randomly and independently from $[0, 5k_BT]$. The coefficients as well as positions

for leftmost and rightmost basins and the barrier positions for each surface can be found in our supporting information (SI), and a representative set is illustrated in Figure 8.1. The dynamics of each system was Euler-Murayama integrated Brownian motion with a diffusion constant of 1 and a timestep of 0.0001. Each surface was simulated for 32 time units in 16 independent runs each initialized from the bottom of the leftmost and deepest basin in the potential in each case shown here. The shallower rightmost basin was taken as the second primary basin. we also performed rightmost-initialized, leftmost-secondary simulations with similar results that we have chosen not to present because they add little further insight and confirm the same essential points.

Each system's metadynamics used Mcgovern-de Pablo boundary-corrected Gaussian hills[224] with a height of 0.05 and a width of 0.025 deposited every 20 timesteps. The relative performance of WTMetaD and TTMetaD was not substantially affected by these choices. The TTMetaD runs used $\Delta T = 2k_BT$ while three sets of WTMetaD runs used $\Delta T = 8k_BT$, $\Delta T = 16k_BT$, $\Delta T = 32k_BT$ respectively. The choice for TTMetaD made little difference between k_BT and $6k_BT$; the choices for WTMetaD reflect a good choice for optimal convergence rate for small-barrier systems, $8k_BT$, optimal convergence rate overall, $16k_BT$, and a choice more tilted towards driving barrier crossing quickly, $32k_BT$.

The results of these simulations confirm that TTMetaD is more robust, more accurate, and faster to drive barrier crossings than WTMetaD. Figure 8.2 shows that the time to first barrier crossing is lower in TTMetaD whether the WTMetaD ΔT parameter is chosen properly for average convergence, $16k_BT$, or chosen relatively high for fast barrier crossing, $32k_BT$. This confirms that it drives barrier crossing faster than WTMetaD systematically, robustly with respect to multi-well barrier structure. Figure 8.3 shows that TTMetaD is systematically more precise than WTMetaD. In each figure, the spread of the TTMetaD results is narrower than the spreads for the WTMetaD results, confirming that the performance of TTMetaD is more robust to variance in the shape of the underlying potentials than WTMetaD. This is especially true for the low ΔT WTMetaD runs because, as is known in

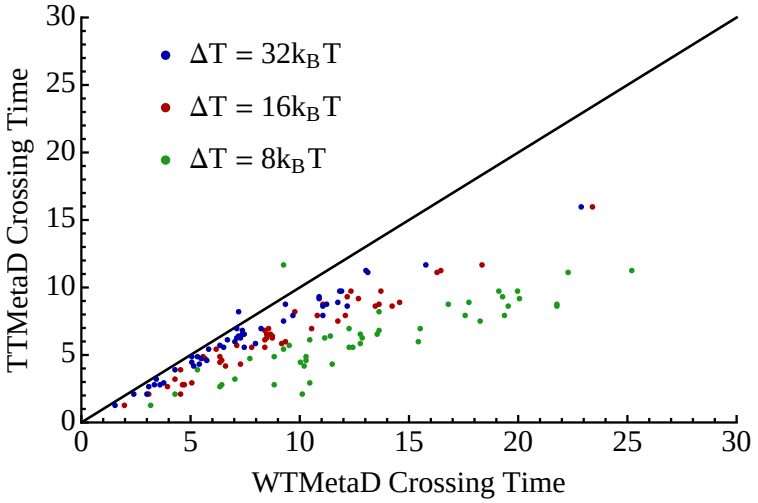

Figure 8.2: Comparison of first crossing times for TTMetaD with $\Delta T = 2k_BT$ and WT-MetaD with a variety of ΔT parameters. With two exceptions in 144 comparisons, TTMetaD drives barrier crossing faster. On average the speed up increases as the first crossing time increases.

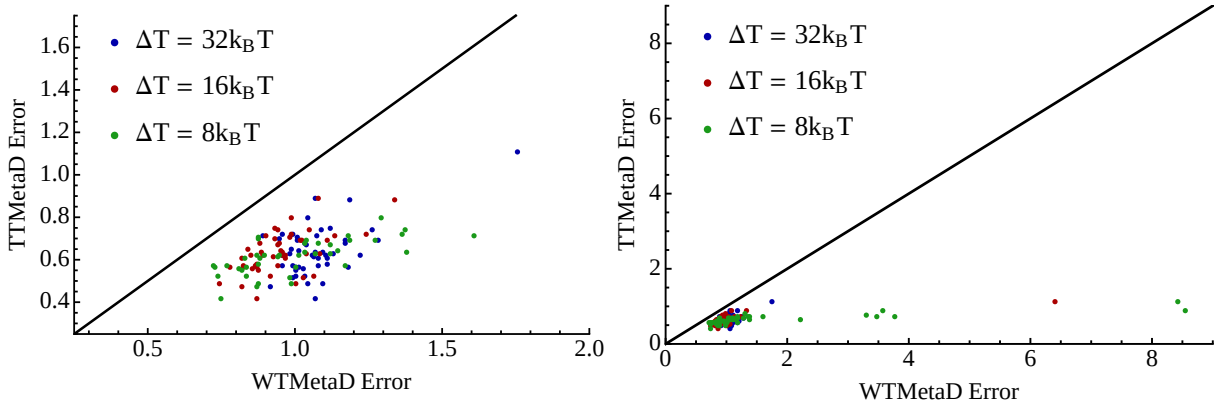

Figure 8.3: Comparison of accuracy for TTMetaD with $\Delta T = 2k_BT$ and WTMetaD with a variety of ΔT parameters. Error is measured as the maximum absolute value of the difference of the estimated PMF from the true potential in the between-basin region after both have been standardized to match in average potential value between the basins. The left panel shows a zoom on the main cluster of results and does not include all data points, while the right shows all results. With no exceptions in 144 cases, TTMetaD is more accurate than WTMetaD. Furthermore, while the spread of the WTMetaD accuracies decreases with higher ΔT, the spread of the TTMetaD accuracies is least of all.

the literature, choosing ΔT higher than necessary makes for more robust convergence, but less accurate convergence when that robustness is not needed. TTMetaD all but eliminates

that tradeoff.

8.3.2 Alanine Dipeptide

Though overdamped dynamics makes a good heuristic guide, it is also necessary to test on more realistic systems. Alanine dipeptide is a common first test system for new enhanced sampling algorithms that exhibits many of the essential complexities common to more challenging biomolecular systems: high dimensionality, adequate but suboptimal choice of CVs, and imperfect separation of time scales between CVs and other variables. These tests cannot demonstrate robustness as directly as the previous tests did, as only one system is simulated. Instead, they are designed primarily to show that TTMetaD remains more precise and faster to drive barrier crossing than WTMetaD in more realistic systems. Alanine dipeptide is a standard first test for the accuracy of metadynamics methods,[27] and will be our primary test of accuracy. The close relation of TTMetaD to WTMetaD and UTMetaD and previous experience with UTMetaD and WTMetaD then suggests that the relative accuracies of TTMetaD, WTMetaD, and UTMetaD on alanine dipeptide will be characteristic of those relative accuracies in more complex systems.

In these tests we simulated biased blocked alanine dipeptide *in vacuo*, modeled by the CHARMM27 force field[208, 209] and held at a constant temperature of 300 K using a Langevin thermostat with a drag of 5 amu/ps. The dynamics were simulated in Gromacs 4.6.1[41, 194, 330, 144] using a stochastic dynamics leapfrog integrator with a timestep of 2 fs, (superfluous) particle-mesh Ewald summation,[116] and SHAKE constraints on all bonds.[288] We performed 60 simulations for each metadynamics method presented. All simulations began from the same initial C_{7eq} conformation and the same pseudorandom Langevin force seed, but each set of 60 runs began from 60 distinct initial velocities pseudorandomly generated from different random seeds.

We compared UTMetaD, WTMetaD, TTMetaD, and ThTTMetaD using hill height, width, and rate parameters and WTMetaD tempering parameters tuned for WTMetaD in

earlier work. The height was 1.2 kJ/mol, the width 0.35 radians, hills were deposited every 120 fs, and the well-tempering parameter ΔT of 4 $k_B T$. For TTMetaD, we simply used a ΔT of 2 $k_B T$ and for ThTTMetaD a ΔT of 2 $k_B T$ and a threshold of 1 $k_B T$. The CVs were the ϕ and ψ angles, and the transition wells for TTMetaD were $(-1.25, 1.25)$ and $(1.0, -1.25)$ in CV coordinates: a rough, unoptimized choice of transition wells.

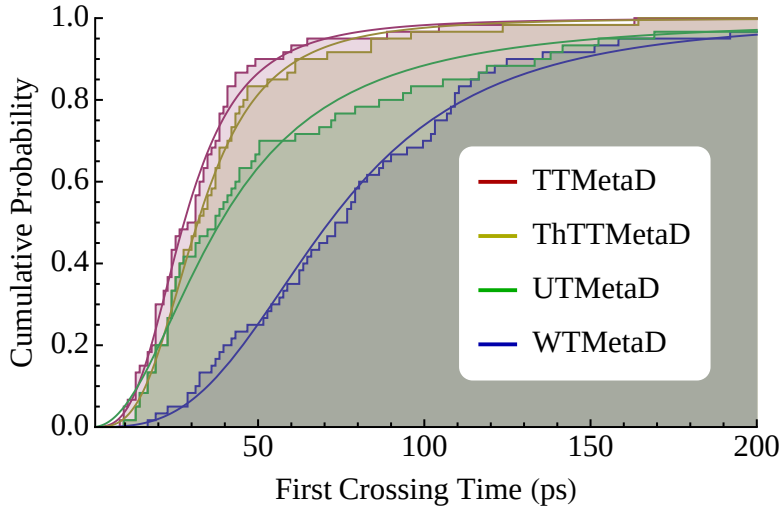

Figure 8.4: Cumulative distribution functions of first crossing times for blocked alanine dipeptide simulated using TTMetaD with $\Delta T = 2k_B T$, ThTTMetaD with $\Delta T = 2k_B T$ and $V_{th} = 1k_B T$, WTMetaD with $\Delta T = 4k_B T$, and metadynamics without tempering (UTMetaD), overlaid by maximum likelihood best fit log-logistic cumulative distribution function curves to guide the eye. The time to hitting the second basin is defined to the precision of the hill width, .35 radians, and the hill addition rate, 120 fs, for consistency with the underlying metadynamics algorithm.

The results of these simulations confirm that TTMetad remains more precise and faster to drive barrier crossing than WTMetaD when applied to more realistic systems. Figure 8.4 shows the cumulative distribution function for first barrier crossing times for the various methods and parameters tested. These times closely follow log-logistic distributions for each tempered metadynamics method, as shown by the maximum likelihood best fits superimposed on each of the empirical distribution functions. The fits have two parameters each, a shape parameter that is the same to within error for all methods and a time scale

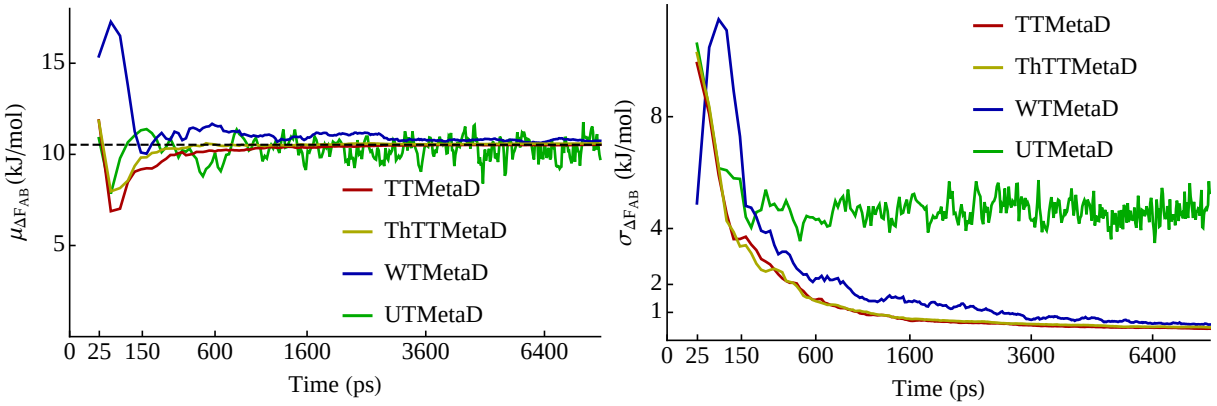

Figure 8.5: Time series of the means and standard deviations of the between-basin free energy difference estimates for the simulations of blocked alanine dipeptide using TTMetaD with $\Delta T = 2k_BT$, ThTTMetaD with $\Delta T = 2k_BT$ and $V_{th} = 1k_BT$, WTMetaD with $\Delta T = 4k_BT$, and metadynamics without tempering (UTMetaD); dashed line in left panel shows target result. Times are plotted on a square root scale to emphasize initial transients. Though TTMetaD and ThTTMetaD are roughly equally precise at all times, ThTTMetaD is more true earlier. WTMetaD is systematically less true and less precise and has higher confounding autocorrelation; UTMetaD does not converge.

parameter that is X for all the TTMetaD tests and Y for optimized WTMetaD, indicating that TTMetaD is 2.5 times as fast to drive barrier crossings than WTMetaD in this test case. Surprisingly, we find that TTMetaD is actually faster to drive barrier crossing than UTMetaD, not just as fast, especially in the tail of the distribution. Looking at the high variability of the bias in Figure 8.5, it seems possible that this may be due to excitation of orthogonal variables during relaxation after depositing large hills in high-energy regions. An orthogonal variable excitation could perturb the effective short-time mean forces and therefore perturb the effective barrier between wells, possibly preventing exploration of the target minimum. However, a full exploration of this surprising observation is beyond the scope of this work; our primary hypothesis, that TTMetaD is not slower to drive barrier crossing than UTMetaD, is confirmed by this test.

Figure 8.5 shows the means and standard deviations of the distributions of between-basin free energy estimates for each of the methods. The final standard deviation of the TTMetaD estimates is 0.41±0.05 kJ/mol, while the standard deviation of the WTMetaD estimates is

0.6±0.1 kJ/mol, indicating that TTMetaD offers an approximately 30 percent improvement in precision after 8 ns; the improvement at earlier times is comparable. The trueness of the TTMetaD estimates, especially the ThTTMetaD estimates, is also better than that for the WTMetaD estimates after 500 ps when the precision becomes roughly equal to k_BT, but precision rather than trueness is always the limiting factor on accuracy after this point. ThTTMetaD trueness is superior compared to TTMetaD because it shows less of a problem with undershooting the barrier after the first overshoot, which corresponds to the idea that the threshold allows it to drive recrossing more reliably. However, this does not appear to be a major effect in this system. Finally, examining the curves in Figure 8.5 shows that the WTMetaD data show coarser, more visible noise, indicating higher autocorrelation in the basin-basin free energy difference estimates. Such autocorrelation can confound free energy estimation, so a less-correlated estimator is preferable. The difference in autocorrelation may arise from the fact that TTMetaD fully flattens the free energy surface, whereas WTMetaD only partially flattens it. The relative autocorrelations of the estimators should reflect the relative autocorrelations of the biased peptide dynamics.

8.3.3 Myoglobin

Finally, we compared the methods in tests on a small protein, myoglobin, because blocked alanine dipeptide does not exhibit the full dynamical richness of larger biomolecules and methods validated on alanine dipeptide do not always transfer well to more complex systems. Our target for the test was a rotameric transition in histidine 64 that has been discussed as gating the diffusion of gas molecules into and out of the heme binding pocket. As with the blocked alanine dipeptide simulations, these tests do not demonstrate robustness directly, as only one system is simulated. These tests are specifically designed to show that TTMetaD is faster to drive barrier crossing than WTMetaD, even more in complex systems than in the dipeptide, and to check the cost of pathfinding in a complex calculation. They use unoptimized CVs for which we are not interested in a PMF, so we do not run the tests to

convergence.

In these tests we simulated oxygenated, heme-bound biased myoglobin in water, with protein governed by the CHARMM27 force field[208, 209] and water by the TIP3P model,[167] held at a constant temperature of 300 K using a stochastic velocity rescaling thermostat[55] with a time constant of 5 ps. The dynamics were simulated in Gromacs 4.6.1[41, 194, 330, 144] using a leapfrog integrator with a timestep of 2 fs, particle-mesh Ewald summation,[116] and LINCS constraints on all bonds.[229, 143] We performed 24 simulations for each metadynamics method and parameter set presented. All simulations began from the same initial conformation and the same pseudorandom stochastic velocity rescaling force seed, but each set of 24 runs began from 24 distinct initial velocities pseudorandomly generated from different random seeds. The initial conformations were generated from the PDB structure 1MBO,[261] using Gromacs built-in tools to add hydrogens and 9055 solvent waters and to minimize the energy of the system after adding hydrogens and solvent.

In this test we compared only WTMetaD and TTMetaD. The CVs were the χ_1 and χ_2 angles of histidine 64. We chose to simulate this using similar parameters as for alanine dipeptide, as the CVs are similar and the transition wells for TTMetaD were the same as before, $(-1.25, 1.25)$ and $(1.0, -1.25)$ in CV coordinates. A single WTMetaD run of 20 ns validated the choice of transition wells before running the TTMetaD. However, because the protein is a more sensitive dynamical environment, we made the initial rate of bias addition an order of magnitude slower. The hill heights were 0.3 kJ/mol, the widths 0.35 radians, hills were deposited every 300 fs, and the well-tempering ΔT was 4 $k_B T$. For TTMetaD, we once again used a ΔT of 2 $k_B T$.

Once again we found that TTMetaD is systematically faster to drive transitions than WTMetaD, and is in fact no slower than UTMetaD, the best-case result, as shown in Figure 8.6. In addition, in these tests we can compare running times meaningfully; here the performance measured in ns/day unexpectedly increased by approximately 1% from WTMetaD and UTMetaD to TTMetaD, but more expectedly the difference does not apear significant.

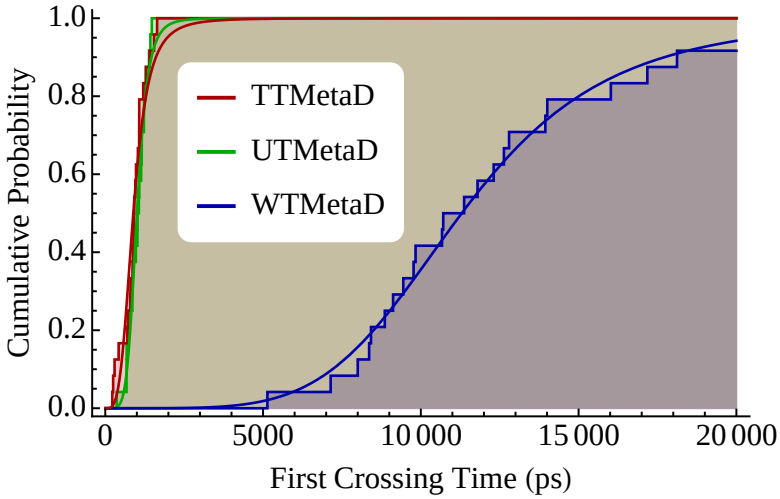

Figure 8.6: Cumulative distribution functions of first crossing times for simulations of myoglobin using TTMetaD with $\Delta T = 2k_BT$, WTMetaD with $\Delta T = 4k_BT$, and untempered metadynamics, overlaid by maximum likelihood best fit log-logistic cumulative distribution function curves to guide the eye. The time to hitting the second basin is defined to the precision of the hill width, .35 radians, and the hill addition rate, 300 fs, for consistency with the underlying metadynamics algorithm.

By design these tests cannot demonstratively confirm that any of the methods is more accurate or more precise than the others, but we note that there is no set of parameters for which the UTMetaD simulation would converge, so the fact that TTMetaD and UTMetaD is no slower for exploration than UTMetaD is strong support for its overall superiority for free energy estimation.

Figure 8.6 shows that in this case TTMetaD is an order of magnitude faster to drive a barrier crossing than WTMetaD. This likely indicates that the WTMetaD ΔT is too small. However, the fact that the TTMetaD ΔT proves more transferable is one of the advantages of the method, so the comparison is representative of practice. UTMetaD is a limit of TTMetaD and WTMetaD, so this indicates that TTMetaD's exploration rate is insensitive to ΔT over the range $[2kT, \infty]$ while WTMetaD's varies by an order of magnitude over the range $[4kT, \infty]$. WTMetaD is thus more difficult to tune correctly. Notably, the WTMetaD times remain log-logistically distributed, while TTMetaD and UTMetaD appear to give

rise to much narrower long-time tails, suggesting that TTMetaD may offer disproportionate improvements in reproducibility of barrier crossing times compared to its improvements in speed when barriers are higher. This advantage is especially important for preliminary exploratory research in which one would like to limit the amount simple randomness can confound one's investigation of the proper choice of CVs and parameters.

8.4 Conclusions

This chapter introduced a new form of metadynamics that uses an on-the-fly estimate of the free energy between basins (V^*) to tune the size of the hills in a way that ensures the basins are escaped as fast as possible relative to the original metadynamics without compromising on convergence rate relative to WTMetaD. In tests the method consistently improves upon WTMetaD in both speed and accuracy. Moreover, it is more robust and can be parameterized using more commonly available prior knowledge. Instead of requiring an estimate for the energy barrier, to first approximation it requires only the CV-space coordinates of one point in each of the basins separated by the barrier. These advantages would make TTMetaD a superior choice for studying transitions with known end states but an unknown mechanism, and especially superior in the case that the choice of CVs is not certain. In such a case it is essential to get a quick estimate of the biased barrier crossing mechanism to check the quality of the CVs. TTMetaD reveals the barrier-crossing mechanism for a given choice of CVs more quickly than WTMetaD would, and potentially more reproducibly if our findings for myoglobin generalize to other highly complex systems. These efficiency gains will provide the highest impact in situations where simulation is most expensive, for instance in *ab initio* molecular dynamics and QM/MM sampling.

The novel transition-based on-the-fly tuning strategy used in TTMetaD may also be translatable into strategies for improving other adaptive biasing methods such as Adaptive Biasing Force,[92] Wang-Landau,[341] or Orthogonal Space Random Walk.[369] We have in fact implemented a transition-tempered OSRW strategy with Michael Schnieder. It is closely

related in concept to the Wang-Landau flat histogram criterion strategy, which has previously been used with metadynamics,[228] but it operates in a more continuous manner and automatically flattens only an adaptively determined minimal metabasin region rather than a preselected CV region containing that minimal metabasin. Therefore it can offer computational savings and requires less foreknowledge of the system to apply. One interesting possibility would be to use V^* to trigger the second phase of flux-tempered metadynamics[302] after using TTMetaD during the first phase. A method based on that hybrid approach may converge yet more rapidly relative to WTMetaD than the current method. A second would be to use V^* to tune the nonequilibrium driving forces in driven metadynamics,[232] a method that may drive first barrier crossing events even more rapidly and increase convergence rate even further by decreasing state-to-state round trip time scales.

In developing TTMetaD we were careful to make sure that it retained the full flexibility of WTMetaD. TTMetaD can be used with adaptively sized hills,[49] field-coordinate CVs,[326] boundary-corrected hills,[224] driving forces,[232] and bias-exchange,[263] among others; few advances for WTMetaD cannot also be used with TTMetaD. In fact, we implemented it in PLUMED[47] and PLUMED2[324] to make it easy to use and familiar for current practitioners of metadynamics, and we implemented it in a modular way so that it can already be used with all of the enhancements described above that are included in PLUMED or PLUMED2. Our implementation code is freely available on request and is open for use and modification.

However, whatever the cost of an enhanced sampling method in raw computational resources, and whatever the potential efficiency gains, it is nearly always the case that the effort required to understand and apply it—the time and training of students—is in fact the limiting resource that must be economized most carefully in research efforts in this field. With that in mind, TTMetaD has been designed to be as easy to use as metadynamics, easier to use than WTMetaD, and compatible with existing metadynamics workflows to the maximum extent possible; with existing technology any lab already using metadynamics and PLUMED2 can begin using TTMetaD simply by reinstalling PLUMED2 with any compat-

ible MD engine, changing one line of an existing PLUMED2 input file, and running the corresponding simulation otherwise as normal.

CHAPTER 9

IMPROVING STABILITY: METABASIN METADYNAMICS

This paper substantially reproduces the material of Dama, J. F., Hocky, G. M., Sun, R., & Voth, G. A. (2015). Exploring Valleys without Climbing Every Peak: More Efficient and Forgiving Metabasin Metadynamics via Robust On-the-Fly Bias Domain Restriction. Journal of Chemical Theory and Computation, 11(12), 56385650. I am indebted to Glen and Rui for their help in testing my code and running many of the simulations; Glen set up and performed the actin simulations and Rui helped run the ClC simulations after I had begun them. Their input in the writing and analysis also cannot be ignored.

This section now describes a way to ensure stability for a metadynamics simulation against effectively irreversible dynamics, by restricting the domain on which bias is applied.

We said earlier that our imp would be foiled if his unfortunate bachelor were sufficiently strong-willed to resist dissatisfaction. Now, the bachelor himself is unlikely to be so strong-willed just from his reading and study, but the great danger is that he may find religion and there find salvation and the strength to resist the imp forever. It is vital to the imp that, however dissatisfied he can make his bachelor, he never encourage the man to explore the woods and mountains where goodhearted monks are secreted away. He therefore whispers against these holy hinterlands at the same time he whispers against the cities, ensuring that the bachelor loses all preferences between cities and locales in cities but never driving him out into the roughs where he might find his salvation. The imp must be very careful to ensure that his whispers are perfectly balanced, or the bachelor may be driven from the cities—or pulled to wander at exactly the edge of the hinterlands.

9.1 Motivation

Metadynamics works by using a choice of reduced coordinates called collective variables (CVs) to iteratively build a bias that increases the rates of transitions between metastable

energy wells; increased transition rates imply decreased sampling autocorrelation and thus improved PMF estimates. The degree to which a bias can actually promote those transitions, however, depends on how well the CVs capture the true reaction coordinates. When the CVs are imperfect the results may not approach the true PMF rapidly, and this is common enough that it is often considered to be the single most relevant limitation preventing application of metadynamics to the study of complex systems.[50, 325, 27, 313, 326] Furthermore, the cost of building the bias also depends on the complexity of the CVs, scaling with the volume of CV space–i.e., exponentially with CV number. All enhanced sampling methods that rely on CVs share these drawbacks to greater or lesser extents.[3, 199, 339] This chapter describes a new variant of metadynamics, metabasin metadynamics (MBMetaD), that I designed to suffer less from the use of poor quality CVs and to make more practical the use of larger numbers of CVs by judiciously restricting the bias's domain in CV space. However, in order to discuss the features of metadynamics that I wished to improve with the new method, I must first compare to an older alternative, window-based umbrella sampling.[321, 172]

Window-based umbrella sampling is stratified sampling applied to simulation.[321] It is accomplished by running many simulations with different energetic biases that keep each simulation restrained within a different small region, or window, of CV space. These windows are constructed so that the sampled distributions in the windows overlap with one another to cover all of the phase space of interest in a given investigation; this typically involves choosing 1) a scale for each stratified dimension of CV space to set the separation of window centers from one another and 2) a single energy scale to set the strengths of the restraint. Once these are chosen, one then runs simulations in each window, perhaps with some form of replica exchange among the biased walkers.[172, 3] After this, the sampling in all of the windows forms a patchwork covering of the CV space that can then be sewn together into a single overall PMF estimate using a method such as Gaussian process regression[308], the weighted histogram analysis method,[178] or the multistate Bennet acceptance ratio.[298]

Metadynamics, on the other hand, is an auxiliary distribution sampling approach. It

functions by iteratively building a bias away from previously visited points to accelerate escape from metastable basins and thereby decrease autocorrelation of sampling on the CV space. In tempered metadynamics,[28, 88] the type of metadynamics that converges asymptotically,[89] the bias is constructed by adding Gaussian hills of bias of progressively shrinking height at regular intervals. Overall, a typical case requires choosing 1) the length scale of the hill in each dimension, 2) a single hill energy scale, and 3) a single energy scale for the convergence rate of the bias. After these are chosen, one then runs one or more simulations under an iteratively growing bias, again with the possibility of replica exchange.[56, 263, 135] Once this is done, one can calculate PMFs via their direct connection to the bias in the asymptotic regime or via one of several nonequilibrium reweighting estimators.[46, 49, 320]

When stratifying based on a set of CVs is not enough to ensure good sampling over the CV range, one says that there are hidden slow variables (HSVs). When these are present, replica exchange can be inefficient and sampling in the windows often does not accurately represent true Boltzmann sampling on accessible simulation time scales.[172, 282] The free energy differences between windows spaced far apart are pieced together using information from windows between them, making the errors induced by poor sampling in a single window nonlocal. This nonlocality follows the topology of the PMF in a way such that sampling errors in transition regions have large effects on relative free energy estimates between the basins they connect, while *errors in basin regions and high energy regions have a more local effect.*[316] However, unless the HSVs prevent replica exchange or lead to obvious unphysical features in a PMF, the erroneous sampling can easily be mistaken for correct.

The presence of HSVs in metadynamics, on the other hand, typically causes sampling hysteresis that is clearly visible in the bias.[56, 27, 313] When the CVs are open for modification, this is an advantage because it gives a clear signal that the CVs need improvement. As in window-based umbrella sampling, the effect on PMF estimates is nonlocal, but unlike in that method, *the nonlocal effects from HSV sampling in high-energy regions can affect*

the estimates of the PMF across all other regions. Specifically, the nonequilibrium bias can drive undesired changes in HSVs that never relax on the time scale of simulation, for instance driving undesired boiling of solvent or an irreversible fluctuation of a protein conformation. That spoils the sampling, making it useless for prediction. Therefore, although hysteresis in the presence of HSVs is a desired behavior in regions of CV space relevant to the process of interest, it is an inconvenience in high-energy regions. It is especially troublesome in complex systems with sensitive dynamical environments and many potential HSVs, e.g., almost all proteins.

Regardless of the presence of HSVs, the simplest form of windowed umbrella sampling scales poorly with the dimensionality of CV space because one typically covers the space with the fixed-volume windows, and the cost of covering space with fixed-volume sets scales exponentially with dimension. This can be ameliorated somewhat by taking a metadynamics-like approach in which one runs simulations only in windows in low-energy regions encompassing the main basins of interest, i.e., a metabasin, gradually learning the shape of the metabasin along the way.[357] The metabasin manifold has a lower volume—and often a lower effective dimensionality—than its bounding box, so covering it with windows is less prohibitive than covering the full span of the CV space. No advance knowledge of the metabasin's shape is required, only an energy scale of interest.

This chapter describes a way to emulate this approach for metadynamics. However, the means and ends differ. The primary goal of MBMetaD is to prevent driving in high-energy regions to prevent the excitation of HSVs that irreversibly alter sampling. Greater efficiency with many CVs is another benefit and an important one, but it is currently a less pressing issue in the application of metadynamics than the problem of irreversible driving. The strategy differs from adaptive umbrella sampling because metadynamics already acts to fill metabasins from the lowest point out—an initialization logic—and our change is therefore rather to cause metadynamics to drive escape no further and converge once it has filled the metabasin—a termination logic.

The most common approach to this termination in the literature has been to self-limit by slowing down the addition of bias, for instance in self-healing umbrella sampling,[219] flat histogram metadynamics,[228] and well-tempered metadynamics.[28] However, though these can be effective, in such methods the time scales of updating the bias and the energy scale of the self-limiting are tightly coupled, which can make it impossible to simultaneously choose acceptable values for both in complex systems. Most often, one must simply choose to add energy very slowly or, similarly, decrease the rate of energy addition very rapidly to prevent spoiling the dynamics. In either case, this can make the simulations prohibitively expensive. There is therefore a need for a scheme that takes a different approach, sampling a specific energy scale regardless of how quickly bias is added, to empower researchers to make more practically efficient investigations of sensitive systems with metadynamics.

With MBMetaD, we solve the longstanding problem of designing metadynamics to flatten only a region of arbitrary shape, without restricting sampling to that region, in order to design a metadynamics that flattens only up to a chosen energy level and no further regardless of how quickly new bias is added. The approach we take is compatible with all tempering strategies,[28, 89, 88, 353] bias exchange,[263] multiple walkers,[275] and more,[302, 326, 232, 135] functioning as a modular enhancement that can and should be used together with other advances in metadynamics methodology. The task of filling a finite domain has been a pernicious challenge throughout the history of metadynamics.[85, 23, 224] When used without proper understanding it can suffer from boundary artifacts. Solutions are available for 1D intervals,[85, 23] but the first simple, implemented solutions for rectangular boxes with hard wall boundary conditions in any number of dimensions were first published by McGovern and de Pablo only two years ago.[224] We independently derived the same corrections in the course of studying the convergence of metadynamics and here we extend the McGovern-de Pablo approach to time-dependent domains of arbitrary shape with open boundaries. Our method consists of this generalized boundary-corrected metadynamics together with adaptive rules for defining time-dependent domains that represent the metabasins of genuine physical

interest in practical metadynamics applications. These rules do not require any *a priori* estimate of barrier heights in the system. Section 9.2 details this new method, Section 9.3 presents results and discussion of three example applications to biomolecular simulation, and Section 9.4 concludes the chapter.

9.2 Method Definition

Tempered metadynamics is proven to converge like a specific quasiequilibrium differential equation that depends only on the method and not on underlying system dynamics.[89] Therefore, to describe the design of the new approach that converges on a finite, open-boundary domain, we first define a new shape of hill that gives rise to a differential equation that flattens the energy on any domain. Second, because the domain of interest will only rarely be known ahead of time, we next discuss how to adaptively define these domains over the course of simulation using either a known target basin energy level or an adaptively learned energy level corresponding to the minimal metabasin containing several known points of interest.

In the original metadynamics,[181] the adaptive bias is constructed as a sum of Gaussian hills centered at previously visited points in CV space, i.e., in 1D,

$$V(s,t_n) = \sum_{i=0}^{n} h e^{-(s-s_{t_i})^2/2\sigma^2} \tag{9.1}$$

where h is an energy-valued hill height, σ is a CV-valued hill width, and the s_{t_i} are the sequence of points in CV space visited at each t_i. To ensure convergence one must sequentially decrease the hill height, for instance using the well-tempered metadynamics (WTMetaD)[28] rule

$$V(s,t_n) = \sum_{i=0}^{n} h e^{-(s-s_{t_i})^2/2\sigma^2} e^{-V(s_{t_i},t_{i-1})/\Delta T} \tag{9.2}$$

where ΔT is a parameter that controls the rate of height decrease. A general theoretical form for metadynamics that can be proven to converge[89] is

$$V(s,t_n) = \sum_{i=0}^{n} G(s,s_{t_i}) w[V(s,t_n)] e^{-V(s_{t_i},t_{i-1})/\Delta T} \tag{9.3}$$

where $G(s,s')$ is called the hill kernel and encapsulates the hill shape and $w[V(s,t_n)]$ is called the global tempering rule and serves as a complement or substitute to the WTMetaD rule for decreasing the hill sizes as time goes on; see for example transition-tempered metadynamics (TTMetaD)[88] or the global tempering used in experiment directed metadynamics.[353] The new method is therefore modular with respect to the various tempering rules and only involves redefinition of $G(s,s')$.

Specifically, consider the case that one knows of some hill kernel $G_0(s,s')$ that could flatten an entire CV space but would like to flatten a domain D of that CV space instead. Then the MBMetaD rule is to use a hill with two parts,

$$G_{MB}(s,s') = G_{Int}(s,s') + G_{Ext}(s,s') \tag{9.4}$$

where $G_{Int}(s,s')$ are hills that would flatten the interior of the domain if it had hard walls, and $G_{Ext}(s,s')$ raises the bias level of the exterior in such a way to exactly match the exterior bias level to the bias level of the domain boundary while also exactly counteracting the parts of $G_{Int}(s,s')$ that slope into the exterior of the domain. Both are based only on $G_0(s,s')$ and the domain D. These hills are illustrated in Figure 9.1.

For the first part, we use a close relative to the multiplicative McGovern-de Pablo[224] rule

$$G_{McGdP}(s,s') = G_0(s,s')/I(s;D,G_0) \tag{9.5}$$

where $I(s;D,G_0)$ is a boundary-normalized integral of $G_0(s,s')$ over D for its s' argument,

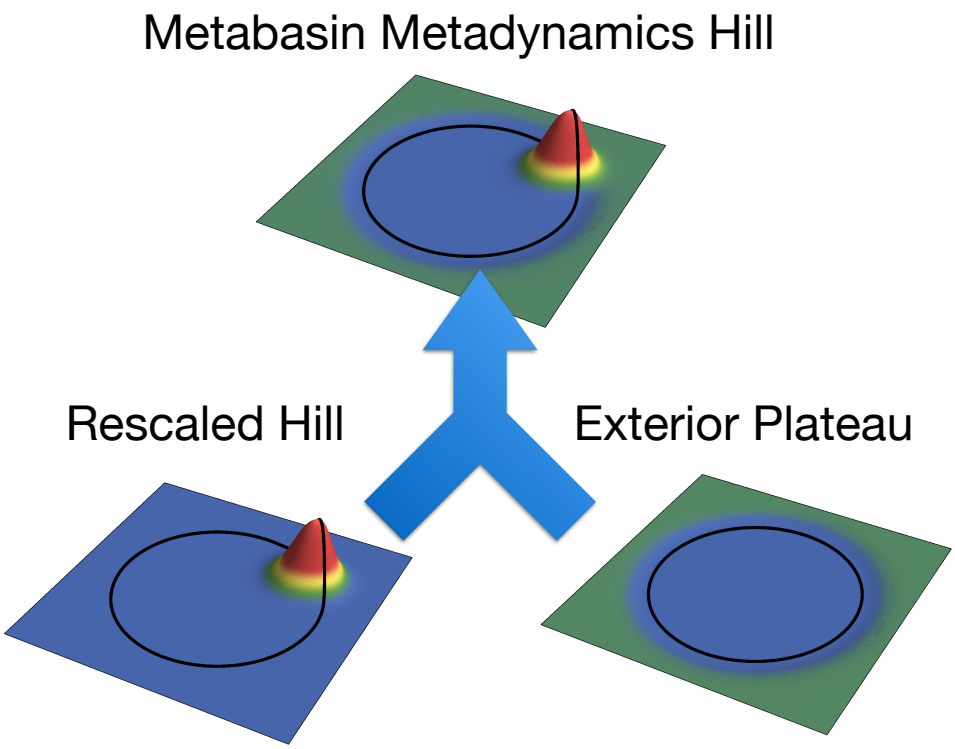

Figure 9.1: An illustration of a metabasin metadynamics hill function (top, Equation 9.4) on the unit circle near the domain boundary (black lines denote $r = 1$) and its decomposition into a rescaled hill (left, Equation 9.7) and a low plateau outside of the domain (right, Equation 9.9). The hill is based on a Gaussian of width $1/\sqrt{10}$ placed at $r = 0.9$. The effects of adding rescaled hill and exterior plateau contributions evenly over all points on the unit circle are exact complements, leading to bias updates that are flat everywhere when sampling is flat on the unit circle.

i.e.,

$$I(s; D, G_0) \equiv \int_D ds' G_0(s, s') \left(\frac{\int_{\delta D} ds \int_D ds' G_0(s, s')}{\int_{\delta D} ds} \right)^{-1} \tag{9.6}$$

with δD the boundary of domain D. This integral does not have a closed form solution for domains of general shape, but for Gaussian hills on a rectangular domain it is given by the sum of error functions provided in the paper of McGovern and de Pablo.[224] The rule (Equation 9.5) has the undesirable property that as s' approaches the boundary of D from the interior, the values of $G(s, s')$ in the exterior diverge to infinity, making it unsuitable for metadynamics on an open boundary domain. This occurs because $I(s; D, G_0)$ goes to zero

quickly as s goes beyond the boundary, so an expedient solution is to modify $I(s; D, G_0)$ such that it falls to a fixed value greater than zero. In MBMetaD, therefore, our form for the first term in Equation 9.4 is

$$G_{Int}(s, s') = G_0(s, s')/f(I(s; D, G_0)), \tag{9.7}$$

with $I(s; D, G_0)$ as above and in the implementation described here, $f(x)$ is the simple piecewise polynomial $f(x) = x$ for $x \geq 1$ and $f(x) = x + .5(1 - x)^2$ for $x < 1$. This choice of f is continuous and differentiable even at $x = 1$, ensuring that all bias forces will be continuous. $I(s; D, G_0)$ is constant in the interior of the domain and decays to zero outside of the domain, so dividing by $f(I(s; D, G_0))$ has the effect of leaving hills in the interior unchanged and causing them to bulge out of the domain subtly more than they otherwise would when close to its boundary.

Using only that first part of the hill, the differential equation from the convergence proof[89] reproduced in Chapter 7,

$$\frac{dV(s,t)}{dt} = r(t) \int ds' G(s, s') p_{wt}(s') \tag{9.8}$$

would predict that the interior of the region would be flattened, but the bias inside would grow indefinitely, forming a steep-walled, flat-topped mesa in the PMF. This would drive all sampling out of the metabasin of interest and is clearly undesirable. The second term in Equation 9.4 ensures that this does not occur. Specifically, we use

$$G_{Ext}(s, s') = (1 - I(s; D, G_0)/f(I(s; D, G_0)) \frac{\int_{\delta D} ds G_{Int}(s, s')}{\int_{\delta D} ds} \tag{9.9}$$

which plateaus in the exterior of the region with a plateau height equal to the boundary average value of the first part of the new bias. An example is illustrated in the bottom right of Figure 9.1. Because its height is set to match the bias added to the boundary, it ensures

that the exterior bias level always matches the average bias level of the domain boundary. In other words, this part of the bias update pulls sampling into the domain everywhere along the boundary exactly enough to cancel out how the first part of the hill would push sampling out of the domain close to the new hill. Moreover, the plateau is perfectly shaped so that flat quasiequilibrium sampling on the interior of the domain leads to an exactly flat increase of the bias everywhere in CV space. Finally, this term, Equation 9.9, is only nonzero when a hill is added close to the boundary (s' is close to the boundary), so MBMetaD functions just like normal metadynamics away from those boundaries.

While this scheme might appear technical at first glance, it has a simple multiscale physical interpretation. On a coarse scale, we look at the system through a two-state lens: either the sample is inside or outside the domain. On a fine-grained scale, we look at the points in CV space to a resolution comparable to the lengthscale σ of the original hill. The metabasin rule, then, corresponds to looking at the CV space using a topology in which all the points inside the domain are resolved to a fine-grained level, while all of the points outside are collapsed into one generic state. The first part of the hill is added to a fine-grained neighborhood of the interior, while the second is added to the entire coarse-grained exterior state—also a single neighborhood, but using the multiscale topology described above instead of the usual metric topology of the CV space.

It is often argued that bias hills should match the Green's functions for the dynamics, in at least a loose sense,[149, 57, 49, 354] and this is no exception. The simplest way to understand this new hill shape in those terms is to recognize that these hills are also like Green's functions, but where boundary conditions have been adjusted so that the walls are only partially reflective and any particles that exit the boundary from the interior must be balanced by particles crossing the boundary from the exterior. In the exterior, the constant profile in the far field corresponds to a well-mixed, fixed concentration boundary condition at infinity. The second part is thus related to the quasistationary distribution of a well-mixed domain exterior and therefore retains a natural connection to dynamics.

Up to now we have spoken in terms of a single connected domain, but the development above applies to any disconnected domain just as easily. Simply divide the domain D into its n components D_i and whenever a sample would be added inside a domain D_i, use the above definitions with D_i in the place of D. In the discussion of its physical meaning, consider a 2^n-state coarse-graining rather than a two-state coarse-graining; the new set of states is the Cartesian product of all the sets of two states (interior and exterior) per domain.

When one's goal is to flatten a PMF in a domain without causing discontinuities, the only domains that can be flattened completely are those with an isoenergetic boundary. Therefore it is natural to define valid domains by defining a boundary free energy and letting the domain be the set of all CV points with a lower free energy than that boundary energy. Since the zero of free energy is arbitrary, these definitions must be made in terms of an energy difference rather than an energy alone. For instance, in this work we showed the use of domains like "all points less than 45 kJ above the minimum free energy", a known basin free energy level, and "all points less than 15 kJ above the transition barrier between points A and B", an implicitly defined metabasin energy level. Of course, the choice of domains D_i is rarely possible to make before a simulation. It is often one of the things one would like to discover using metadynamics. For that reason, we also add an on the fly mechanism to discover D_i defined in terms of simpler criteria like the examples above.

We solve this problem similarly to as in TTMetaD,[88] where we used the bias as an on the fly free energy estimator. However, in this case we seek metabasins rather than barriers, and now the bias will not properly approximate the free energy anywhere outside of a previously defined domain. Therefore we use the nonequilibrium umbrella sampling estimator of Branduardi, Bussi, and Parrinello[49] to approximate the domain locations on the fly rather than the bias alone. This allows for the discovery of extensions of a domain outside a previously chosen domain, possibly including new disconnected components, regardless of initial conditions. The free energy estimates provide minima, transition barriers, and on the fly region selections as described in the next paragraph.

The recipes above have straightforward numerical implementations. $I(s; D, G_0)$ can be calculated by numerical integration on any domain D simply by summing many hills $G_0(s, s')$ evenly over D, and once it is available $f(I(s; D, G_0))$ and $I(s; D, G_0)/f(I(s; D, G_0))$ are simple to calculate. The nonequilibrium umbrella sampling estimator is evaluated using a running histogram accumulated during simulation alongside the bias, free energy minima are trivial to calculate, and the transition barriers are found via breadth-first path search as in TTMetaD.[88] These minima and barriers, plus an offset, define the target region energy. The region of interest is then the total set of grid points under that free energy in the current on the fly estimate. That set of grid points is then split into connected domains with a standard flood-fill connected components labeling algorithm.[83] Once these domains are defined, we compute $I(s; D, G_0)$ for each domain and then continue the simulation. Our implementation updates the domain at fixed time intervals with an option not to update the domain each time, but to instead also wait until the free energy estimate in the exterior of the region has changed by some set amount. The latter corresponds to a logarithmic update schedule that cuts down on unnecessary computational overhead at late times when the domain definition is stable. Each change in domain definition only affects future hill addition, so the previously deposited bias continues to be used unchanged.

Our implementation is a public fork of the PLUMED2 package.[324] It provides MBMetaD with a minimum based domain or a transition barrier based domain (with transition barriers defined as described for TTMetaD in Chapter 8), using only one additional required input number, the energy offset from the transition or minimum. Two further parameters to control the update schedule and one controlling accuracy of the numerical integration are also available for fine-tuning, but these have general defaults and can be considered optional.

9.3 Application and Discussion

The purpose of MBMetaD is to fundamentally change the tradeoffs of metadynamics with respect to other methods by giving it a self-limiting mechanism, and thus the remainder of

this chapter focuses on investigating the self-limiting behavior in three examples. First, we show that it functions as intended all the way through PMF convergence for a simple model biomolecule with imperfect CVs, alanine dipeptide, without negatively affecting convergence efficiency or accuracy. Second, we show that it correctly self-limits in a more complex example, monomeric actin, where the self-limiting is used to acheive more stable and reproducible exploration of conformational space. Finally, we apply it to a membrane transport protein, ClC-ec1, using mediocre CVs to examine its failure modes and in particular show that it retains metadynamics' attractive property of displaying hysteresis in the domain of interest (rather than harder to recognize errors) even as it prevents the nonequilibrium driving from spoiling the dynamics in other domains.

Our focus in these examples is the self-limiting behavior and reproducibility rather than convergence efficiency or accuracy; self-limiting behavior aimed at improving reproducibility is the sole novel feature introduced in MBMetaD and the method otherwise behaves much like whatever other metadynamics method it augments. Though the accuracy of metadynamics in practice does deserve further study, it is not a focus of this particular work.

9.3.1 Alanine Dipeptide

A standard first test case for new metadynamics methods is blocked alanine dipeptide *in vacuo* with ϕ and ψ dihedral CVs because the long literature on metadynamics provides excellent guidance on transferring knowledge gained with this model to understanding of the method in more complex systems.[181, 28, 49] Alanine dipeptide provides a simple case of a system with imperfect separation of time scales between CV dynamics and other relaxations, and is thus a minimal test for any adaptive enhanced sampling method that is meant to be robust to memory in the CV dynamics.

These tests modeled alanine dipeptide using the Amber03 force field[105] and at a constant temperature of 300 K using a Langevin thermostat with a drag of 5 amu/ps. The dynamics were integrated in Gromacs 4.6.1[41, 194, 330, 144] patched with a customized ver-

sion of PLUMED2[324] and using a stochastic dynamics leapfrog algorithm with a timestep of 2 fs, particle-mesh Ewald summation,[116] and SHAKE constraints on all bonds.[288] We performed 32 simulations for each method and parameter set presented. Each simulation began from the same C_{7eq} conformation, but each set of 32 runs began from 32 distinct initial velocities pseudorandomly generated from distinct random seeds and using 32 distinct Langevin thermostat seeds.

We compared WTMetaD, TTMetaD, and transition-tempered MBMetaD using hill height, width, and rate parameters and tempering parameters as in our previous work introducing TTMetaD.[88] The height was 1.2 kJ/mol, the width 0.35 radians, hills were deposited every 120 fs, the TTMetaD parameter ΔT was 2 $k_B T$, and the transition wells for TTMetaD were $(-1.25, 1.25)$ and $(1.0, -1.25)$ in (ϕ, ψ) CV space; for WTMetaD ΔT was 4 $k_B T$. When using metabasin metadynamics we compared the minimum-based domain definition with relative energy levels of 35.0 kJ/mol and 42.5 kJ/mol and transition-based domain definition with relative energy levels of 7.5 kJ/mol and 15.0 kJ/mol. We chose these values in order to roughly match the two methods based on the approximation that the apparent transition barrier relative to the minimum in these CVs is 27.5 kJ/mol; this makes for clear and direct comparison between the two different styles of domain specification. The higher energy values were chosen as a representative test of the method. The lower energy values were chosen to illustrate how the two domain definitions differ in behavior as the energy excess above the transition barrier becomes low and the PMF is barely flattened over the barrier between the transition wells.

The results for higher energy limits confirm that MBMetaD with either domain definition retains the accuracy and efficiency of TTMetaD in the low-energy regions it is designed to study while giving up all resolution of the high-energy regions it is designed not to explore. Figure 9.2 shows the replicate-averaged estimated PMFs of regular TTMetaD and the two MBMetaD variants after 8 ns of simulation for comparison; it is evident that the low-lying contours are essentially indistinguishable among the three methods while the high-energy

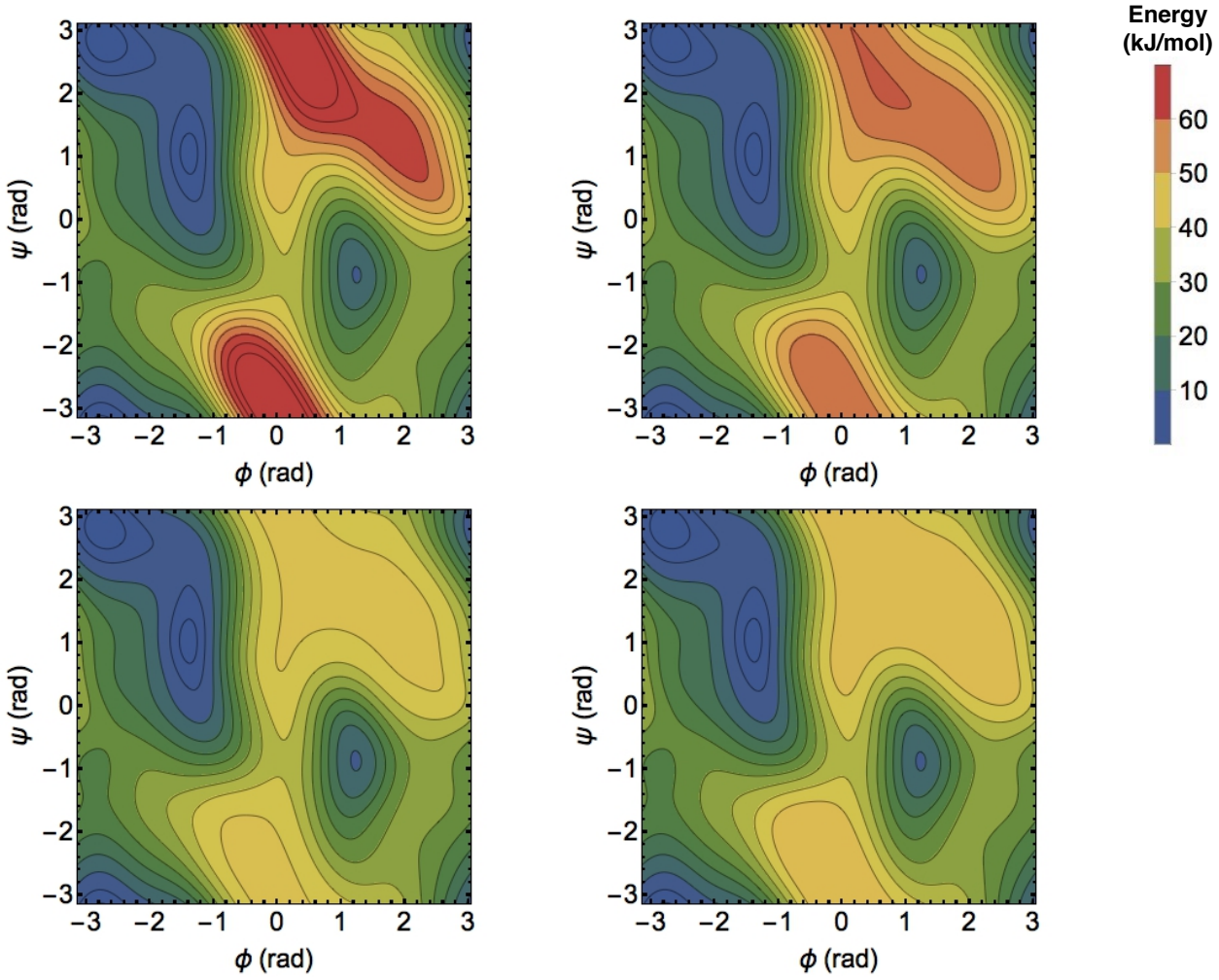

Figure 9.2: Bias-based free energy estimates averaged over 32 runs of 8 ns for four different metadynamics choices designed to give different levels of resolution of the high-energy regions. **Clockwise from top left:** WTMetaD, TTMetaD, 15 kJ/mol transition-referenced MBMetaD, 42.5 kJ/mol minimum-referenced MBMetaD. All methods use parameters as defined in the text. Contours are placed every 5 kJ/mol. The estimates are all but identical in low-energy regions despite stark differences in the high-energy regions. WTMetaD is least controllable and gives the most resolution of high energy regions, TTMetaD is more controllable but would give more and more resolution of high energy regions given additional convergence time, and both forms of MBMetaD are fully controllable, converging without resolving the high energy regions.

contours present in the first are simply missing in the latter two. Also as expected, the contours outside of the domain but close enough to be biased are somewhat random and are not optimized over the course of convergence. Figure 9.3 compares the convergence rates for these methods on the domain of interest in terms of trueness and precision.

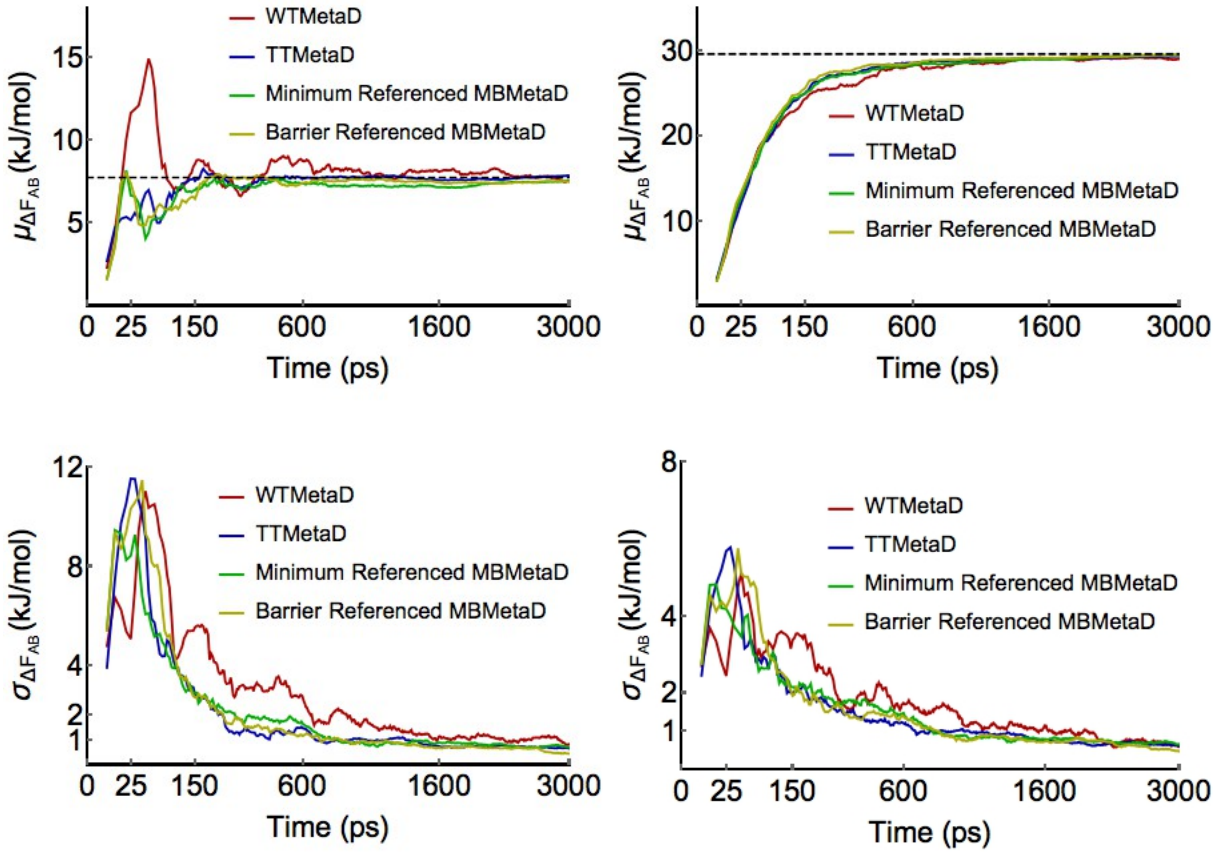

Figure 9.3: Convergence of MBMetaD bias compared to TTMetaD and WTMetaD. Figures on the left show convergence results for the estimated free energy difference between points (-1.25,1.0) and (1.25,-1.0), the two primary basins, while figures on the right show convergence results for the estimated free energy difference between points (-1.25,1.0) and (2.1, -1.5), the deeper basin and the lowest free energy barrier point between the two basins. Figures on top show the averaged free energy difference estimates as a function of time, while the lower figures show the standard deviation of the estimates as a function of time. All methods are as described in the text, with the statistics calculated over 32 runs per method. The MBMetaD methods in these figures used energy offsets of 42.5 kJ/mol (minimum referenced) and 15 kJ/mol (transition referenced). For these free energy difference estimates inside the metabasin domain, MBMetaD incurs no convergence accuracy penalty.

However, with lower energy limits, the transition-referenced domain definitions become clearly superior to minimum-referenced domains. This is shown in Figure 9.4. This is because in the normal case that CVs are imperfect, hidden barriers in HSVs can cause the bias to grow too high in one basin before exploring the next basin.[27] When this occurs in minimum-referenced MBMetaD, the energy differences from the minimum can therefore appear higher

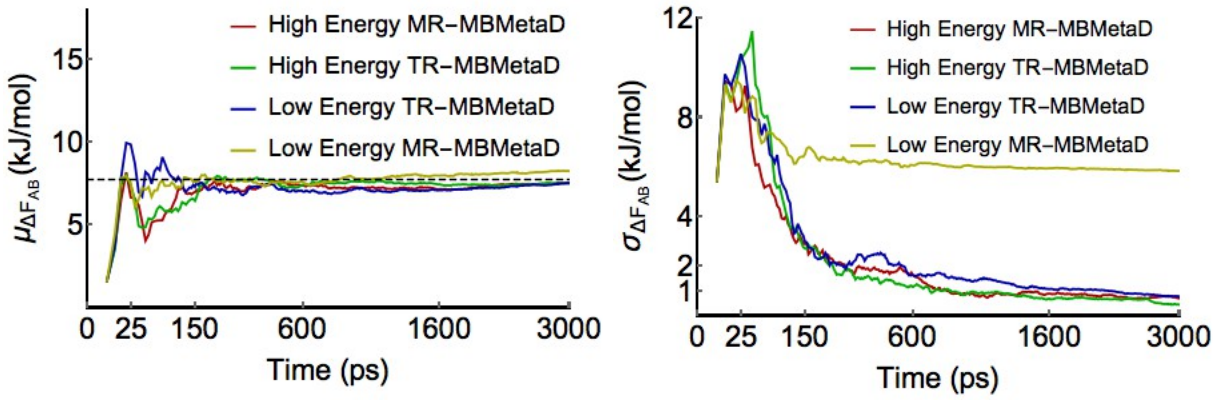

Figure 9.4: Comparison of the convergence of MBMetaD biases using different domain specifications. The MBMetad methods in these figures used energy offsets of 42.5 kJ/mol (high energy minimum referenced) and 35 kJ/mol (low energy minimum referenced), 15 kJ/mol (high energy transition referenced), and 7.5 kJ/mol (low energy transition referenced). The trueness of the MBMetaD methods are all fairly consistent, but the precision of transition-referenced MBMetaD is more robust with respect to choice of domain specification due to the early time cancellation of errors discussed in the text.

than they should, the estimated domains then become too small. If the intended domain was not much larger than the true metabasin of interest, the bias may not fully flatten all the way between the states of interest until the overestimation corrects itself. On the other hand, when similar overestimation occurs in transition-referenced MBMetaD, the transition barrier estimate also overshoots the true value, so that the domain is made larger by the same amount that the previous error would make it smaller. This systematic cancellation of errors ensures greater stability and better flattening before convergence, and thus better, less autocorrelated sampling and faster convergence. We always recommend using the transition-based rule for adaptive domain restriction when seeking to connect basins.

In these tests, the desired regions of interest are single large connected domains including both wells. In the high energy tests, the on the fly regions begin with large connected domains including both wells and small disconnected satellite domains due to sampling noise in the free energy estimate. As the simulation proceeds and the free energy estimate is refined, new satellites spring up, some disappear, and some connect with and absorb into the main

domain. The behavior is the same in the low energy transition-referenced simulation as well. However, in the low energy minimum-referenced simulations, the regions often begin with one primary domain around the basin the simulation is initialized in. In these runs, when a transition to a new well occurs there is a delay before the umbrella sampling estimator registers the discovery of a new low-energy region, but then a domain begins to grow around the new basin. Depending on vagaries of sampling, these domains can either remain separate or join together before the simulation terminates. This reconnection is not efficient, so once again, we always recommend using the transition-based rule for adaptive domain restriction when seeking to connect basins.

9.3.2 Actin Flattening

Actin is a key component of the cytoskeleton,[267] but the structural bases and mechanisms of its regulation are only partially understood, even at the monomer level. The monomer is also called globular actin or G-actin, and aspects of it have been studied using metadynamics[259, 223] and umbrella sampling.[290] However, G-actin is a dynamically sensitive protein with a high degree of allostery that can be difficult to design good CVs to investigate and can resist enhanced sampling analysis.[191]

A previous paper[290] used umbrella sampling to investigate nucleotide-dependent G-actin conformational distributions measured in terms of an interdomain distance and torsion (Figure 9.5) but was unable to reach convergence, presumably due to the presence of HSVs. Attempting to use metadynamics with the same torsion and distance CVs showed hysteresis, confirming the presence of HSVs, but this hysteresis typically involved unexpected and apparently irreversible behavior. Hypothesizing that the CVs were adequate in low-energy regions but inadequate in higher energy configurations, we applied MBMetaD to the problem in order to sample the lowest-lying metastable basins well and learn enough to refine the CVs before investigating further.

To do this, we performed 8 simulations of WTMetaD with and without adaptive domain

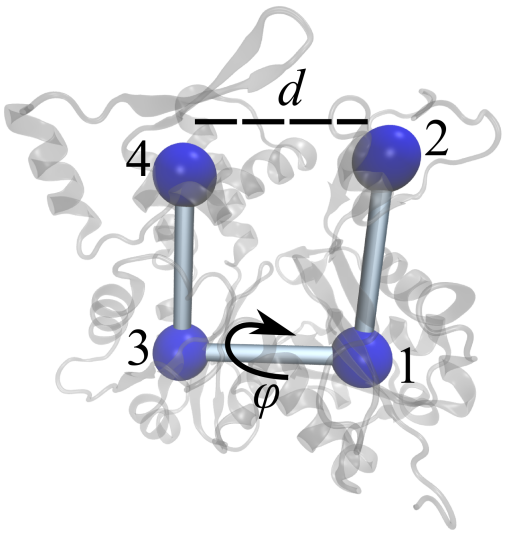

Figure 9.5: Schematic illustration of the CVs used to investigate nucleotide-dependent actin dynamics, consisting of the distance between subdomains 2 and 4 and the torsion angle between subdomains 2, 1, 3, and 4.

restriction for 50 ns each. The simulations were set up as in previous work,[290] with proteins simulated in NAMD[260] patched with a customized version of PLUMED2[324] using the CHARMM22/27 force field with CMAP.[209] Our CVs were similar to those in the previous work, consisting of the 2-1-3-4 torsion between the cores of the four actin subdomains—the flatness of the monomer—and the distance between subdomains 2 and 4—the width of the nucleotide binding cleft. However, for the sake of computational efficiency the positions of these subdomains were calculated as the centers of mass of only the backbone and C_β carbons in the subdomain cores rather than the centers of mass of all of the subdomain core atoms. All simulations were initialized from the flattened conformation like that of filamentous actin (PDB ID 2ZWH).[250] Metadynamics added hills of initial height 0.004184 kJ/mol and widths 0.3 radians and 0.02 nm every 200 fs, with heights sequentially adjusted according to the WTMetaD rule with a bias factor of 10; the MBMetaD domain was defined as every point less than 25.104 kJ/mol above the current estimate of the free energy minimum, and was updated every 200 ps if the exterior bias had increased by 2.092 kJ/mol since the last update. A minimum reference is appropriate because the simulations do not study a particular transition, but rather the fluctuations about a single basin.

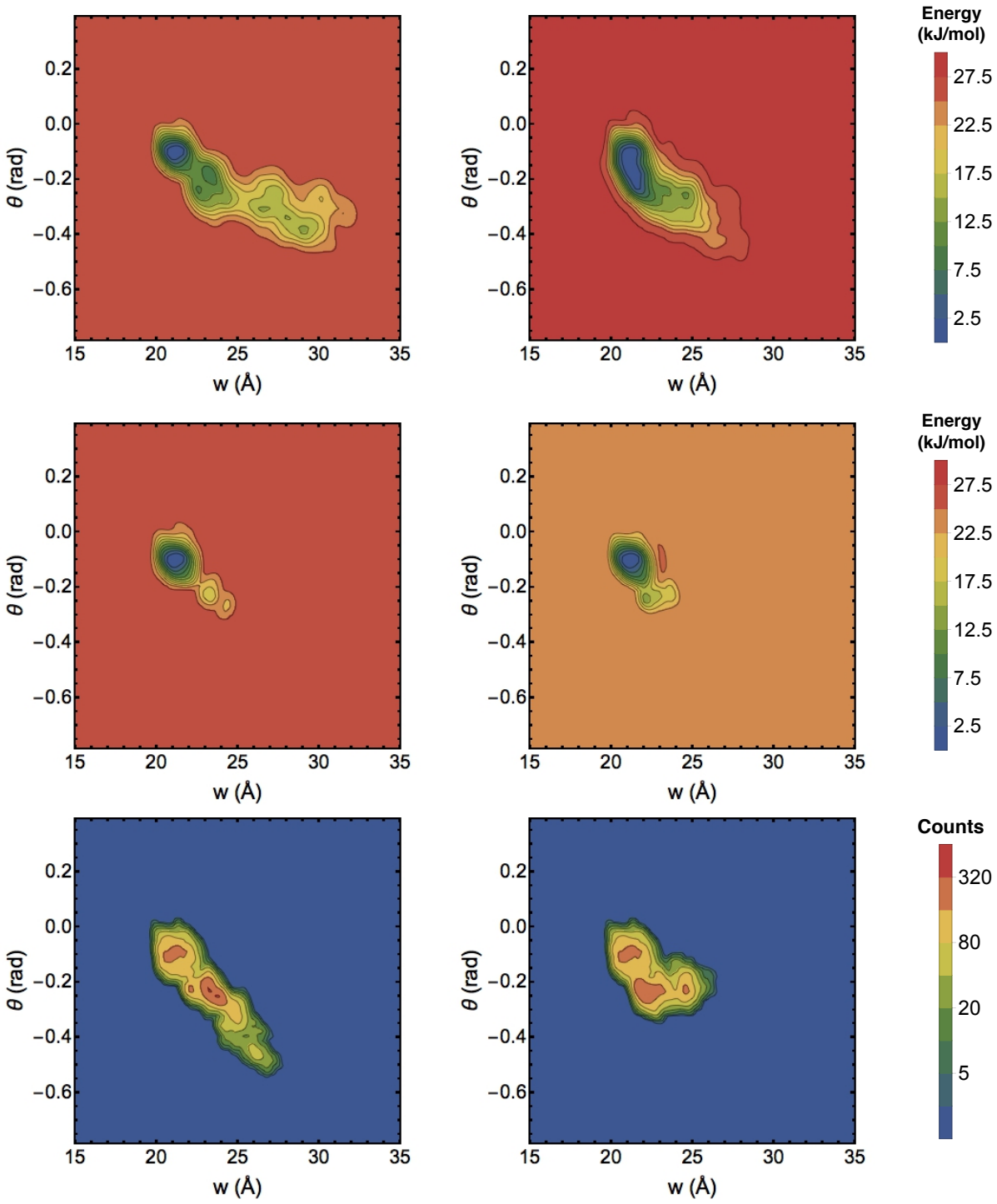

Figure 9.6: Representative sampling from unrestricted WTMetaD (top row) and well tempered MBMetaD (lower two rows) with the same initial conditions (columns) as a function of cleft width (w) and twist angle (θ). In the unrestricted case only bias PMF estimates are shown because sampling and bias are directly related; for MBMetaD both bias PMF estimates (middle row) and sampling histograms (bottom row) are shown.

Representative biases and sampling histograms at the end of 50 ns are shown in Figure 9.6. It is evident that the MBMetaD rule successfully limits the bias to a smaller metastable basin of CV space than the unrestricted WTMetaD does without artificially restricting the sampling to the same basin. The regions selected by MBMetaD correspond to the biased regions. They consist of single primary domains around the starting states together with smaller disconnected domains that spring up wherever the simulation occupies another state long enough. However, the primary purpose of this test is to demonstrate greater simulation repeatability for MBMetaD compared to the WTMetaD reference, investigated in Figure 9.7.

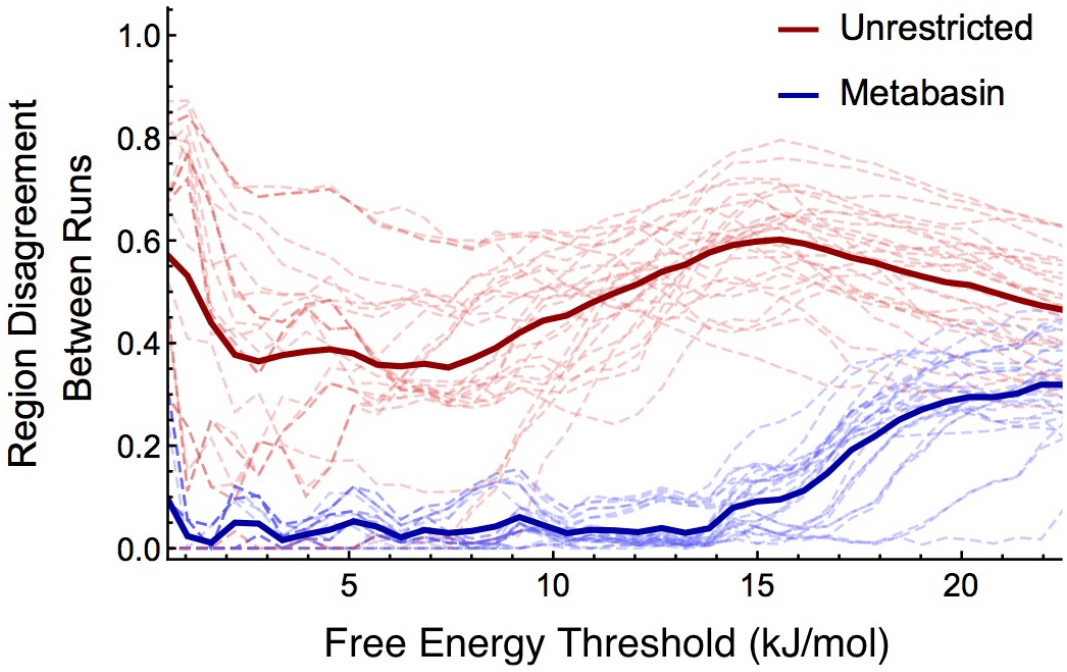

Figure 9.7: Figure showing the disagreement between the free energy estimates of different runs of like type. Disagreement (y axis) is measured as one minus the ratio of the area of the intersection of the two regions predicted by two different runs to be below a given free energy threshold (x axis) to the area of their union. All PMF estimates are standardized so that their minima are zero. Dashed lines represent the disagreement as a function of free energy threshold for individual pairs, while solid lines indicate their averages. The unrestricted WTMetaD estimates disagree with one another more than the MBMetaD estimates at all free energy levels, showing much improved repeatability for the latter.

This figure demonstrates that overlaps of the biases of different randomly initialized runs are greater in MBMetaD than WTMetaD, i.e., that the sampling is more similar across runs of MBMetaD than WTMetaD, just as desired. The disagreement between predictions of which regions of CV space are below a threshold energy shown in this figure correspond to one minus the ratio of the area of the intersection of the regions to the area of their union; expressed as an integral, this is $1 - \int \min(I(F_i(s) < F_{th}), I(F_j(s) < F_{th}))ds / \int \max(I(F_i(s) < F_{th}), I(F_j(s) < F_{th}))ds$, with F_i and F_j the PMF estimates from runs i and j, F_{th} the threshold energy, and $I(f(s) < g)$ the indicator of the function f being less than scalar g at point s.This indicates greater stability and more physically meaningful—or at least more easily interpreted—exploration. The nonreapeatable free energy basin predictions seen with unrestricted WTMetaD do not correspond to states that appear biologically relevant or well-populated at equilibrium, but rather appear to correspond to metastable traps found only because of overzealous nonequilibrium kicks. The MBMetaD data proved more useful for refining the choice of CVs and designing new simulations, but because refining CVs is a system specific process rather than a matter of MBMetaD methodology, describing that process will be left for future publication focused on the structure of G-actin. For our purposes here, it is sufficient to demonstrate that the method enhances sampling out of well-understood basins repeatably and without applying risky nonequilibrium driving forces in less well understood regions of CV space.

9.3.3 ClC Antiporter Chloride Transport

Our final example concerns the ClC-ec1 Cl^-/H^+ exchanger,[6] a paradigmatic protein of the ubiquitous ClC family.[210, 108, 161, 311] In this example we will focus on investigating one particular part of the molecular mechanism of chloride transfer: motion of chloride anions in the channel when a key central acidic residue, E148, is protonated. The previous examples showed that MBMetaD is an accurate and effective self-limiting mechanism when the CVs are appropriate in a metastable basin. However, it is often the case that CVs will

not be ideal even in the domain of interest, and with this final example we intend to show that MBMetaD continues to have both desirable self-limiting behavior and dynamics with desirable hysteresis in that case. Therefore, in this case our CVs are simply the displacements along the membrane normal direction, z, from each of two chlorides, one closer to the cytosolic side of the membrane, z_{cyt}, and the other closer to the extracellular side, z_{ext}, to the center of mass of a central group of protein atoms that is chosen only as a stable reference for the protein frame of reference and has no other intended physical significance. This leaves motion of E148, water fluctuations, helix motion, and other protein environment fluctuations as potential HSVs.

We aim here to compare the function of MBMetaD with unrestricted WTMetaD in estimating PMFs of simultaneous translation of these chlorides throughout the protein, with the mechanism presumed to involve five key chloride states identified in previous structural studies.[108, 109, 120] First, when z is greater than 1 nm, the chloride is said to be extracellular. Second, when z is near 0.6 nm, the chloride is expected to be in its external binding site. Third, when z is approximately -0.5 nm, it is expected to be in its central binding site. Finally, when z is less than -1, this coordinate does not distinguish well between whether the chloride is in its internal binding site or unbound on the cytosolic side of the membrane.

The ClC model system was set up as a dimer of the wild-type ClC-ec1 structure (PDB: 1OTS) embedded into a lipid bilayer (163 POPE) and solvated with 11,000 TIP3P waters[167] in a 92 x 92 x 79 $Å^3$ box under periodic boundary conditions. Protein and lipids were modeled with CHARMM27 force field;[208, 119] long-ranged electrostatic interactions were treated with the Particle Mesh Ewald (PME) method,[91] and the cutoff distance for the short-ranged interactions from both the Lennard-Jones and the real-space Coulomb interaction was set to be 12 Å. The classical MD simulation was performed using the Gromacs MD package[144] patched with a customized version of PLUMED2[324] and a 2 fs timestep. The initial configuration of the system was taken from our previous work,[340] where the system was first equilibrated for 7 ns in the NPT ensemble with a temperature of 310 K and a pres-

sure of 1 atm, then further equilibrated for 10 ns in the NVT ensemble with a temperature of 300 K. Residues E113 in monomers A and B and D417 in monomer A were protonated as suggested by previous calculations,[118] and E148 in monomer A was also protonated to allow chloride translations from the external to the central site. All other residues were set to their default protonation states.

Metadynamics was applied using the CVs described above for the chloride ions in monomer A only, as previous experiment[279] showed that each monomer carries out Cl^-/H^+ exchange independently. Monomer B was still included in the system and simulated as normal, not subject to any direct bias. Hills of initial height 0.2092 kJ/mol and widths 0.035 nm in both dimensions were added each 1 ps. For WTMetaD runs, the well-tempered bias factor was 13, whereas for transition-tempered MBMetaD runs the transition-tempered bias factor was 5 using transition wells (0.45,1.17) and (1.75, 0.12). MBMetaD domains were targeted to the transition barrier energy between those points plus 12 kJ/mol and were updated every 100 ps if the exterior bias level had changed by 1 kJ/mol. To restrict sampling to cases where the chlorides were in the protein in the appropriate locations in the protein interior, the coordinates of the extracellular and cytosolic-side chlorides were both restrained. The cytosolic-side chloride was restrained to the region $-3.24 < x < -0.93$; $-2.58 < y < -0.38$; $0.09 < z < 2.00$, while the extracellular-side chloride was restrained to the region $-2.90 < x < 0.70$; $-2.94 < y < 0.21$; $-1.49 < z < 0.0$. The restraints were implemented as half-harmonic walls with force constants of 200 kJ/(mol $Å^2$).

The free energy estimates given by three simulations using each method after various times, shown in Figure 9.8, do not correspond to converged potentials of mean force. They reflect the essential randomness of attempting enhanced sampling while neglecting key HSVs and show that MBMetaD prevents the nonequilibrium driving from spoiling the dynamics as it does so frequently in the unrestricted simulations. In the unrestricted case, the metastable band structure seen in the MBMetaD outside of the starting basin is overshadowed by other features in two of the three WTMetaD replicates.

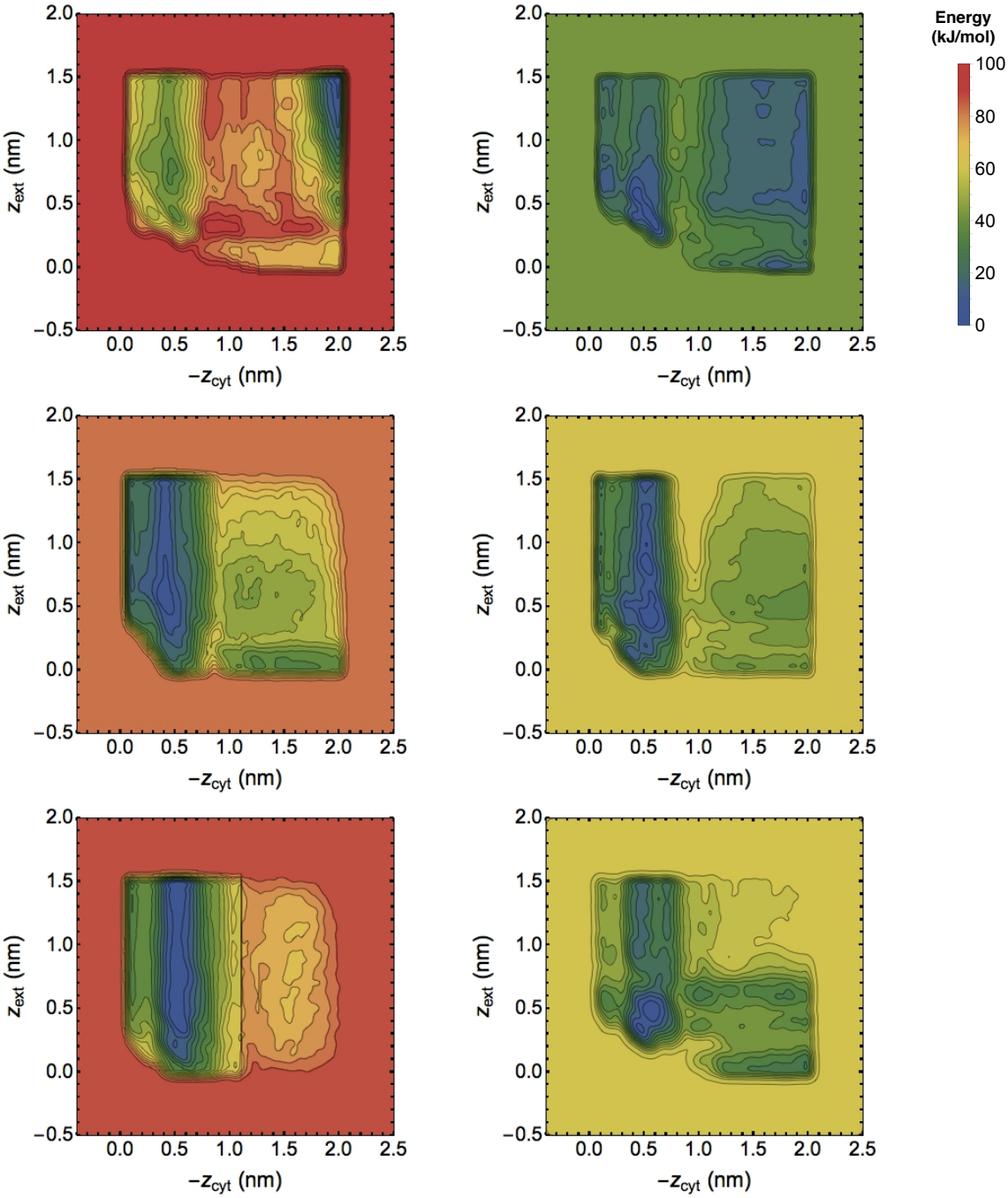

Figure 9.8: Bias-based free energy estimates from unrestricted WTMetaD (left column) and transition-tempered MBMetaD (right column) over three runs (rows), each run until features stabilized qualitatively or appeared to be irreparably incorrect. Clockwise from top left, these PMF estimates correspond to 250 ns, 300 ns, 300 ns, 300 ns, 350 ns, and 400 ns of simulation. None appear fully converged, yet the latter are more physically plausible and the former show distinct signs of dynamics gone irrevocably astray in the top and bottom runs (see text for descriptions of the atomistic details behind these unphysical features).

The top left energy profile shows an unexpected and implausible prediction of a sharp well where both chlorides are outside of the protein that is much deeper than any of the expected binding sites, while the bottom left shows a sharp line near $z_{cyt} = 1$ that is not seen in any other simulations and also does not correspond to any plausible physics: furthermore, the estimate beyond this line appears to simply be a blob without the expected basin structure for z_{ext}. In the simulation corresponding to the top unrestricted WTMetaD figure, the trajectory shows that the cystosolic chloride becomes stuck inside the protein between helices that are normally firmly bound together. In the simulation corresponding to the bottom unrestricted WTMetaD figure, one of the main helices unfolds and extends into the solvent. The chloride associates with the helix and then the nonequilibrium bias ratchets the helix apart by pushing on the chloride, unfolding it turn by turn, before the chloride dissociates and goes into the solvent, with effectively permanent damage done.

Judging by the noticeably different energy scale, it is evident that MBMetaD has expected self-limiting behavior even in this highly challenging and poorly tuned test case, and judging by the fact that the MBMetaD runs do not exhibit the odd behavior seen in the others, it also appears that self-limiting behavior improves robustness to CV quality as predicted. However, it is also clear that the MBMetaD runs in the right column of Figure 8 do not fully agree amongst one another, showing that it does not cover up core deficiencies in the CVs that can be seen by comparing multiple independent replicates.

The disagreement occurs at the level of which regions of interest are selected on the fly, as well. The two topmost right panels substantially agree. However, in the bottom right panel, the final region of interest does not include the upper right corner of the CV space. At the time of termination, the simulation was sampling the upper right region but not adding bias. Given time, the histogram correction would ensure that this region would properly register as part of the CV space to bias. Still, as seen in the alanine dipeptide example, it is better to use good CVs and a robust region definition than to rely on long-time convergence of the domains. This reflects the inefficiency of any CV-based enhanced sampling in the presence

of HSVs.

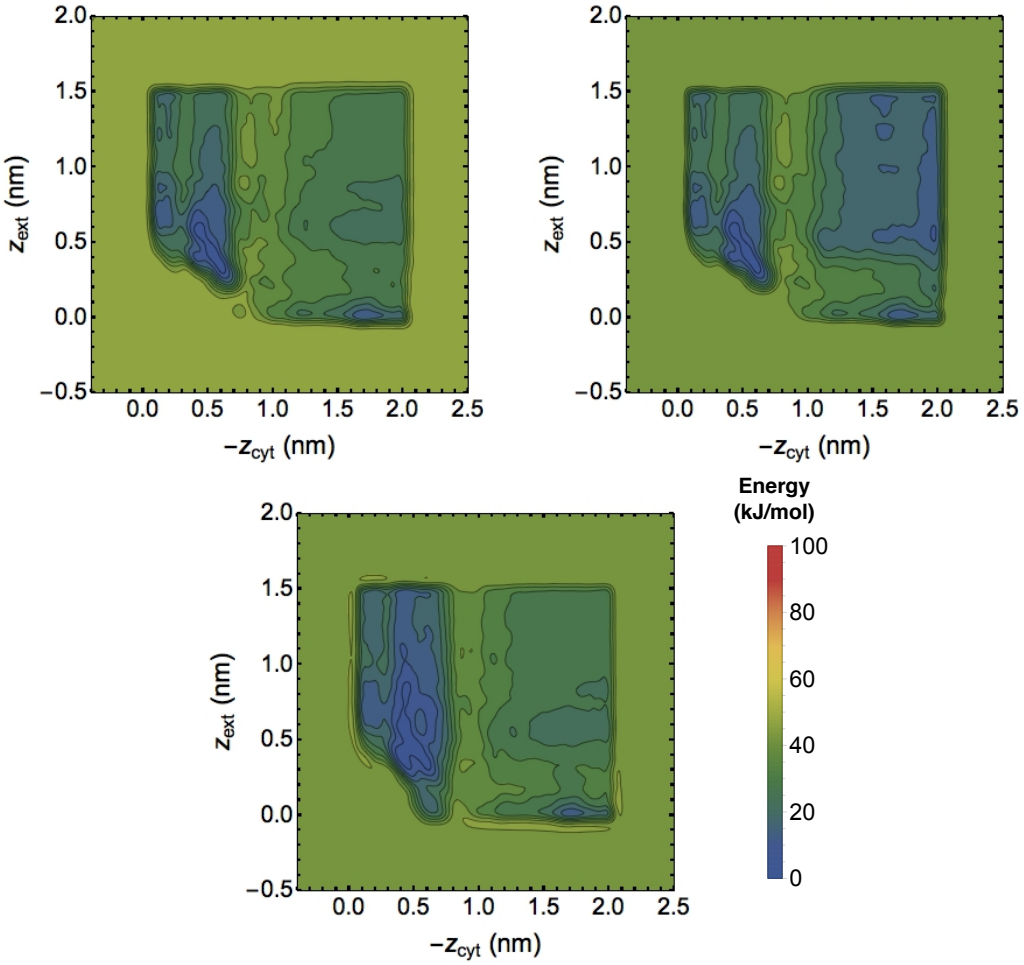

Figure 9.9: Biased-based free energy estimates from a branched MBMetaD trajectory. The bias-based PMF estimate of the primary trajectory at 200 ns (left), is compared with estimates at 300 ns from direct continution of that trajectory (middle) and continuation from a new random equilibrated initial conformation (right). The prominent differences between the two show that sampling remains highly autocorrelated on a 100 ns time scale even when biased, a hysteresis that properly indicates the presence of hidden barriers.

Next, mimicking the usual process of checking whether one run might have luckily converged to the correct result despite the gauntlet of possible issues that evidently affected the other runs, we initialized another MBMetaD replicate of length 100 ns beginning with the bias and stored histogram of the top run just before 200 ns to check if the PMF estimate of this replicate would agree with the original copy of the top run after 300 ns. As Figure 9.9 makes evident, that is not the case. The original run and its branched replicate differ; the

former samples configurations in which both chlorides are far from the center of the protein, whereas the other samples configurations in which the cytosolic-side chloride resides in the central pocket. The new replicate primarily explores the initial basin whereas the original primarily explored the other basins. Motion between the basins is not fully facilitated by the bias because key HSVs are involved in the true barriers between them, and the bias reports that faithfully. MBMetaD correctly reports the presence of these HSVs even as it prevents the less physically interesting and implausible dynamics seen in the unrestricted WTMetaD simulations. The domain restriction rule provides increased safety without covering up essential problems.

9.4 Conclusions

The new method MBMetaD fundamentally eliminates one of the primary limitations of metadynamics, that adding energy indefinitely causes undesired and irreversible change in many sensitive dynamical systems, by providing an effective and convenient self-limiting mechanism that causes it to fill up to a flexibly defined free energy level and no farther—requiring no *a priori* estimate of barrier heights, but rather estimating them on the fly as necessary. This allows for more focused study that should be especially practical when using many CVs at once and incidentally demonstrates a solution to the problem of boundary artifacts in metadynamics on domains of any shape in any number of dimensions. Most importantly, it means that using MBMetaD makes designing good CVs for metadynamics simpler. If energy is no longer added in regions of CV space that are irrelevant to transition mechanisms, the CVs no longer must be carefully tuned there.

Furthermore, having a controlled energy level makes setting the hill height and the tempering rates for WTMetaD and TTMetaD simpler. Current practice requires users to choose their desired energy level by balancing the level of tempering with the initial rate of biasing,[28, 27] but the tempering must also be matched to only partially known CV hysteresis time scales and the biasing rate must also be matched to the only partially known

rate of dissipation in the system. Thus the parameters are overdetermined in a manner depending on unknowns with opaque tradeoffs; using MBMetaD to disentangle the choice of final energy level from the choice of tempering rate and hill height can make the choices more straightforward. Additional technical complexity of the method for the implementer of it is therefore counterbalanced by decreased intuitive complexity for actual application.

By restricting our changes to the hill function alone, we ensure compatibility with many other advances in metadynamics methodology that cannot be discussed in depth here. For instance, the new procedure is compatible in principle with untempered,[181] well-tempered,[28] and transition-tempered[88] metadynamics, multiple walkers[275] and bias exchange,[263] driven metadynamics,[232] experimentally-directed (ensemble-biased) metadynamics,[353, 215] concurrent metadynamics,[135] multiple time steps,[121] and any choice of CV. It is not compatible with adaptive Gaussian hills,[49] but is compatible with field-coordinate metadynamics.[326]

Moreover, there is little reason to think that this strategy applies only to metadynamics. All other adaptive enhanced sampling methods such as the adaptive biasing force[92, 93] and orthogonal space random walk[369] approaches can similarly suffer hysteresis effects related to adding reckless driving forces when their parameters are set too aggressively. It may be that adding region-limiting mechanisms like the one presented here may allow for more of those aggressive parameter choices to be used safely, with substantial potential efficiency gains. In each case, one could use the bias and an auxiliary histogram together to determine regions of interest on the fly based on the simple definitions we propose here, such as 'everything less than 2 kT above the barrier between these two states.'

Finally, we find the physical underpinnings of MBMetaD unexpectedly natural and we hope they will inspire further thought in the field. The new hills can be understood as approximate Green's functions of diffusion inside the domain of interest plus an approximate quasistationary distribution in the exterior of the domain. Mathematical work on the foundations of accelerated dynamics[188, 189] is finding deep theoretical power in using qua-

sistationary distributions in lifting the Green's function of a coarse-grained Markov process onto a fine-grained configuration space. Therefore the MBMetaD hills appear to emerge as approximate mixed-resolution Green's functions, connecting the worlds of Markov state modeling, accelerated dynamics, multiscale modeling, and CV-based adaptive enhanced sampling in a surprising way.

CHAPTER 10

ADDRESSING FALSE CONVERGENCE: HALF-TEMPERED METADYNAMICS

This paper substantially reproduces material from a paper being prepared for publication, to be submitted under the authorship of Dama, J. F., Sun, R., & Voth, G. A.; I am indebted to Rui for both running the simulations that I designed and programmed and for detailed discussions of the findings.

Finally, we consider that our imp worries that, like Faust, the bachelor might die of old age and escape the torment before the design is complete. He finds it is better, overall, not to slow his whispers too quickly: it is better to risk overcompensating for the bachelor's preferences rather than to risk allowing them to persist without being countered by dissatisfaction before it is too late. And, to the bachelor's greater misfortune, our gleeful imp finds that even if his overcompensation may no longer entirely and exactly negate the bachelor's natural preferences every time, what remains of the preferences can still be guaranteed to eventually shrink to an infinitesimal distribution—a trivial mockery of real desire that the imp hardly minds leaving our tragic wretch.

10.1 Motivation

The proof provided earlier showed that well-tempered metadynamics[28] (WTMetaD) converges to the exact PMF asymptotically under very broad conditions compatible with any exponential or stretched exponential CV relaxation dynamics. Moreover, in unpublished work we have applied the same style of proof to establish the asymptotic convergence of the adaptive biasing force method[92, 93] and the $1/t$ Wang Landau method[33] with the same weak assumptions on the CV dynamics; all prove to be quite robust, in theory.

However, despite the proofs that these converge, practical experience shows that the methods do not always actually produce good PMF estimates even when those conditions

are satisfied.[27, 100, 302, 313, 339] This could formerly be attributed to the supposed approximate and nonequilibrium nature of the methods, but now that that has been ruled out by mathematical proof, this explanation is no longer tenable. Instead, the incorrectness of the PMF estimates must be due to finite time, pre-asymptotic behavior, i.e., the appearance of convergence before actual convergence has been reached. The problem of false convergence in metadynamics is in fact widely recognized to the point that some researchers advise against seeking convergence of the metadynamics bias at all.[85, 22, 134] These researchers instead recommend untempered metadynamics that does not converge, but rather continues driving the system at the same rate indefinitely, then using time averages of the indefinitely fluctuating bias to estimate a PMF instead of requiring the bias itself to converge.

This chapter will show that while both camps, the one favoring convergent tempered metadynamics and the one favoring robust fluctuation, have merits, there is in fact an ideal middle way between the two that provides both excellent convergence and superior robustness. To introduce it, I will first need to discuss three essential factors that govern the duration of the finite time, pre-asymptotic regime in metadynamics.

First, no metadynamics method can converge more rapidly than the biased trajectories decorrelate given fixed potential surfaces. Numerous statistically independent samples are always required to construct a good PMF, whether with adaptive biasing or not. The rate that simulation can create independent samples given an ideal bias on the CVs limits the asymptotic convergence rate for a convergent adaptive method. The quality of CVs is therefore crucial.

Second, the method cannot converge more quickly than the sizes of the bias updates, the so-called metadynamics hills, go to zero. So long as there is a finite hill height, the bias will fluctuate.

Third, the method cannot converge more quickly than the new bias can correct for the effects of unbalanced sampling due to initial conditions at early times. If the hill sizes are not large enough later on, then the bias may seem to converge but will in fact represent only

early-time sampling rather than a true PMF.

We addressed the first and second with our recent work on metabasin metadynamics[87] and transition-tempered metadynamics (TTMetaD),[89] respectively. This chapter will address the third.

In the field of stochastic approximation, it is well-known that to prove asymptotic almost-sure convergence, as we did for well-tempered metadynamics, the latter two factors must be balanced by allowing the size of updates (e.g., metadynamics hills) to an approximant (the bias) to decrease proportionally to $1/t$ in the asymptotic regime,[278, 179] as we proved for well-tempered metadynamics. Furthermore, for maximum rate of convergence, the specific rate at which update sizes decrease should be matched to the variance and autocorrelation of the sampling on the nearly-converged bias surface. However, it is also well-known that $1/t$ methods are rarely robust, and that in almost every case it is superior to use a $1/t^{\alpha}$ rule with α between one half and one, then to average the iterates over time.[179]

When a $1/t$ stochastic approximation method converges almost surely, it is typically the case that $1/t^{\alpha}$ methods with α between one half and one will converge in distribution and the time averages of the iterates will converge almost surely,[179] a scheme best known as Polyak-Juditsky averaging.[268] Moreover, the slower decrease in hill height at long times provides significantly greater robustness against false convergence, and thus greater reliability of single replicate approximations. $\alpha = 1/2$, or what we will call half-tempering, is generally taken to be a good default since it is the version most robust against false convergence that is still expected to converge in distribution.

This can be applied for any stochastic approximation scheme and thus any adaptive enhanced sampling scheme, from metadynamics[181, 28, 27, 313] to adaptive biasing force[92, 93, 145] to the orthogonal space random walk[369, 198] to variationally enhanced sampling.[329, 294] In this work we propose a specific implementation for metadynamics, calling it α-tempered metadynamics, and demonstrate its relative performance to both $1/t$ and constant hill height rules. The remainder of this chapter will first flesh out the new proposal in

the next section, then present the results of applying the new rule to calculate permeation PMFs of ethanol through a POPC membrane, then end with conclusions and an appendix describing combination rules for tempering schemes.

10.2 Method Definition

In metadynamics, one begins with a system that samples configurational states x that can be coarse-grained to CV states s via a map σ; the CVs must be defined so that the dynamics can be altered by adding biases $V(\sigma(x))$. Metadynamics is the process of iteratively updating a time-dependent bias $V(\sigma(x), t)$ at each regular time increment $t = k\Delta t$ for some chosen Δt according to the current state x_t and then generating the next new sample by continuing the previous trajectory under the newly biased system Hamiltonian.

The updates are called hills and take the form $h(t)G(s, \sigma(x_t))$; $h(t)$ is called the hill height and $G(s, \sigma(x_t))$ is called the hill kernel. $G(s, \sigma(x_t))$ often takes the Gaussian form $\exp(-||s - \sigma(x_t)||^2/2\delta^2)$ with δ some chosen width parameter, but it is important to choose it carefully when the bias domain has boundaries.[85, 22, 224, 87] In untempered metadynamics[181] $h(t) = h$ for some chosen h, in WTMetaD,[28] $h(t) = h \exp(-V(\sigma(x_t))/\Delta T)$ for some initial height h and energy scale ΔT, in transition-tempered[88] $h(t) = h \exp(-V_t^*/\Delta T)$ with V_t^* an estimate of bias at the highest barrier on the PMF between several CV state points of interest, and in global-maximum-tempered[353] $h(t) = h \exp(-\max_s(V(s,t))/\Delta T)$ where $\max_s$ is the maximum over all points s in CV space.

These rules where $h(t) = h \exp(-\hat{V}/\Delta T)$ for some $\hat{V}$ that increases by the same amount as the overall bias whenever a constant $\bar{V}$ is added to the bias give rise to $1/t$ rules, as shown in the proof of convergence.[89] The reason is that $h(t) \sim d\bar{V}/dt$ and $d\bar{V}/dt \sim \exp(-\bar{V}) \implies \bar{V} \sim \log(c + t) \implies d\bar{V}/dt \sim 1/(c + t) \to 1/t$. To get $d\bar{V}/dt \sim 1/t^\alpha$, or "$\alpha$-tempering," one needs $d\bar{V}/dt \sim \bar{V}^{(1-\alpha^{-1})^{-1}}$, or, in a smoother form that has the correct limits $\exp(-\bar{V}/\Delta T)$

as $\alpha \to 1$ and 1 as $\bar{V} \to 0$,

$$h(t) = h(1 + (\alpha^{-1} - 1)\hat{V}(t)/\Delta T)^{(1-\alpha^{-1})^{-1}}. \tag{10.1}$$

Using an $\alpha = 1/2$, this, then, is half-transition-tempered metadynamics:

$$h(t) = h(1 + V_t^*/\Delta T)^{-1}. \tag{10.2}$$

Figure 10.1 compares the hill height curves for a small set of values of α and ΔT as a function of the distinguished bias measure $\hat{V}$ and the curves of bias level as a function of time for the same set.

Adding thresholding corresponds simply to

$$h(t) = h(1 + \max(0, V_t^* - V_{th})/\Delta T)^{-1} \tag{10.3}$$

with V_{th} the chosen threshold bias before which no tempering will occur.

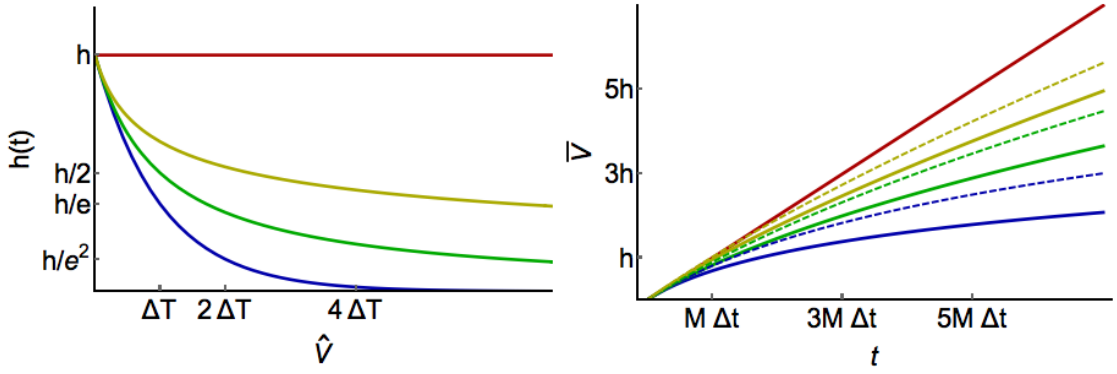

Figure 10.1: Behavior of the hill height (left) and idealized bias level evolution (right) for a small set of α values: 1 (blue), 0.5 (green), 0.25 (yellow), and 0 (red). All hill height curves follow Equation 10.1. In the idealized bias level evolution figure, solid lines correspond to choosing $\Delta T = h$, while dashed lines correspond to $\Delta T = 2h$, to illustrate representative shapes, Δt is the time between hill additions, and M is the number of $G(s, s')$ hills that must be added to the domain to achieve an average value of 1 on the domain (a ratio of the domain size to the raw hill shape's size). Typically $\Delta T \gg 2h$, but the qualitative behavior in the right panel is identical in that limit.

For better or worse, this changes the final convergence behavior when applied to WT-MetaD. Using the rule $h(t) = h(1 + V(\sigma(x_t))/\Delta T)^{-1}$ does not lead to convergence to a partially-flattened landscape, but convergence to a fully-flattened landscape. Writing it more like WTMetaD we see it is

$$\begin{aligned} h(t) &= h \exp(-\log(1 + V(\sigma(x_t))/\Delta T)) \\ &= h \exp(-V(\sigma(x_t)/(V(\sigma(x_t)) \log(1 + V(\sigma(x_t))/\Delta T)^{-1})) \end{aligned} \tag{10.4}$$

which is a version of well-tempering with a steadily increasing $\Delta T(t)$ scaling as ΔT at first and then $\bar{V}/\log(\bar{V}/\Delta T)$ in the asymptotic regime where $\bar{V} \gg \Delta T$. Intuitively, it starts by converging to a partially flattened PMF but moves to more and more flattened PMFs as time goes on. In the asymptotic regime it converges to the endpoint it would have if it were WTMetaD with an infinite bias factor, the exact PMF instead of a flattened PMF.

This is an extremely simple formula to add to existing implementations of tempered metadynamics. Given a code programmed like Plumed2[324] or the LAMMPS colvars package,[122] it can be as simple as changing a single line of the tempering method's algorithm and adding a single new input parameter α to the code's interface. Implemented in such a way, it can be entirely agnostic to the hills and CVs used and can also equally well be used with bias-exchange[57, 263] and concurrent metadynamics.[135] All of the various tempering rules can in fact be mixed and matched to an astonishing degree; a few essential rules for making these combinations are provided in the appendix to this chapter. Our Plumed2 implementation provides α-tempering rules for WTMetaD, TTMetaD, and GMTMetaD in such a way that all of them can be used at once with independent α values, thresholds V_{th}, and ΔT parameters, if so desired.

10.3 Application and Discussion

To demonstrate the new method, we applied it to the calculation of a 2D PMF of the permeation of ethanol through a POPC membrane. This is a challenging and realistic test system, but one that allows for high simulation replication and thus confident analysis of the dependence of each method on initial conditions, in particular on false convergence, the primary target for improvement for the α-tempering methods. The tests focus on demonstrating the greater relative robustness and precision of the $1/\sqrt{t}$ rule for transition tempering, called half-transition-tempered metadynamics (half-TTMetaD), compared to untempered and fully-transition-tempered metadynamics (full-TTMetaD).

The MD simulations were carried out using GROMACS-5.0.4[41, 194, 330, 144, 271] patched with Plumed2[324] modified by the Voth research group to perform α-TTMetaD using McGovern-de Pablo boundary conditions as a special case of metabasin metadynamics.[224, 87] The lipids were modeled using the CHARMM36 force field, while ethanol was modeled using the CHARMM general force field (CGenFF).[333] As suggested for CHARMM36[174] and CGenFF, the solvent followed the TIP3P water model.[167] The lipid bilayer was made up of 16 x 2 POPC molecules matching the article of Ghaemi et al.;[134] the bilayer was placed in the x-y plane and surrounded by 2064 solvent water molecules. The simulated system had dimensions of approximately 3.2 nm by 3.2 nm by 9.75 nm with periodic boundary conditions enforced along each dimension. An initial structure of the POPC lipid bilayer was prepared using the CHARMM-GUI membrane builder[164] then equilibrated for 100 ns in solvent at 323 K, chosen for consistency with the system of Ghaemi et al.[134] The thickness of lipid bilayer and area per head group were compared to the experimental value to confirm the bilayer equilibrated properly. Then the small molecules were randomly packed into the system using PACKMOL.[220] Initial velocities were randomly sampled from a Boltzmann distribution with a temperature of 323 K and re-equilibrated. The lipid and water plus small molecules were individually coupled to two 323 K heat baths using velocity rescaling with a stochastic term.[55] Pressure control was applied via a Berendsen barostat[40] that rescaled

the box every 5 ps. The cut-off distance for the short-range neighbor list was 12 Åand the neighbor list was updated every 40 steps. Long-range electrostatic interactions were approximated using the Fast Smooth Particle-Mesh Ewald (SPME) method.[116] All hydrogen bonds were constrained by linear constraint solver[142] (LINCS) and the MD integration time step was 0.2 fs.

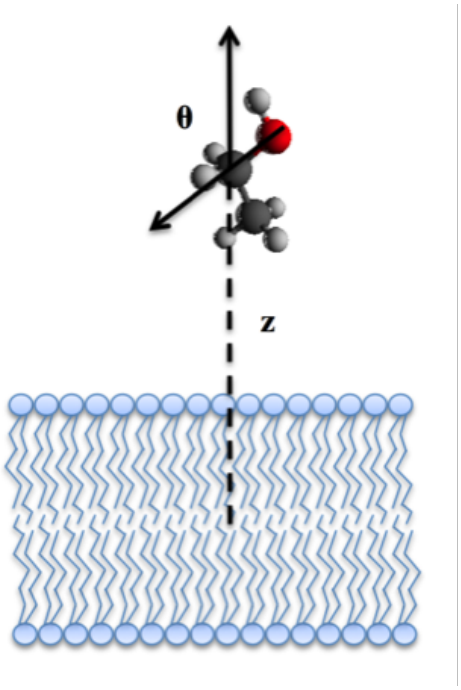

Figure 10.2: An illustration of the collective variables used in the application of half-tempered metadynamics to the calculation of ethanol permeation PMFs through POPC membrane.

The two chosen CVs were the box-scaled z-component distance, z, between the centers of mass of a single chosen ethanol and the membrane and the orientation, θ, between the vector from the chosen ethanol's hydroxyl oxygen to its adjacent carbon and the vector normal to the lipid bilayer, calculated as the vector between the centers of mass of the lipid head groups in one leaflet and the tail groups in the same leaflet. These are shown in Figure 10.2. The same base Gaussian hill shape was used in each MetaD simulation; only tempering pattern varied. The initial height h of the Gaussian function was 0.02 kJ/mol and the widths were 0.2 nm and 0.35 rad along the z and θ directions, respectively. New Gaussians were added once every picosecond in all simulations. These values were chosen in accordance with Ref.[134]

where the same system was studied using untempered bias-exchange metadynamics. In every case the TTMetaD CV-space points of interest were where the ethanol was at the center of the solvent region and perpendicular to the membrane, $z = \pm 0.5$ and $\theta = 0$, and where the ethanol was in the center of the membrane and perpendicular to it, $z = 0$ and $\theta = 0$.

The first set of simulation conditions compared half-tempering ($\alpha = 0.5$) to full-tempering ($\alpha = 1.0$) across three ΔT values: 3 kT, 5 kT, and 8 kT. 8 independent replicates with different randomly sampled initial ethanol positions and different randomly sampled initial system velocities were performed for each. The evolution of the estimates for the central barrier height for each of these and the consensus estimate is shown in Figure 10.3. The lowest ΔT values show clear signs of false convergence for $\alpha = 1.0$ that are far less pronounced for $\alpha = 0.5$, consistent with our expectations of improvement. The spread of the replicate estimates is significantly improved using $\alpha = 0.5$, as shown in Figure 10.4, and more importantly, the lower autocorrelation for each walker implies that each walker's result is a more reliable individual indicator of convergence when using half-tempering.

The traces in Figure 10.3 show an undesirable pattern of outliers, however. Frequently, it appears that one run will estimate a high or low barrier and appear to be stuck with that estimate even if the others show lower autocorrelation. Looking into the history of the bias in these cases, we saw that these outliers occured when the width of the hill allowed the bias to connect from the solvent region into the membrane region of CV-space before an actual barrier crossing occurred, tricking TTMetaD into the convergence regime too early. This effectively freezes the bias due to initial conditions, causing false convergence. Happily, this is exactly what the thresholding rule introduced in the first TTMetaD paper[88] is designed to address, so in our next trials we compared half-tempering to full-tempering once again, but this time with 1 kT threshold values (V_{th}) to prevent premature entry into the convergence regime, this time comparing with only the lower ΔT values.

The results are the highest-quality, most reproducible and reliable convergence traces we have yet seen using metadynamics on a test system of this complexity, shown in Figure 10.5,

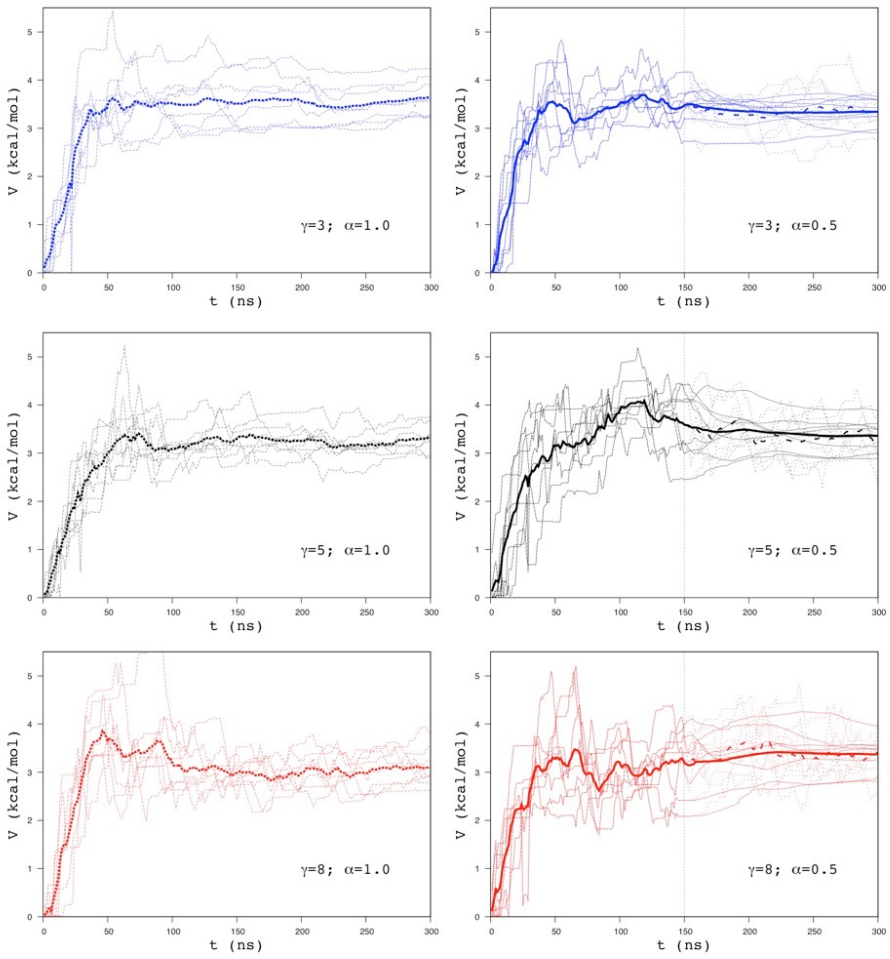

Figure 10.3: Evolution of the estimate of central permeation barrier height as a function of time using fully transition-tempered ($\alpha = 1.0$) and half-transition-tempered ($\alpha = 0.5$) metadynamics with a variety of tempering parameters ΔT, written in the legend in terms of $\gamma = 1 + \Delta T/kT$. Thin lines indicate traces of individual runs, while thick lines indicate overall averages; dashed lines indicate raw traces, while solid lines indicate traces that are time-averaged from 150 ns. Time averaging is not appropriate with full tempering.

with the corresponding spreads shown in Figure 10.6. The full-tempering rule still shows false convergence, though the outliers are less severe, while the half-tempering rule shows excellent bunching without any apparent false convergence. Note that 1 kT is the first and

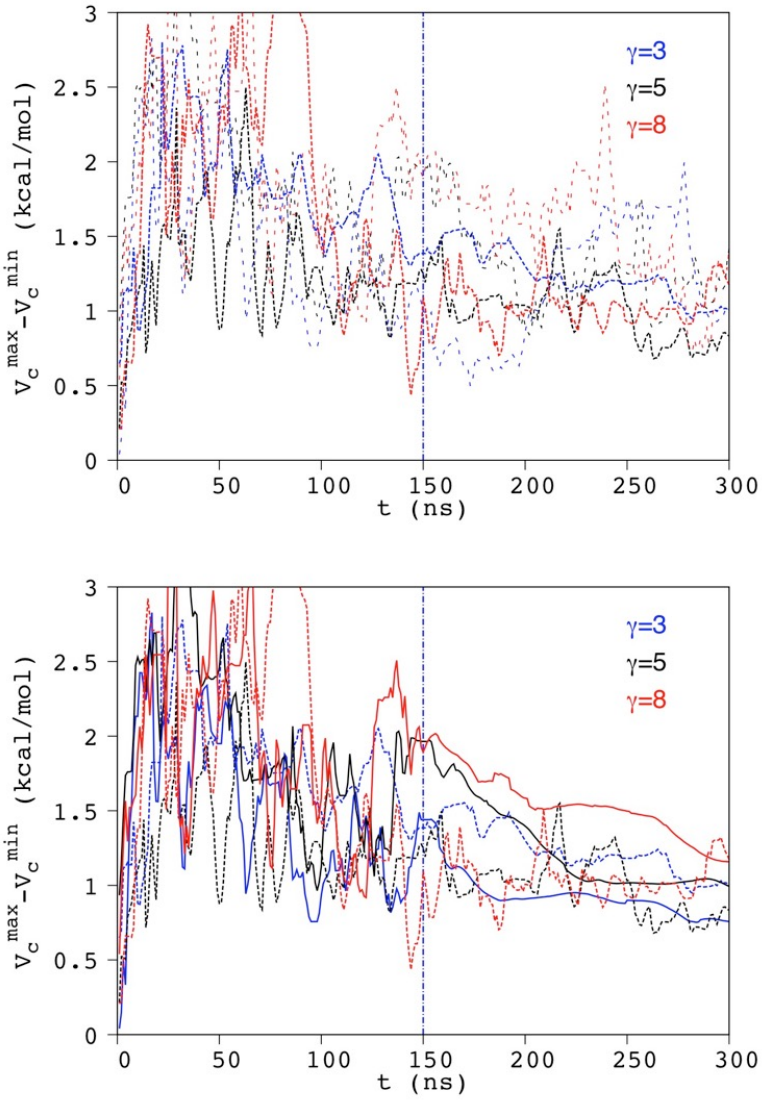

Figure 10.4: Evolution of the spread of central permeation barrier height estimates as a function of time using fully transition-tempered (closely dashed lines) and half-transition-tempered metadynamics (top, sparsely dashed; bottom, solid) with a variety of tempering parameters ΔT, written in the legend in terms of $\gamma = 1 + \Delta T/kT$. In the top panel, the spread of half-tempered traces is calculated among raw traces; in the bottom panel the spread is calculated among traces time-averaged from 150 ns for the half-tempered runs. Time averaging is not appropriate with full tempering.

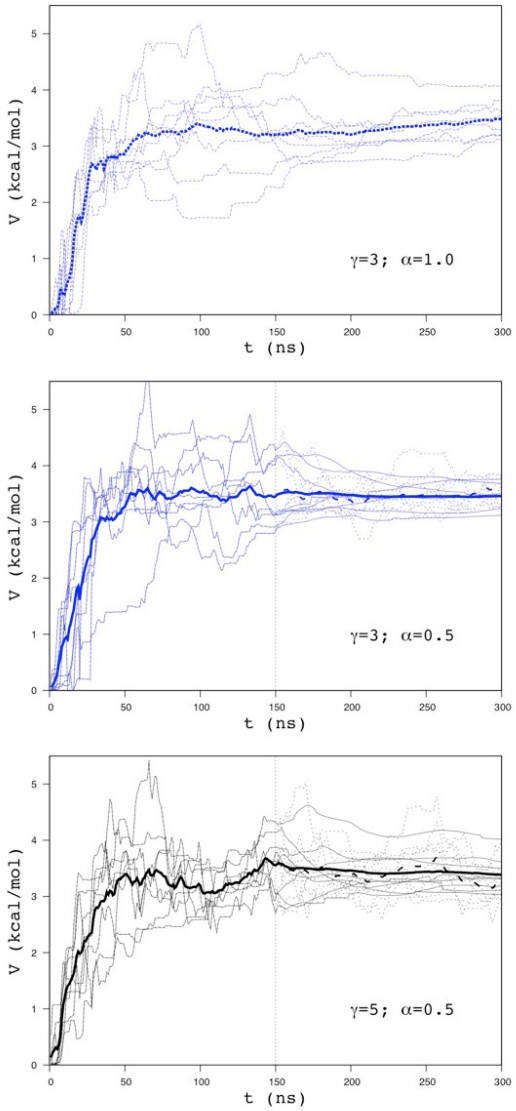

Figure 10.5: Evolution of the estimate of central permeation barrier height as a function of time using fully transition-tempered ($\alpha = 1.0$) and half-transition-tempered ($\alpha = 0.5$) metadynamics with a variety of tempering parameters ΔT, written in the legend in terms of $\gamma = 1 + \Delta T/kT$. Thin lines indicate traces of individual runs, while thick lines indicate overall averages; dashed lines indicate raw traces, while solid lines indicate traces that are time-averaged from 150 ns beginning at 150 ns. Time averaging is not appropriate with full tempering.

only threshold value we examined; essentially no tuning of this value appears to be required.

In the previous test cases, the methods compared were all convergent either in distribution

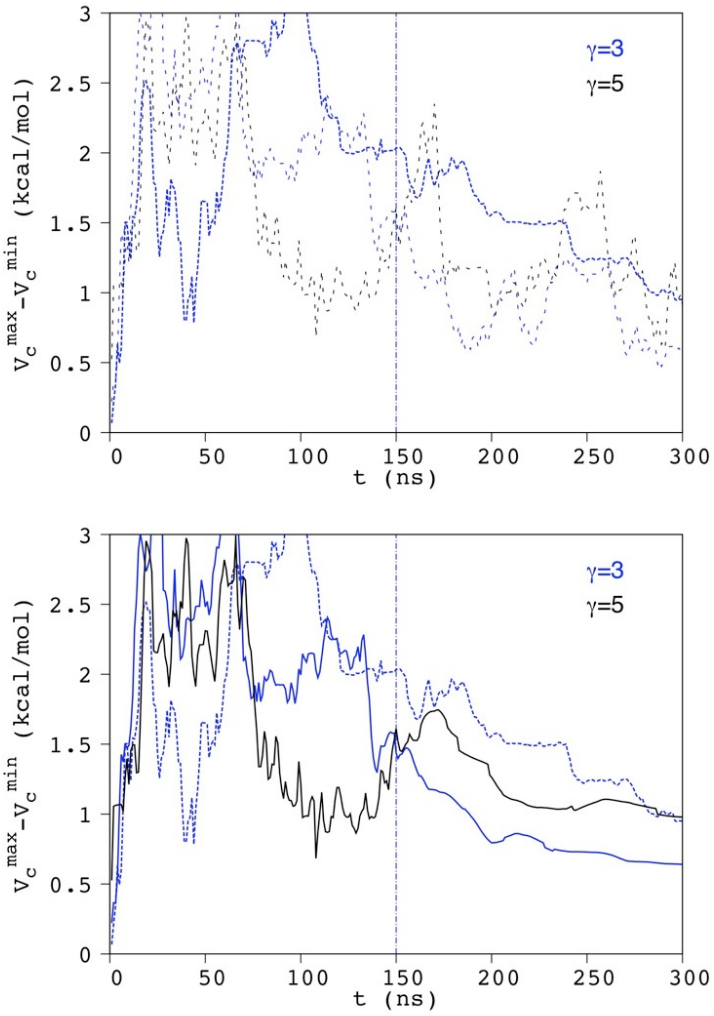

Figure 10.6: Evolution of the spread of central permeation barrier height estimates as a function of time using fully transition-tempered (closely dashed lines) and half-transition-tempered metadynamics (top, sparsely dashed; bottom, solid) with a threshold of 1 kT and a variety of tempering parameters ΔT, written in the legend in terms of $\gamma = 1 + \Delta T/kT$. In the top panel, the spread of half-tempered traces is calculated among raw traces; in the bottom panel the spread is calculated among traces time-averaged from 150 ns for the half tempered runs. Time averaging is not appropriate with full tempering.

or with probability one. In the next set of tests, we compared to untempered and 30%-tempered metadynamics ($\alpha = 0.3$) to verify that half-tempering is in fact in a sweet spot

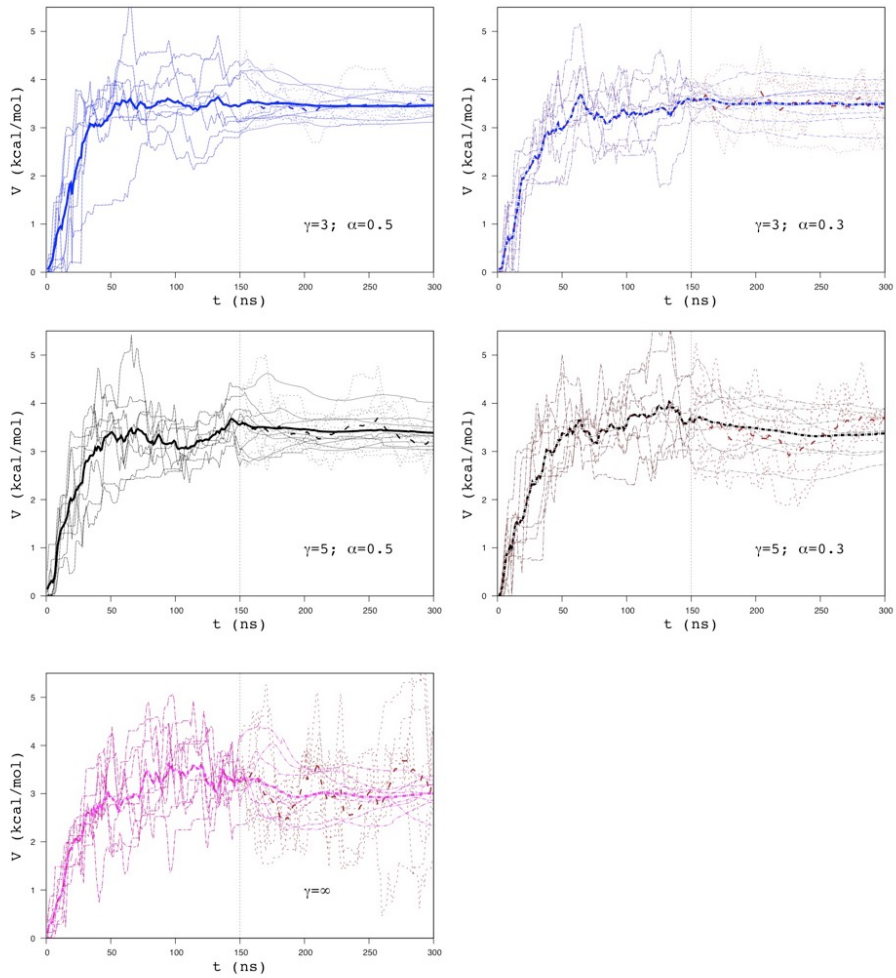

Figure 10.7: Evolution of the estimate of central permeation barrier height as a function of time using half transition-tempered ($\alpha = 1.0$), 30% transition-tempered ($\alpha = 0.3$) metadynamics, and untempered metadynamics ($\gamma = \infty$) with a variety of tempering parameters ΔT written in the legend in terms of $\gamma = 1 + \Delta T/kT$. Thin lines indicate traces of individual runs, while thick lines indicate overall averages; dashed lines indicate raw traces, while solid lines indicate traces that are time-averaged from 150 ns beginning at 150 ns.

maximizing robustness and rate of convergence in terms of precision. The 30% tempering rule is intermediate between untempered and half-tempered metadynamics, and therefore might be suspected to offer more robustness and lower autocorrelation than half tempering at an

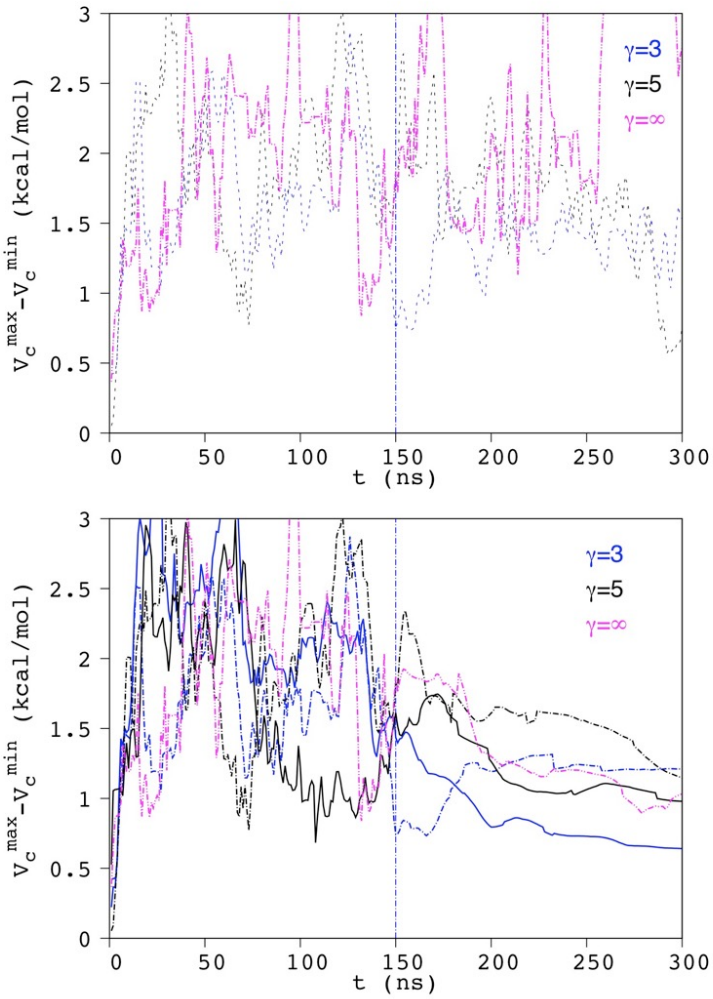

Figure 10.8: Evolution of the spread of central permeation barrier height estimates as a function of time using fully transition-tempered (closely dashed lines) and half-transition-tempered metadynamics (top, sparsely dashed; bottom, solid) with a threshold of 1 kT and a variety of tempering parameters ΔT, written in the legend in terms of $\gamma = 1 + \Delta T/kT$. In the top panel, the spread of half-tempered traces is calculated among raw traces; in the bottom panel the spread is calculated among traces time-averaged from 150 ns.

acceptable cost in precision. However, as shown in Figures 10.7 and 10.8, in fact the 30% tempering offers no benefit in terms of autocorrelation or robustness against half-tempering with a default threshold.

Comparing these various results, it seems evident that above $\alpha = 0.5$, the sampling autocorrelation becomes a dominant source of error over the bias due to initial conditions, whereas below $\alpha = 0.5$, the false convergence that retains bias from initial conditions becomes more dominant; while we have not fully explored the region right around $\alpha = 0.5$, half-tempering appears to be an excellent mean between the imprecision of untempered metadynamics and the false convergence of fully tempered $1/t$ metadynamics when using a low ΔT. Also, it is gratifying to see that thresholding does in fact have a significant effect in this more complex application despite making little difference in the simpler test cases in the initial TTMetaD paper;[88] this validates our speculation to that effect in that paper.

In general, the choice of metadynamics parameters that is strictly best will always depend on the time and energy scales of the system of interest. In our experience, however, half-tempering with a ΔT of 3 kT and a threshold of 1 kT has proven to be an excellent default for membrane permeation and small peptide dynamics. However, whenever studying any new class of systems we highly recommend to replicate runs at least 5 times each to detect false convergence and the types of outliers encountered above that indicate a need for a higher threshold.

10.4 Conclusion

Well-tempered metadynamics has long been known to have problems with convergence when parameters are chosen poorly. Reproducibility and false convergence become especially serious in cases that hills shrink too rapidly, but precision and stability present serious challenges when hills do not shrink quickly enough.[302, 27, 313, 88] This has been the basis of a long-lived controversy between users of WTMetaD and the original metadynamics. Our final position, which may make neither camp happy or may provide a basis for reconciliation, admits the shortcomings and advantages of both and suggests a happy mean between the two extremes of convergence and fluctuation: half tempering.

It is our hard-learned experience that no PMF estimate based on a single run of a $1/t$

metadynamics method should be considered reliable without special validation using either the biased sampling histogram[49] or multiple independent replicates; we always recommend using these methods to estimate error instead of appealing to slow change of the bias as a sign of convergence. However, the fact that the original metadynamics does not converge[181, 183, 57] is also a serious limitation. It can be less dynamically stable[87] and it is not clear that averages of the fluctuating bias converge to the true negative PMF when the CV dynamics is not Brownian.[57]

The present method appears to offer significantly improved security against false convergence, but without sacrificing precision, stability, or theoretical accuracy. It can still suffer from false convergence somewhat, so care is still required when applying it. However, the evidence provided above and experience in the field of stochastic approximation suggest that the problem is significantly ameliorated using the $1/\sqrt{t}$ rule instead of the $1/t$ rule, even if not fully solved.

Happily, the new method is also extremely easy to use. It is implemented in a free to download, public fork of Plumed2[324] for WTMetaD, TTMetaD, and more. It can be used simply by adding a single keyword, e.g., TTALPHA=0.5 for TTMetaD or WTALPHA=0.5 for WTMetaD. No additional post-processing is strictly required, though one can use simple time-averaging of the sort already used for the original metadynamics to achieve increased precision if desired.

The half-tempering method builds on known methods in a very simple, modular way. It works with WTMetaD[28] and TTMetaD[88], among others.[89, 353] It is compatible with any choices of hills, from fixed width[181] to adaptive width[49] to metabasin open-boundary-corrected.[87] The idea also applies to more distinct methods: to use it with adaptive biasing force methods[92, 93] one would alter the averaging rule to increase the weight of new samples by the square root of the the number of previous samples; in variationally-enhanced sampling,[329] one would simply adjust the step size differently.

Adaptive enhanced sampling benefits immensely from automatic tuning to cleanly sepa-

rate the convergence regime from the exploration regime, like in TTMetaD,[89] from safety mechanisms to enhance stability of the non-equilibrium driving process, like metabasin metadynamics,[87] and increased robustness from accepting convergence in distribution rather than almost sure convergence, as in the current work. In our hands, these have transformed our capabilities with metadynamics, but each of these principles is theoretically applicable to any other adaptive enhanced sampling methods as well.

10.5 Appendix: Combining Tempering Rules

The proof of convergence[89] for WTMetaD and globally tempered metadynamics (GT-MetaD) allowed for $h(t)$ to be proportional to a well-tempering factor (WTF), $\exp(-V(\sigma(x_t))/\Delta T)$, times any global-tempering factor (GTF), a functional $w[V(s)]$ of $V(s)$ such that $w[V(s) + \bar{V}] = w[V(s)]O(\exp(-\bar{V}/\Delta T'))$, where Ω indicates a tight scaling relationship in the asymptotic regime $\bar{V} \gg \Delta T'$ and the prime simply indicates that the two ΔT parameters for the WTF and GTF may be different. This allows for astounding flexibility and quite a bit of mixing and matching.

For instance, the $1/t$ rules for transition-tempering, well-tempering, and so on can be multiplied and added together, and assuming each factor j of each term i has its own ΔT_j, tempering using any sum of products of these rules is theoretically convergent so long as $\bar{\Delta T}_i = [\sum_{j \in i} \Delta T_j^{-1}]^{-1}$ satisfies $\infty > \bar{\Delta T}_i > 0$ for every i. When there are distinct ΔT_j for different well-tempered terms, the one in the product term with greatest $\bar{\Delta T}_i$ will determine the degree of flattening of the final convergence target.

An example application would be well-tempering with a high Delta T together with transition-tempering with a lower ΔT and lower α. In this case the WTMetaD behavior would dominate the early and late stages of convergence, but transition tempering could take over the convergence in an intermediate regime so that using a high well-tempered ΔT for the initial and late convergence phases does not force either a small initial hill height or a very long late phase in which fluctuations shrink very slowly.

When combining $1/t^{\alpha}$ rules, the behavior is quite a bit more complicated because the exponential product rule cannot be used to simplify the terms. However, assuming that all of the α are between 0 and 1 and all ΔT_j are greater than zero, so long as each term has at least one term with $\alpha \geq 0.5$ one expects it to converge in distribution.[179] Furthermore, if all of the terms have factors with $\alpha = 1$, then it will converge as a $1/t$ method using the greatest $\bar{\Delta T}_{i,\alpha=1} = [\sum_{j \in i:\alpha_j=1} \Delta T_j^{-1}]^{-1}$ to control the convergence rate.

WTFs make the behavior especially tricky. If any of the WTFs have $\alpha = 1$, the system will flatten according to the one in the factor i with the greatest corresponding $\bar{\Delta T}_{i,\alpha=1^-}$ except in very carefully balanced cases where $\bar{\Delta T}_{i,\alpha=1} = \infty$. Specifically, one can multiply two perfectly counterbalanced well-tempered and globally-tempered terms with $\alpha = 1$, $\Delta T_{WT} = -\Delta T_{GT}$, to create a tempering that changes the target PMF without shrinking the hills if only the level of the bias changes. Then, adding a third term to the product with a chosen α, one has designed a method with a partially flattened target but an $\alpha < 1$. An example application would be to only half-flatten a surface, as in an aggressively tempered WTMetaD, but with rapid bias addition and robustness against false convergence very unlike a low-ΔT WTMetaD. This could be particularly useful in well-tempered ensemble replica exchange calculations, for example, or when used in tandem with metabasin metadynamics in highly dynamically sensitive biomolecules.

These combination rules allow one to substantially decouple previously intractable trade-offs between initial convergence speed, stability, degree of flattening, and final convergence rate, providing an extensive design space for tackling the most challenging free energy calculations with the most complex CV dynamics.

CHAPTER 11

COARSE-GRAINING MANY COMPLEX MOLECULES: VARIATIONAL ULTRA COARSE-GRAINING

This chapter substantially reproduces the material of the paper Dama, J. F., Sinitskiy, A. V, McCullagh, M., Weare, J., Roux, B., Dinner, A. R., & Voth, G. A. (2013). The Theory of Ultra-Coarse-Graining. 1. General Principles. Journal of Chemical Theory and Computation, 9(5), 24662480. Anton Sinitsky and I shared first authorship, but I developed the theory independently; our collaboration was primarily in writing.

The theory of ultra coarse graining adds discrete states to particle models. If we were to complicate our social analog from before in a similar way, we might consider what happens when a family has visitors or does not: what is the effect on the neighbors? What is their preference for the family after the new visitor arrives? And the family with the visitor may see its preferences for interactions with various neighbors change as well. If a census taker were attempting to understand the dynamics of the households in an area, it would always be prudent to ask on the survey who the residents are and how recently their residences have changed. If not, then households that change in composition would erroneously be modeled as one single family type or another.

In coarse-graining, we are often in the position of having a clear picture of the "households" occupying an area, but without knowing the "family structure" inside: we see a protein's gross conformation at a coarse-grained level, for instance, but can easily miss its allostery.

In some cases, these family structures can be defined crisply. Nuclear families and a few singles might occupy an old sitcom suburb. However, in general cases it can be very hard to tell from the outside what exactly a family's structure is—is Grandma over for a couple of weeks or indefinitely? Are the Smiths giving each other a little space or separating? Often, as a parent, it might be easiest to explain these cases to children as mixtures of the multiple

state. Grandma may or may not be staying indefinitely, we'll find out. We can imagine the kids thinking about their neighborhood in terms of households with definite locations but occasionally uncertain family structure.

The same will be true of our proteins in this chapter. We may know just how the gross structure of the protein is arrayed, but be unable to pin down exactly which state an allosteric switch is in. The chapter describes this further and introduces bottom-up variational methods for developing such models.

11.1 Introduction

Many key biomolecular processes implicated in cellular growth, function and disease cannot be studied effectively by means of atomistically detailed computer simulation despite impressive recent progress in computational hardware, software, and methodology.1-9 Furthermore, even when these processes can be simulated at the all-atom resolution level, it is often unclear whether these simulations are necessary or efficient uses of computational resources and whether or not the resulting data are so detailed that they obscure the essential physics of the process being studied. For these reasons, the computational biophysics community has devoted considerable effort to develop efficient coarse-grained (CG) models that describe the processes of interest in reduced detail and at lower computational cost.6,10-20

An important subclass of these CG models is developed by using systematic CG methods to create reduced-detail approximations of systems by direct comparison to existing fine-grained (FG), for example, all-atom, models.21-36 These methods have the goal of reducing the computational cost of obtaining CG statistics from the FG model and of elucidating the essential details needed to reproduce the physics of the FG model at a CG level; they can be contrasted to methods in which CG models are developed to reproduce experimentally known properties directly rather than through agreement with a FG reference model.37-40 The two approaches each have complementary merits, with one particular advantage of systematic CG methods being that improvements in FG models can be used to improve

systematically derived CG models and that this class of CG models can be validated against reference models more thoroughly and transparently. A disadvantage is that the connection of these models to experiment is more indirect and they can inherit the systematic errors of their base FG models; however, FG models contain a wealth of experimental data and accumulated physical insight that cannot always be incorporated in a CG model through direct fits to experiment. Therefore, while the connection to experiment for systematic CG models may be more indirect it can also be more extensive and varied, especially in systems for which detailed experimental data is sparse. Which approach should be applied to test a particular hypothesis will depend on the relative availability of reliable FG models and directly relevant experimental data and cannot be judged without expert knowledge of the particular system in question. Previous researchers have argued that systematic CG methods can provide quantitatively accurate, computationally efficient models for a wide range of complex molecular phenomena including biological processes.6,14,19,23,28,41 Two very different systematic coarse-graining approaches show promise and have attained significant early successes.

In the first of these, CG proceeds by collapsing groups of atoms into single effective particles, also called CG sites, and then using statistically averaged collective forces from the atomistic model to define interactions between these effective particles.[154, 246, 247, 366] This approach is inspired by a structural biology view of biomolecular function in which the most important fluctuations occur as smooth deformations such as twisting, compression, and stretching of common structural motifs. Systematic methods of this type applied to biomolecular simulation include, among others, reverse Monte Carlo43,44,[240, 241] Boltzmann inversion,[276, 328, 17] elastic network modeling (ENM),[363, 206, 24] multiscale coarse-graining (MS-CG)[154, 153, 246, 247, 371] and relative entropy minimization.[296, 63] Most CG models of this type described in the literature have the resolution level of one effective CG particle per several heavy atoms, while more coarse models with one CG site per several amino acid or nucleotide residues, on the order of a hundred atoms per effective

particle,[366, 104, 15, 367, 368, 303, 304] are not as numerous.

In the second approach, a "coarse-graining" of sorts proceeds by collapsing sets of atomistic configurations into single effective states. Statistics from the atomistic model are then used to define transition rates and free energy differences between the discrete states.[75, 270, 257, 187, 336] This approach is inspired by a more biochemical view of biomolecular function in which the most important changes, such as binding events, chemical reactions, and protein domains folding or unfolding, occur in the form of discrete kinetic transitions between states with particular characteristics. Often an additional assumption that the system retains no memory of previous transitions is introduced so that the CG models are Markov state models (MSM).[75, 270, 257, 160, 243, 256, 274, 184, 173]

Effective particle bottom-up coarse-graining (CG) methods are well-suited to modeling systems in which all events of interest at the CG scale can be naturally explained in terms of smooth structural transitions, whereas effective state models are better suited to modeling systems that seem to evolve exclusively through more abrupt transitions between metastable states when viewed at the CG level. However, neither of these methods is generally applicable to all possible systems.

In key, paradigmatic biomolecular systems, for any level of coarse-graining beyond a few CG particles per residue, there also exists the possibility that some effective CG particles should be able to undergo protonation or other reactions (e.g. ATP hydrolysis), or major conformational changes "within" the CG site. These internal CG site state transitions should be connected to the effective CG particle configurations. Though systematic CG methods for combining different metastable protein conformations in a single model have been previously employed in the literature, no previously published systematic CG method appears capable of constructing models that can represent the essential physics at an "ultra coarse-grained" (UCG), tens to hundreds of atoms per CG degree of freedom, scale as just described. This chapter aims to fill that gap, which is an important step in multiscale theory and simulation that must first be taken to link molecular scale interactions with cellular scale phenomena

involving many interacting biomolecules.

This chapter will demonstrate that the effective state and effective particle coarse-graining can be combined naturally within a single UCG framework, enabling the development of models that combine discrete transitions with more continuous motion. The methods presented here promise to combine the strength of these two types of models, by allowing CG particles to have internal states and by allowing state transitions to couple to continuous structural deformations within individual models, using a similar level of rigor to that currently attainable in CG strategies at a higher level of resolution. Our strategy is to split configuration space into discrete volumes based on the values of collective variables either within or between CG particles, then simultaneously optimize a CG model for each volume using volume-restricted FG particle distributions as individual reference models. A CG model similar in spirit has been developed for lipids in the past,[241] but to our knowledge this is the first general, systematic derivation of such state-dependent CG models from generic atomistic FG reference systems in the canonical ensemble.

The remainder of this chapter is organized as follows. Section 2 will discuss motivation of the UCG methodology in the case of a specific and particularly illustrative example, the actin filament. Section 3 will introduce notation for FG models, CG models, and UCG schemes. Section 4 will define consistency between the models and derive useful mathematical formulations of consistency in terms of effective free energy surfaces, forces, and masses. Section 5 will then illustrate the theory by providing examples of applying UCG consistency to simple models that do not explicitly require an approximate treatment. Section 6 will next describe variational approximations to consistency between free energy surfaces, while Section 7 will complete the body of the chapter with a summary of the requirements of the method and a discussion of its relationship to previously introduced approaches. Finally, Section 8 will conclude the paper.

11.2 Motivation

As a paradigm example motivating UCG methodology, consider the problem of constructing a model of actin filaments (F-actin). These filaments are a key component of the cellular cytoskeleton and play a crucial role in the dynamics of the cell.[266, 267] A single monomer (Figure 11.1a) contains 2883 atoms,[254] and at least 13 monomers are required to approximate an infinite filament using periodic boundary conditions (Figure 11.1b).[79] As a result, standard atomistic molecular dynamics (MD) simulations of F-actin with explicit water represent a significant challenge even with the use of modern supercomputing resources; the physical time scale of these simulations is currently limited to on the order of 100 ns.[289, 117] This time scale is insufficient for full sampling of the filament dynamics,[289, 117] which motivates the use of enhanced sampling techniques with biophysically motivated collective CG variables.

Previous work has considered a highly CG model of actin monomer with four CG sites (Figure 11.1c).[366] In this model, each of the four CG beads was placed at the center of mass of a corresponding domain, and assignment of amino acid residues to one of the four domains was performed by essential dynamics coarse-graining (ED-CG)[366],54,55 based on the spatial correlation of thermal fluctuations of the atoms in the protein. This model reflects many (though not all) important aspects of F-actin filament dynamics.[289, 117] To improve the performance of the model, one possible way is to add more CG sites and, respectively, to introduce more continuous CG variables, which may be defined as the centers of mass of specific fragments of the actin monomer.[289, 117]

However, an alternative approach is possible. Experimental studies have established that polymerization of G-actin monomers leading to F-actin, depolymerization of F-actin releasing back G-actin, and interactions of F-actin with other proteins are modulated by the states of the nucleotides bound by the monomers in the filament (ATP, ADP or ADP + phosphate) and, probably, by the states of the D-loops and other residue conformations in the monomers (folded or unfolded).[266, 79, 351, 107] This behavior suggests that successful

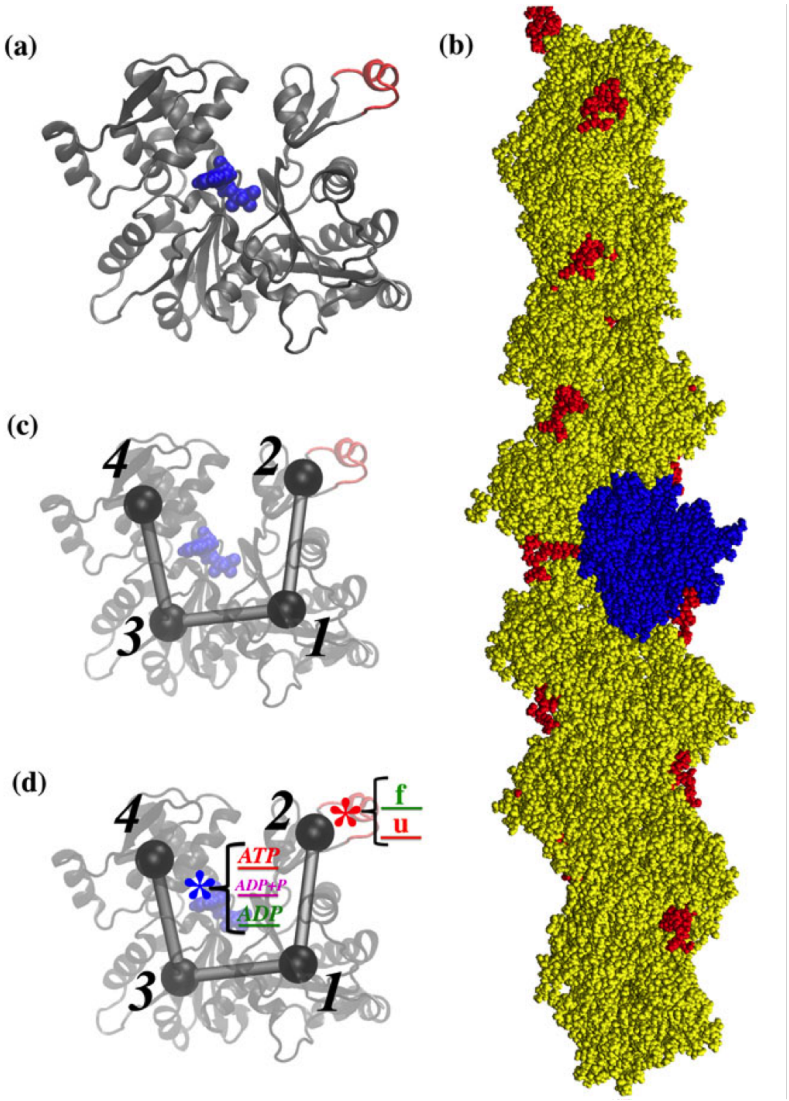

Figure 11.1: (a) Actin is a 375-residue protein of vital importance for living cells (D-loop shown in red, nucleotide in blue). (b) To simulate F-actin filament dynamics, one must use periodic boundary conditions with 13 monomers in a unit cell, which makes straightforward MD simulations computationally expensive. (D-loops are shown in red, one monomer in blue) (c) A simple CG model of actin with four structureless CG sites per monomer as studied in previous work.[366] (d) Addition of two discrete CG variables indicating the states of the nucleotide and the D-loop is expected to significantly improve the performance of the resulting UCG model at the expense of only a slight increase in the number of collective variables.

UCG models of F-actin should explicitly distinguish between different states of the bound nucleotides and between different states of the D-loops within the CG sites. To improve the four-site CG model without increasing the number of sites, one could, for example, add two

discrete CG variables per monomer to indicate the states of the nucleotide and the D-loop in the monomer or other protein conformations.

In the resulting model, GC bead number 2 has two internal states defined by whether the D-loop in it is folded or not. As for the other discrete CG variable, it indicates different states of the actin monomer as a whole rather than a specific CG site in it, since the nucleotide was not explicitly included into any of the four CG domains in the original model.[366] The resulting heuristic UCG model (Figure 11.1d) better fits with biochemical and biophysical reality as compared to the original model with just four structureless CG beads. Given the intuitive attractiveness of this approach and its ability to still scale up to much larger length and time scales relative to all-atom or more highly resolved CG models, the question arises of how to build such an UCG model in a systematic way, strictly from the viewpoint of statistical mechanics. This problem is discussed in the subsequent sections.

11.3 Model Definitions

The goal of this paper is to describe an UCG methodology suited to creating simple, intuitive highly CG models that blend discrete state changes with continuous structural motion from detailed, complex all-atom models with many continuous degrees of freedom. Our starting point is to formalize the intuitive processes that structural biologists and computational biochemists use in practice. It is worthwhile to sketch out this process here before proceeding to the formal description.

The intuitive process proceeds as follows. First, a researcher selects a detailed FG model, for example an all-atom model[131] or perhaps a more highly resolved CG model such as a G model,[138] to simulate and then visualize the molecular system under investigation. Next, the researcher looks for typical large-scale motions and other interesting transitions using principal component analysis, clustering, foreknowledge of the system, or any of many other tools to aid this search.[67, 272, 370, 36] Once the large-scale motions and states are identified, the researcher will then examine their coupling with more subtle motions, perhaps matching

a mixed ENM81 to their all-atom data or analyzing their Markov chain model states[257] in terms of which CG schemes work best in each state. In this way, the important motions within each state and between each linked group of states are characterized separately. The purpose of the present section will be to describe each step in this process mathematically in a way that treats the states and models for each state together to form a single unified UCG model.

The first step is to describe an FG model. As in previous work on the MS-CG method,[154, 153, 246] an FG model is a system of n structureless point particles in three-dimensional Euclidean space with a set of positions $\mathbf{r}^n \equiv r_i$ and velocities $\dot{\mathbf{r}}^n \equiv \dot{\mathbf{r}}_i$, where i enumerates individual particles, held at constant temperature and evolving ergodically on an energy surface defined by the Hamiltonian function $h(\mathbf{r}^n, \dot{\mathbf{r}}^n)$:

$$h(\mathbf{r}^n, \dot{\mathbf{r}}^n) = t(\dot{\mathbf{r}}^n) + u(\mathbf{r}^n) \tag{11.1}$$

where $t(\dot{\mathbf{r}}^n)$ and $u(rn)$ are kinetic and potential energies, respectively. Then, the classical partition function for the system can be written (ideal gas normalization is understood) as

$$z = \int d\mathbf{r}^n d\dot{\mathbf{r}}^n e^{-\beta h(\mathbf{r}^n, \dot{\mathbf{r}}^n)} \tag{11.2}$$

and the probability density at a given point in phase space is separable in the following sense

$$p(\mathbf{r}^n, \dot{\mathbf{r}}^n) = z^{-1} e^{-\beta h(\mathbf{r}^n, \dot{\mathbf{r}}^n)} = z^{-1} e^{-\beta u(\mathbf{r}^n)} e^{-\beta t(\dot{\mathbf{r}}^n)} = p_r(\mathbf{r}^n) p_{\dot{r}}(\dot{\mathbf{r}}^n) \tag{11.3}$$

The second stage is the process of identifying and assigning discrete states. In this process, one first examines a numerically simulated (or otherwise acquired) ensemble of FG configurations and then chooses a set of distinct states by picking regions of FG configuration space that exemplify qualitatively different behavior, for instance in the case of actin, a region in which a transient helix forms in the flexible D-loop and a region where it is completely

disordered. For these regions, the state can be determined unambiguously. However, in most other regions, say when the transient helix is only half-formed, the configuration will appear to be in between well-defined discrete configurational states. The problem of setting precise boundaries on gradual qualitative transitions has a long history as the sorites paradox;[180] in practice, these configurations between well-defined states are assigned somewhat fuzzily between the likely states, with slightly different assignments being possible. Instead of trying to avoid this fuzziness, the present approach formalizes it in the following way. All possible states are described as members ν of a discrete set Σ of latent (hidden) states, where ν is a label, and every configuration $\mathbf{r}^n$ of the FG model has a probability $p_\Sigma(\nu;\mathbf{r}^n)$ to be assigned to each state ν. Such probabilities are also known as membership functions. This latent variable distribution varies with both ν and $\mathbf{r}^n$ and is normalized for every value of $\mathbf{r}^n$. In the case of the actin D-loop described above, this could be accomplished by using two states labeled 'folded' and 'unfolded' and a latent state probability written as a Fermi switching function of the path collective variable '$s(\mathbf{r}^n)$' described in a previous study by Pfaendtner et al.[259]

$$p_{\{\text{folded,unfolded}\}}(\text{folded};\mathbf{r}^n) = 1/(1+e^{(s(\mathbf{r}^n)-s_t)/\sigma}) \tag{11.4}$$

$$p_{\{\text{folded,unfolded}\}}(\text{unfolded};\mathbf{r}^n) = 1/(1+e^{-(s(\mathbf{r}^n)-s_t)/\sigma}) \tag{11.5}$$

where s_t is the value of the path collective variable at the midpoint of the transition and σ is the width of the transition in collective variable coordinates. As another example, a protein that may be ligand bound or unbound (b,u) and protonated or not protonated (p,np) would have four states bp, up, bnp, unp for bound and protonated, unbound and protonated, bound and not protonated, and unbound and not protonated, respectively and if the state changes reflect local changes that are far apart on the protein, the latent state variables might be

independent so that

$$p_(bp, up, bnp, unp)(bp; \mathbf{r}^n) = p_{b,u}(b; \mathbf{r}^n) p_{p,np}(p; \mathbf{r}^n) \tag{11.6}$$

which can make specifying and reasoning about the model much simpler. The probability for classifying a configuration as protonated might be a function of the distance of the closest proton from the protonation site, and the probability for classifying a ligand as bound might be a function of the number of hydrogen bonds between the protein and the ligand. In any example, the probability of finding each state is the expectation over the FG configurations of the probability of finding it in any configuration,

$$P_\Sigma(\nu) = \int d\mathbf{r}^n p_r(\mathbf{r}^n) p_\Sigma(\nu; \mathbf{r}^n) \tag{11.7}$$

This quantified-uncertainty approach to mapping from FG description to discrete states has a long precedent in both biophysical modeling (see, e.g., Ref.[257]) and the general field of statistical inference (e.g., in logistic regression[226]), and taking advantage of smoothness in $p_\Sigma(\nu; \mathbf{r}^n)$ can be important for developing robust, practical methods for parameterizing UCG models. However, much of what will follow is also valid when $p_\Sigma(\nu; \mathbf{r}^n)$ is not smooth, so the use of state maps with no fuzziness can be seen as a special case of this framework. The state-specific FG distribution $p_r(\mathbf{r}^n) p_\Sigma(\nu; \mathbf{r}^n)/P_\Sigma(\nu)$ plays an important conceptual role in this work. As stated in the introduction, our strategy is to split up the FG configuration space into a multiplicity of state volumes, then simultaneously optimize a CG model for each of those volumes using the volume-restricted FG models as individual reference models. The normalized distribution $p_r(\mathbf{r}^n) p_\Sigma(\nu; \mathbf{r}^n)/P_\Sigma(\nu)$ is exactly a volume-restricted FG model's probability distribution, and the approach in this chapter is very closely related to force matching or relative entropy minimizing each of these state-specific FG models separately. The potential associated with these state-specific models is $u(\mathbf{r}^n) - \beta^{-1} \ln p_\Sigma(\nu; \mathbf{r}^n)$, consisting of the usual energy term and an additional state-dependent entropic term; this

quantity will be very important in UCG force matching expressions discussed later. For the example of the folded or unfolded actin D-loop, this state-specific potential is

$$u(\mathbf{r}^n) + \beta^{-1} \ln \left[1 + e^{(s(\mathbf{r}^n) - s_t)/\sigma)}\right] \tag{11.8}$$

which has a simple interpretation along its asymptotes: for the unfolded state ($s(\mathbf{r}^n) \gg s_t$) the state-specific potential is approximately $u(\mathbf{r}^n) + \beta^{-1}\sigma^{-1}(s(\mathbf{r}^n) - s_t)$, which corresponds to the usual system plus a linear restraint pulling the system back to the state transition region, while for the folded state ($s(\mathbf{r}^n)s_t$) the state-specific potential is approximately $u(\mathbf{r}^n)$, which is the usual, unperturbed potential of the system. A good state-specific CG model will have similar behavior. It will match the reference system's behavior in the state it is intended to, and it will keep the system in that state's typical configuration space until there is a state transition.

Next, with states and state map defined, each state must be analyzed in terms of a reduced set of continuous variables specific to that state. In terms of the previous development, for every value of ν the FG coordinates $\mathbf{r}^n$ are mapped onto a set of reduced variables. Though the reduced set of variables may differ between states, it is not necessary that they do for any of the results that follow. For convenience and clarity, in this paper only linear reductions from the FG variables to the CG variables for each state are considered,[246] though there are reasons to expect that generalizations to nonlinear maps will prove feasible.[80, 361]

To make sure that the final UCG model is physically intuitive, each CG degree of freedom must behave as either a particle or an internal mode of a particle under translations of the FG system, the mapped degrees of freedom must rotate properly under rotations of the FG coordinate system, and the model must contain no fully redundant degrees of freedom. Under these limitations, the CG configuration for each atomistic configuration and state assignment will be a system of N_ν linear combinations of FG degrees of freedom, with values $\mathbf{R}_{I,\nu} \equiv \mathbf{R}^{N_\nu}, \mathbf{R}_{I,\nu} = \sum_i c_{Ii,\nu} r_i$ for some scalars $c_{Ii,\nu}$, with the translation condition

requiring that $\sum_i c_{Ii,\nu} = 1$ or $\sum_i c_{Ii,\nu} = 0$ for each continuous degree of freedom index I and the condition that no variables are redundant requiring that all the $c_{Ii,\nu}$ row vectors for each I for a given state are linearly independent. The rotation condition insures that the $c_{Ii,\nu}$ are scalars rather than tensors.

At the UCG level, the physical meaning and importance of inertia, and thus CG momenta, become far less clear. In many cases of biophysical relevance, motions at the CG level are damped, frustrated, caged and otherwise perturbed by fast bath interactions enough that momentum becomes a less useful variable than simple rate of change, i.e. velocity. For this reason, we consider models with mapped velocities $\dot{\mathbf{R}}_{I,\nu} = \sum_i c_{Ii,\nu}\dot{\mathbf{r}}_i = \sum_i (c_{Ii,\nu}p_i)m_i$ instead of mapped momenta $\mathbf{P}_{I,\nu} = \sum_i c_{Ii,\nu}\mathbf{p}_i$ as in MS-CG. Following the notation used for MS-CG, both of these transformations are written in terms of mapping matrices

$$M_{R,\nu}^{N_\nu}(\mathbf{r}^n) \equiv \{\sum_i c_{Ii,\nu}\mathbf{r}_i\} \tag{11.9}$$

$$M_{R_{I,\nu}}(\mathbf{r}^n) \equiv \sum_i c_{Ii,\nu}\mathbf{r}_i \tag{11.10}$$

$$M_{R,\nu}^{N_\nu}(\dot{\mathbf{r}}^n) \equiv \{\sum_i c_{Ii,\nu}\dot{\mathbf{r}}_i\} \tag{11.11}$$

$$M_{R_{I,\nu}}(\dot{\mathbf{r}}^n) \equiv \sum_i c_{Ii,\nu}\dot{\mathbf{r}}_i \tag{11.12}$$

Though a map must be defined for each state in this scheme, and they can differ if necessary, the maps are never required to differ between states for the results of this chapter to hold.

Finally, for the given system define an apparent (effective) UCG energy function

$$H(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = T(\nu, \dot{\mathbf{R}}^{N_\nu}) + U(\nu, \mathbf{R}^{N_\nu}) \tag{11.13}$$

that depends on both state and a point in corresponding CG phase space and can be used to define a Boltzmann distribution on the full CG model space. The partition function for

the full UCG model is

$$Z = \sum_{\nu} \int d\mathbf{R}^{N_\nu} d\dot{\mathbf{R}}^{N_\nu} e^{-\beta H(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})} \tag{11.14}$$

with state-specific partition functions

$$Z_\nu = \int d\mathbf{R}^{N_\nu} d\dot{\mathbf{R}}^{N_\nu} e^{-\beta H(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})} \tag{11.15}$$

and the probability distribution for the full model is

$$P(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = Z^{-1} e^{-\beta H(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})} \tag{11.16}$$

with state-specific probability distributions

$$P_\nu(\mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = Z_\nu^{-1} e^{-\beta H(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})} = P(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) / P_\Sigma(\nu) \tag{11.17}$$

The coordinate and velocity distributions are separable given the form of the energy, allowing the definition of probabilities $P_{\nu,R}(\mathbf{R}^{N_\nu})$ and $P_{\nu,\dot{R}}(\dot{\mathbf{R}}^{N_\nu})$ such that

$$P(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = P_\Sigma(\nu) P_\nu(\mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = P_\Sigma(\nu) P_{\nu,R}(\mathbf{R}^{N_\nu}) P_{\nu,\dot{R}}(\dot{\mathbf{R}}^{N_\nu}) \tag{11.18}$$

Finally, it is also useful to define state-dependent delta functions

$$\delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) \equiv \prod_I \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - \mathbf{R}_{I,\nu}) \tag{11.19}$$

$$\delta(M_{R,\nu}^{N_\nu}(\dot{\mathbf{r}}^n) - \dot{\mathbf{R}}^{N_\nu}) \equiv \prod_I \delta(M_{R_{I,\nu}}(\dot{\mathbf{r}}^n) - \dot{\mathbf{R}}_{I,\nu}) \tag{11.20}$$

which allows for the shorthand notation

$$P(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu}) = \int d\mathbf{r}^n d\dot{\mathbf{r}}^n p_r(\mathbf{r}^n) p_{\dot{r}}(\dot{\mathbf{r}}^n) p_\Sigma(\nu; \mathbf{r}^n) \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) \delta(M_{R,\nu}^{N_\nu}(\dot{\mathbf{r}}^n) - \dot{\mathbf{R}}^{N_\nu}) \tag{11.21}$$

In summary, each point in the FG phase space $(\mathbf{r}^n, \dot{\mathbf{r}}^n)$ is assigned a latent state variable according to $p_\Sigma(\nu; \mathbf{r}^n)$ and then mapped to a point in CG phase space $(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})$ using the state-specific map $M_{R,\nu}^{N_\nu}$. The CG potential must depend on state as well as the other mapped variables. The distribution $p_\Sigma(\nu; \mathbf{r}^n)$ must be a discrete probability distribution for the latent variable ν normalized for each $\mathbf{r}^n$, though it need not be continuous as a function of $\mathbf{r}^n$, while the mapping transformations $M_{R,\nu}^{N_\nu}(\mathbf{r}^n) \equiv \sum_i c_{Ii,\nu} r_i$ must be full-rank and such that $\sum_i c_{Ii,\nu} \in 0, 1$ for each I.

Many other UCG schemes are possible, but this subset of them is particularly convenient for analytical investigation while still allowing the examination of important and practical cases. As with the MS-CG method before, we fully expect that this initial foray into the theory of UCG can be followed up with significant generalizations.[234, 95]

11.4 Consistency

Once the FG and CG models and the maps between them have been defined, one can begin to consider the multiscale relationship between the two models. One particularly important relationship is consistency, or indistinguishability. Put simply, one would ideally like to know how to make a CG model so robust that no sequence of experiments separated by long times that measure only functions of the CG variables in a single snapshot in time could ever distinguish between the FG and CG models. (The subject of dynamic consistency is beyond the scope of this paper.) A necessary and sufficient criterion for this indistinguishability is that the probability of seeing any given set of CG variables in the CG model should be the same as the probability of seeing that same set of CG variables mapped from the FG model; in

other words, the equilibrium joint probability densities for the variables $(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})$ must be identical in both models. In terms of simulation output, it means that any histograms of any combinations of $(\nu, \mathbf{R}^{N_\nu}, \dot{\mathbf{R}}^{N_\nu})$ variables from the CG and FG models should be equal between them.

The probability of seeing a given set of CG configuration variables in the atomistic model is just the combination of the probabilities of seeing that CG configuration given whatever FG configuration it is currently in, so that equality of the equilibrium joint probability densities for configuration can be written as the integral

$$P_{\nu,R}(\mathbf{R}^{N_\nu}) = \int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) p_r(\mathbf{r}^n) \tag{11.22}$$

Equation 11.22 is a necessary and sufficient criterion for the two models to be consistent written in terms of configuration distribution data; it will be referred to as the distributional consistency condition. It can also be written in terms of energies as

$$U(\nu, \mathbf{R}^{N_\nu}) = -\beta^{-1} \ln \left[\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)} \right] + C \tag{11.23}$$

where C is an arbitrary constant. Using the state-specific or volume-restricted potential form introduced earlier, it can also be written as

$$U(\nu, \mathbf{R}^{N_\nu}) = -\beta^{-1} \ln \left[\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) e^{-\beta(u(\mathbf{r}^n) + kT \ln p_\Sigma(\nu; \mathbf{r}^n))} \right] + C \tag{11.24}$$

which is intuitively related to previous CG consistency expressionsone now chooses to CG a biased FG model restricted to a specific region with bias $kT \ln p_\Sigma(\nu; \mathbf{r}^n)$ instead of the full FG model. Satisfying this consistency principle requires that the CG potential reproduce the FG probability distribution at every point in configuration space, which may require the CG potential to be quite complex. There is no general guarantee that a CG potential must be of any specific functional form,[96] or even that it must be smooth, without knowing some

details of the FG model. This equation (11.24) for the apparent CG potential will be used as the basis for a relative-entropy-based variational approach in Section 4.

Equation (11.24) can also be converted to an equation in terms of forces by taking the gradient of the logarithm of each side. In the case that the model only has one discrete state, the resulting equation would be exactly the MS-CG force-matching equation. However, unlike in the MS-CG approach, in UCG there are potentially many distinct mappings, up to one per discrete state. Each state's CG free energy surface can be found by force matching; however, because there is no gradient across state boundaries in the CG model, the constant differences between the surfaces must be estimated by another method, for instance state-only distribution matching. To define each state-dependent force field, hold the state fixed, take the logarithm of both sides, and then take the gradient with respect to any of the continuous CG variables defined in that state; the force on a CG variable $\mathbf{R}_{I,\nu}$ in state ν is described by

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} \ln p(\nu, \mathbf{R}^{N_\nu}) = \frac{\partial}{\partial \mathbf{R}_{I,\nu}} \ln \left[\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) p_r(\mathbf{r}^n) \right] \tag{11.25}$$

The right-hand side can be expressed in terms of atomistic forces by expressing the partial derivative with respect to the CG variable applied to the delta function in eq. (25) as a linear combination of partial derivatives with respect to atomistic coordinates and integrating by parts so that the derivatives act on the potential $u(\mathbf{r}^n)$ after applying the chain rule to the Boltzmann factor $p_r(\mathbf{r}^n)$.

To simplify the above expression, we start with the consistency equation with probabilities in terms of the free energies in each model,

$$Z^{-1} e^{-\beta U(\nu, \mathbf{R}^{N_\nu})} = z^{-1} \int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)} \tag{11.26}$$

Next, take the logarithm of each side to convert the multiplicative partition functions into

an additive free energy offset (recall the introductory chapter on force matching),

$$U(\nu, \mathbf{R}^{N_\nu}) = -\beta^{-1} \ln \left[\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)} \right] + \beta^{-1} \ln (z/Z) \tag{11.27}$$

then take partial derivatives of both sides with respect to the continuous variables for this state,

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} U(\nu, \mathbf{R}^{N_\nu}) = -\beta^{-1} \ln \left[\int d\mathbf{r}^n \frac{\partial}{\partial \mathbf{R}_{I,\nu}} \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)} \right] \ldots$$
$$\ldots + \beta^{-1} \ln (z/Z) \tag{11.28}$$

From the mapping transformation for this state,

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - R_I) = -\frac{1}{c_{Ii,\nu}} \frac{\partial}{\partial \mathbf{r}_i} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - \mathbf{R}_{I,\nu}) \tag{11.29}$$

for any nonzero $c_{Ii,\nu}$, and more generally,

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - R_I) = -\sum_i \frac{d_{Ii,\nu}}{c_{Ii,\nu}} \frac{\partial}{\partial \mathbf{r}_i} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - \mathbf{R}_{I,\nu}) \tag{11.30}$$

for any set of nonzero $c_{Ii,\nu}$ if $\sum_i d_{Ii,\nu} = 1$, which implies

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} U(\nu, \mathbf{R}^{N_\nu}) =$$
$$= \frac{\beta^{-1} \int d\mathbf{r}^n \left(\sum_i \frac{d_{Ii,\nu}}{c_{Ii,\nu}} \frac{\partial}{\partial \mathbf{r}_i} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - \mathbf{R}_{I,\nu}) \right) \prod_{J \neq I}^{N_\nu - 1} \delta(M_{R_{J,\nu}}(\mathbf{r}^n) - \mathbf{R}_{J,\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)}}{\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)}} \tag{11.31}$$

which can be integrated by parts to

$$\frac{\partial}{\partial \mathbf{R}_{I,\nu}} U(\nu, \mathbf{R}^{N_\nu}) = \frac{-\beta^{-1} \int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) \sum_i \left(\frac{d_{Ii,\nu}}{c_{Ii,\nu}} \frac{\partial}{\partial \mathbf{r}_i} p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)} \right)}{\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)}} \tag{11.32}$$

so long as

$$\sum_i \frac{d_{Ii,\nu}}{c_{Ii,\nu}} \frac{\partial}{\partial \mathbf{r}_i} \delta(M_{R_{I,\nu}}(\mathbf{r}^n) - R_{J,\nu}) = 0 \qquad \text{for } I \neq J \tag{11.33}$$

or, using the delta function identities again,

$$\sum_i \frac{d_{Ii,\nu}}{c_{Ii,\nu}} c_{Ji,\nu} = \delta_{IJ} \tag{11.34}$$

This system of equations implies that the transpose of the $d_{Ii,\nu} c_{Ii,\nu}$ matrix must be a right inverse of the matrix $M_{R,\nu}^{N_\nu}$, so as long as $M_{R,\nu}^{N_\nu}$ has a right inverse then the integration by parts can be done. This inverse exists for every UCG model considered in this chapter due to the earlier requirement that the rows of each $M_{R,\nu}^{N_\nu}$ are linearly independent. Representing the matrix with the elements $d_{Ii,\nu} c_{Ii,\nu}$ with a new force map symbol $M_{R,\nu}^{N_\nu +}$, the result can be written compactly as

$$\frac{\partial}{\partial \mathbf{R}^{N_\nu}}) U(\nu, \mathbf{R}^{N_\nu}) = \frac{-\beta^{-1} \int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) M_{R,\nu}^{N_\nu +} (\frac{\partial}{\partial \mathbf{r}^n} (p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)}))}{\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) e^{-\beta u(\mathbf{r}^n)}} \tag{11.35}$$

and this can in turn be simplified into a fixed-state, fixed-CG-configuration expectation:

$$\frac{\partial}{\partial \mathbf{R}^{N_\nu}}) U(\nu, \mathbf{R}^{N_\nu}) = \left\langle M_{R,\nu}^{N_\nu +} \left(\frac{\partial}{\partial \mathbf{r}^n} (u(\mathbf{r}^n) - \beta^{-1} \ln p_\Sigma(\nu; \mathbf{r}^n)) \right) \right\rangle_{\mathbf{R}^{N_\nu}, \nu} \tag{11.36}$$

where the notation

$$\langle f(\mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu} \equiv \frac{\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu;\mathbf{r}^n) p_r(\mathbf{r}^n) f(\mathbf{r}^n)}{\int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu;\mathbf{r}^n) p_r(\mathbf{r}^n)} \tag{11.37}$$

is used for the UCG configuration specific expectation for any function f of the atomistic configurational variables.

This set of equations, one for the force on each CG degree of freedom defined in each state ν, is satisfied if and only if the structural consistency equations specific to each state are also satisfied, making the two equally valid criteria for consistency between two state-dependent UCG models much as in the case of MS-CG. However, two important changes between these equations and the MS-CG equations deserve comment. First, there is a different force field for each state, as discussed earlier, and the force matching cannot specify the model force field completely because it cannot specify the state-to-state energy differences. Second, the matched forces are no longer the usual atomistic forces; instead, those forces are derivatives of an atomistic free energy that incorporates the entropy of the UCG state distribution. The emergence of this term is the reason why it is useful to allow $p_\Sigma(\nu;\mathbf{r}^n)$ to be smooth, as stated in Section 3. If it is smooth, that term can be estimated like any other force from simulation as a new entropic contribution; otherwise, the term contains divergences and cannot be estimated directly from points sampled in simulation. Instead, it must be evaluated as a surface integral on the boundary separating the regions in the FG conformational space with different values of discrete CG variables (details not shown). Though this complicates the implementation of the case where $p_\Sigma(\nu;\mathbf{r}^n)$ is not smooth, the mathematics still holds so long as $p_\Sigma(\nu;\mathbf{r}^n)$ is a distribution. In the section that follows, this consistency equation will be used as the basis for a force-matching variational principle.

Consistency in velocity space can be defined and guaranteed using similar logic. In this case, state-dependence is important for defining which velocities exist in each state, but because the state-map is independent of momentum by construction, the equations for

consistency are closer to the MS-CG results than the configurational consistency equations are. However, the results are not exactly the same, because for these low-resolution models allowing overlaps in the definitions of effective particles is expected to become increasingly useful[104, 73] and internal breathing modes can also be expected to become physically important.[162, 211]

The consistency equation in velocity space can be written as the Gaussian integral

$$P_{\nu,\dot{R}}(\dot{\mathbf{R}}^{N_\nu}) \sim \int d\dot{\mathbf{r}}^n \delta(M^{N_\nu}_{R,\nu}(\dot{\mathbf{r}}^n) - \dot{\mathbf{R}}^{N_\nu}) e^{-\beta \sum_i \frac{1}{2} m_i \dot{\mathbf{r}}_i^2} \tag{11.38}$$

which can be solved straightforwardly to obtain the result

$$P_{\nu,\dot{R}}(\dot{\mathbf{R}}^{N_\nu}) \sim e^{-\frac{1}{2}\beta \sum_I \sum_J \mu_{IJ,\nu} \dot{\mathbf{R}}_{I,\nu} \cdot \dot{\mathbf{R}}_{J,\nu}} \tag{11.39}$$

where μ_ν is the inverse of the matrix with I, J elements equal to $\sum_i c_{Ii,\nu} c_{Ji,\nu} m_i^{-1}$. This derivation recaptures the single-state, no-map-overlap case originally derived for MS-CG; in that case there is no mixing between sites and $\mu_{IJ,\nu} = \delta_{IJ}(\sum_i c^2_{Ii,\nu} m_i^{-1})^{-1}$.

While this kinetic energy is still quadratic no matter how complex the linear mapping, the site velocities can become coupled if their definitions overlap. This behavior need not be taken into account when configuration-space consistency is all that is needed, but in cases where kinetics is of interest, these couplings must be included in a consistent model. They can, however, be made to disappear by designing the map such that the site definitions are orthogonal under the coordinate system with a dot product weighted by the inverse of the atomistic masses, i.e., by specifying that

The results of this subsection define consistency between a FG model and a CG model, but satisfying the configurational consistency equations may require the CG model to be highly complex. No CG method that requires potential terms that depend on each CG variable simultaneously will be of practical use; even CG models with three-body potentials are presently of limited usage.[185, 97, 230] For this reason, full consistency between models

is almost never achieved or even intended. Instead, CG practitioners use approximations to consistency and reduced sets of consistency equations in place of the exact, full consistency equations. Deriving approximate variational methods based on these equations for exact consistency is the focus of Section 6. However, before presenting approximation strategies, it makes sense to examine a few simple models in which no approximations are needed; these will demonstrate the flexibility and power of this new approach to UCG consistency without forcing us from theory into detailed numerics.

11.5 Examples of Consistent UCG Models

At this point, with consistency defined, it appears worthwhile to provide two examples to illustrate how to use these UCG consistent models to clarify essential physics of systems that could be obscured using either effective particle CG or effective state CG alone. (See technical details on these examples in the Appendix.) The first example mimics a system in which the conceptual integrity of a unique effective particle CG breaks down because each CG configuration hyperplane slices through multiple metastable basin manifolds. The second example imitates a system in which the conceptual integrity of a unique effective particle CG breaks down because each metastable basin has different slow and fast variables; the basins are each best described by different CG variables. For the sake of simplicity and to make graphical representation possible, each example starts with a two-dimensional FG model and constructs one-dimensional CG models. Only qualitative behavior of the two examples will be described in this discussion section; the quantitative details of each are available in the Appendix.

As a first example, consider a model with three particles constrained to a line with two bonded and the third not bonded to either of the others as shown in Figure 11.2. This model can be considered a very simple model of the conformational selection ligand binding mechanism, since it involves calculating the approach potential of a nonbonded species to a system with multiple thermally populated metastable states with different configurations.

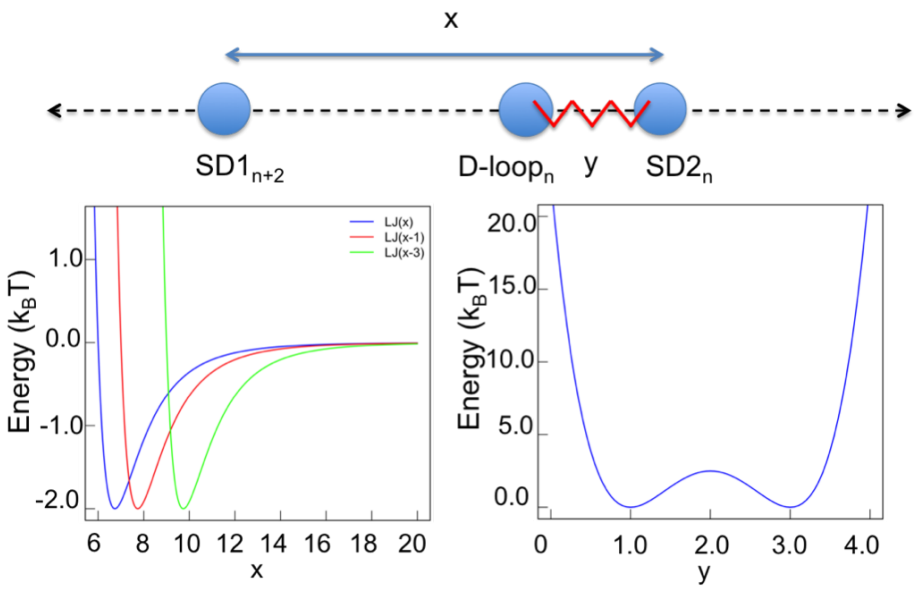

Figure 11.2: Figure 2. A three-particle system interacting through Lennard-Jones nonbonded potentials and a double well bond potential (top) and the potentials operative between them (below). The particles are labeled to suggest similarity to a conformational selection mechanism suspected to influence filamentous actin dynamics; SD1 and SD2 indicate actin subdomains 1 and 2, while subscripts indicate actin monomer position along the filament.

In this example, the two bonded particles interact with the third through a Lennard-Jones potential, while the bonded interaction is a double well potential so that there are two stable bond lengths between the particles. There is no limit to how far the third particle can be from the two bonded particles, while the bonded particles are highly constrained, so it is intuitive to CG the two bonded particles into a single effective particle, especially at long range.

However, if this CG is performed, it obscures the two-state nature of the bonded subsystem (Figure 11.3), and the resulting nonbonded potential is a counterintuitive double well with complex enthalpic and entropic contributions that may not give reasonable kinetic barriers. A UCG model without these drawbacks is easy to conceive; using the bonded particles' separation distance at the midpoint of the double-well potential as a strict dividing line between two states is a particularly natural choice. In the resulting multistate CG model, one can more easily see that the approach of the third particle to the bonded particles

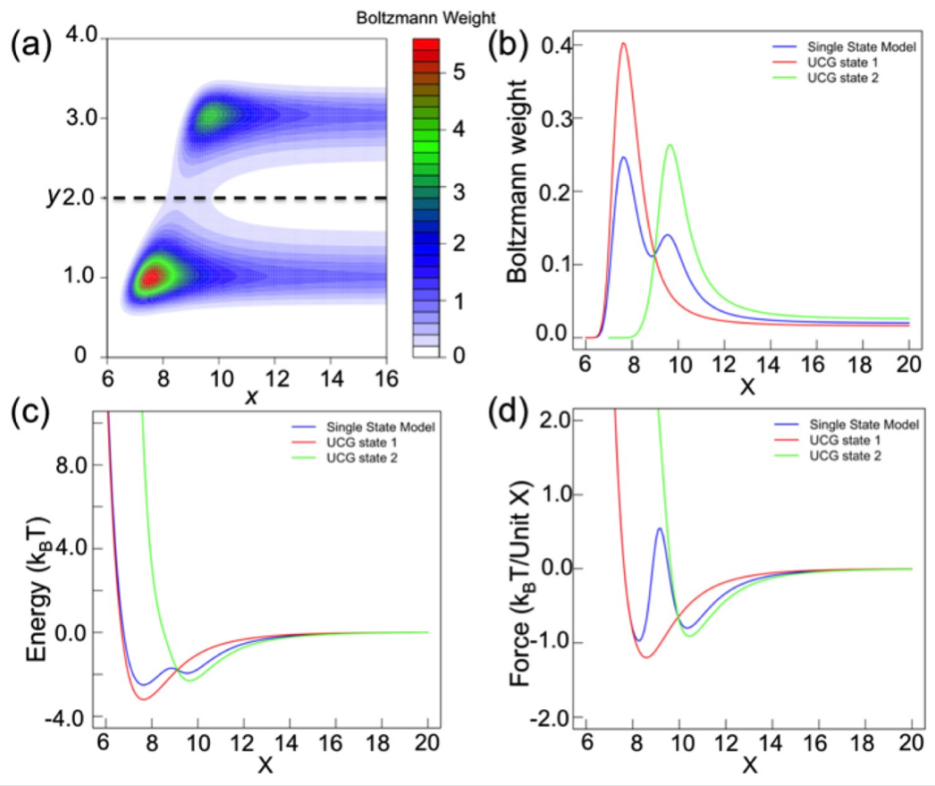

Figure 11.3: Single-state CG and multistate UCG models corresponding to the three-particle model in Figure 11.2. On the left are the FG probability distribution (top, a), with corresponding consistent CG and UCG distributions (top, b), forces (below, c), and potentials (below, d).

proceeds by barrierless approach in either state, and that a state transition to the smaller bond length is promoted as the third particle approaches the other two. In addition, one expects that the UCG force field will converge more quickly than an MS-CG force field for this example because the convergence of UCG force matching depends on convergence within states rather than between states whereas MS-CG depends on convergence of both, and the transition between states is the slowest relaxation in the system. While this example is very simple and low-dimensional, it indicates the advantages of UCG over effective particle CG for representing the physics of a simplistic model of a key mechanism in protein-protein and protein-ligand interaction, conformational selection.

The state assignment probability in the this first example is independent of the CG variable, so the state map entropic force is always mapped to zero, and thus one is free to

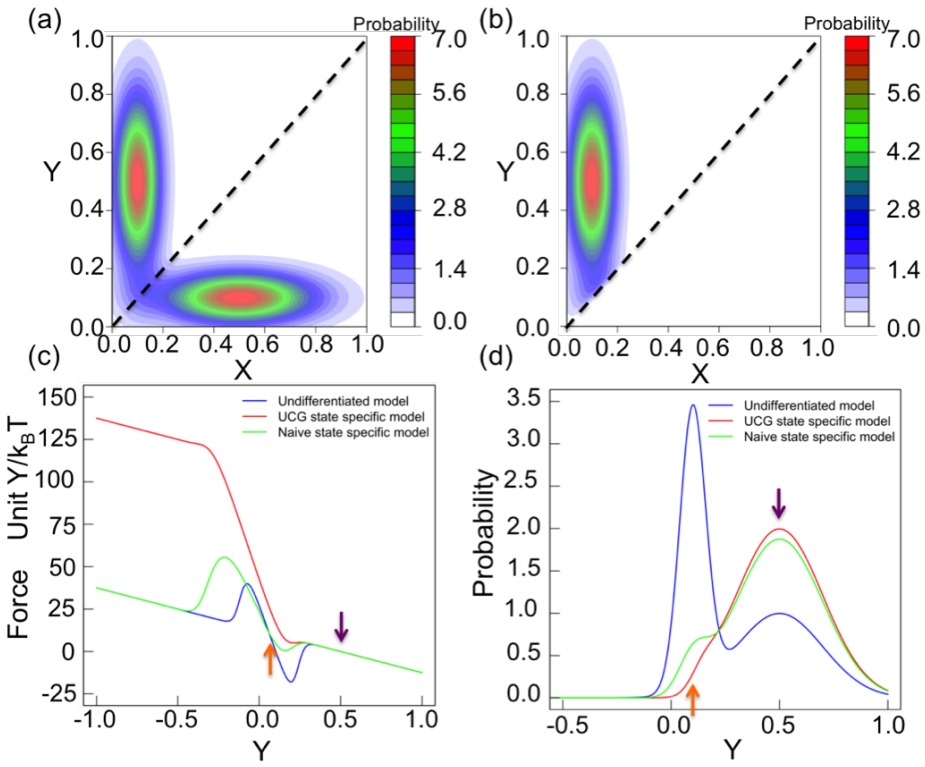

Figure 11.4: Correct description of the forces acting on UCG sites require taking into consideration the entropy contribution arising from the latent probabilities of various discrete states. Above are the FG probability distribution (top, a) and state-specific FG probability distribution (top, b), with corresponding consistent coarse-grained forces (c) and probability distributions (d) for two different CG strategies to the right. Both CG by integrating out a well-transverse variable, but one uses a single, undifferentiated state while the second uses two states. Panels (c) and (d) show the consistent CG model as the UCG state-specific model and the CG model derived without the state map entropic force term as the naive state-specific model; arrows indicate the locations of the peak maxima in the FG probability distributions.

choose a discontinuous, zero-one state assignment with no loss of convenience. The second example will consider a situation where that entropic force term is not trivial and demonstrate how each state may have different CG variables. For this example, consider two long thin wells with major axes at right angles to one another, as shown in the upper left panel of Figure 11.4. In each well, motion along the minor axis is fast and motion along the major axis is slow, so each seems to suggest CG to a variable orthogonal to the slowest CG variable in the other well. If either is chosen outright, then the CG model will appear to have a single shallow broad well and a single deep narrow well. Furthermore, the depth of

the narrow well will be set by entropy and thus highly temperature dependent and poorly transferable. Introducing a state for each well with the line separating the states as the mirror line bisecting the right angle between the wells and then coarse-graining to the major axis variable for each state results in physically more reasonable and transferable models. In this example, the state separation is not independent of the CG variable in either state, so a state map dependent force should be included in the force matching. The bottom left panel of Figure 11.4 illustrates what happens when this term is left out in the naive two-state model. Simply put, models that neglect the entropic state assignment weight force show an unphysical stabilization of configurations in the state transition region. As before, though this model is simple, the intuitive picture corresponds to important, paradigmatic models, for instance inactive smooth muscle heavy meromyosin attached to an actin filament, one head bound to actin and the other blocked, with each head fluctuating differently based on its binding state.[352, 30]

11.6 Variational Methods

The previous section defined consistency between FG and state-dependent UCG models and derived two equivalent conditions for consistency, eqs. (23) and (36). The first was written in terms of equality of distributions and the second in terms of equality of forces. This section will describe two variational approaches to choosing CG models based on FG data, one based on distributional equality and minimizing relative entropies, and the other based on force equality and minimizing mean-squared error.

The distributional consistency equations define an optimal choice of UCG distribution given data for the FG system it represents. By measuring how different the distributions in trial state-dependent UCG models are from this optimal choice, one can determine their relative merits. A natural choice for measuring differences between probability distributions

is the relative entropy,

$$S_{\text{rel},g}[W] = \sum_{\nu \in \Sigma} \int d\mathbf{r}^n p_\Sigma(\nu; \mathbf{r}^n) p_r(\mathbf{r}^n) \ln \left[\frac{p_\Sigma(\nu; \mathbf{r}^n) p_r(\mathbf{r}^n)}{P_{\nu,R}(\nu, M_{R,\nu}^{N_\nu}(\mathbf{r}^n)) g(\mathbf{r}^n; \nu, M_{R,\nu}^{N_\nu}(\mathbf{r}^n))} \right] \quad (11.40)$$

where $P_{\nu,R}(\nu, M_{R,\nu}^{N_\nu}(\mathbf{r}^n))$ is the state and configuration probability under trial CG force field $W(\nu, \mathbf{R}^{N_\nu})$ and $g(\mathbf{r}^n; \nu, M_{R,\nu}^{N_\nu}(\mathbf{r}^n))$ is a weighting function for $\mathbf{r}^n$ given $(\nu, M_{R,\nu}^{N_\nu}(\mathbf{r}^n))$. In previous work g has been chosen to be either uniform or equal to the conditional probability of seeing a FG configuration given the CG configuration, and it amounts to no more than a choice of a FG-model-specific constant term. The relative entropy has been applied as the basis of a variational principle for single-state models by Shell[296] and demonstrated to be a powerful tool for the CG of biomolecular systems.[69, 63] Shell's approach has initially been used to develop CG models with only continuous degrees of freedom, but naturally generalizes to the case of multiple latent states. The generalization is simply a variational method centered on minimizing the quantity $S_{\text{rel},g}$ as written above rather than the single-state version. This quantity differs conceptually from Shell's relative entropy in that the latent state variables must be introduced into the outer sum, so that the relative entropy written as a sum over fine-grained states is converted to a sum over fine-grained states distinguished according to imputed coarse-grained states. It is a weighted sum of the relative entropies for each of the models of the separate volumes of configuration space split out of the FG model using the state map.

The above approach to UCG shares many of the appealing qualities of the relative entropy minimization method in the single-state case. In particular, following the work of Shell et al.,[296, 69] one can readily show that this relative entropy UCG obeys the same minimization conditions and has the same attractive convexity properties when the set of trial CG potentials consists of linear combinations of basis functions.

An equally rigorous and potentially more computationally convenient variational method can be built around the force matching equations.[154, 153, 246, 247, 285] Here, optimal CG

forces are defined by the consistency condition and difference from the optimal forces is measured using a weighted mean-squared error

$$\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) \times$$
$$||\frac{\partial}{\partial \mathbf{R}^{N_\nu}} W(\nu, \mathbf{R}^{N_\nu}) - \langle M^{N_\nu +}_{R,\nu} (\frac{\partial}{\partial \mathbf{r}^n} (u(\mathbf{r}^n) - \beta^{-1} \ln p_\Sigma(\nu; \mathbf{r}^n)))\rangle_{\mathbf{R}^{N_\nu},\nu}||^2 \qquad (11.41)$$

for any arbitrary everywhere-positive normalized probability distribution $\rho(\nu, \mathbf{R}^{N_\nu})$ and trial potential $W(\nu, \mathbf{R}^{N_\nu})$. By construction, this is uniquely minimized by the consistent UCG force field. The single-state version with canonical weights in place of $\rho(\nu, \mathbf{R}^{N_\nu})$ is a natural choice and has attained notable success as the multiscale coarse-graining (MS-CG) variational principle.[154, 153, 246]

The residual χ^2 in (42) is written in terms of restrained expectations, but can easily be converted to an expectation over the entire phase space without restraints and then to an expectation over a FG simulation trajectory much like the MS-CG residual. To see this, we perform the following manipulations, first defining CG and FG force field variables for the sake of brevity

$$\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) \equiv \frac{\partial}{\partial \mathbf{R}^{N_\nu}}) W(\nu, \mathbf{R}^{N_\nu}) \qquad (11.42)$$
$$\mathbf{f}^n(\nu, \mathbf{r}^n) \equiv \frac{\partial}{\partial \mathbf{r}^n} (u(\mathbf{r}^n) - \beta^{-1} \ln p_\Sigma(\nu; \mathbf{r}^n)) \qquad (11.43)$$

Then

$$\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) ||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) - \langle M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu}||^2 \tag{11.44}$$

$$\begin{aligned}\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) (&||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu})||^2 \dots \\ &- 2\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) \cdot \langle M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu} \dots \\ &+ ||\langle M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu}||^2)\end{aligned} \tag{11.45}$$

$$\begin{aligned}\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) (&||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu})||^2 \dots \\ &- 2\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) \cdot \langle M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu} \dots \\ &+ \langle ||M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2\rangle_{\mathbf{R}^{N_\nu},\nu}) + C'\end{aligned} \tag{11.46}$$

$$\begin{aligned}\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) (\langle &||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu})||^2\rangle_{\mathbf{R}^{N_\nu},\nu} \dots \\ &- 2\langle \mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) \cdot M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)\rangle_{\mathbf{R}^{N_\nu},\nu} \dots \\ &+ \langle ||M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2\rangle_{\mathbf{R}^{N_\nu},\nu}) + C'\end{aligned} \tag{11.47}$$

$$\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) \langle ||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) - M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2\rangle_{\mathbf{R}^{N_\nu},\nu} + C' \tag{11.48}$$

$$\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) \langle ||\mathbf{G}^{N_\nu}(\nu, M^{N_\nu}_{R,\nu}(\mathbf{r}^n)) - M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2\rangle_{\mathbf{R}^{N_\nu},\nu} + C' \tag{11.49}$$

$$\begin{aligned}\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} &\rho(\nu, M^{N_\nu}_{R,\nu}(\mathbf{r}^n)) p(\mathbf{r}^n; \nu, M^{N_\nu}_{R,\nu}(\mathbf{r}^n)) \dots \\ &\times ||\mathbf{G}^{N_\nu}(\nu, M^{N_\nu}_{R,\nu}(\mathbf{r}^n)) - M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2 + C'\end{aligned} \tag{11.50}$$

$$\chi^2_\rho[W] \equiv \langle ||\mathbf{G}^{N_\nu}(\nu, M^{N_\nu}_{R,\nu}(\mathbf{r}^n)) - M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n)||^2\rangle_\rho + C' \tag{11.51}$$

The constant term C' is irrelevant for optimization and can safely be ignored. This variational principle is written in terms of an arbitrary everywhere-positive weight distribution. When FG configurations are sampled from a simulation generating a FG distri-

bution that is canonical except for possible bias in the distribution of the CG variables, $\rho_r(\mathbf{r}^n) = \rho_r(M^{N_\nu}_{R,\nu}(\mathbf{r}^n))p(\mathbf{r}^n; M^{N_\nu}_{R,\nu}(\mathbf{r}^n))$, for instance in constant NVT molecular dynamics or umbrella sampling using a CG variable as collective variable, then this expectation can be written as the time expectation

$$\chi^2_\rho[W] = \lim_{T\to\infty} \sum_{t=0}^{T} \sum_{\nu\in\Sigma} \rho_\Sigma(\nu;\mathbf{r}^n)||\mathbf{G}^{N_\nu}(M^{N_\nu}_{R,\nu}(\mathbf{r}^n_{t,\rho})) - M^{N_\nu+}_{R,\nu}\mathbf{f}^n(\nu,\mathbf{r}^n_{t,\rho})||^2 \tag{11.52}$$

where $\rho_\Sigma(\nu;\mathbf{r}^n) = \rho(\nu,\mathbf{R}^{N_\nu})/\rho_r(\mathbf{R}^{N_\nu})$ and $\mathbf{r}^n_{t,\rho}$ are samples from a simulation with an equilibrium distribution of ρ. In the case that $\rho(\nu,\mathbf{R}^{N_\nu})$ is the consistent $P_{\nu,R}(\nu,\mathbf{R}^{N_\nu})$, a natural choice corresponding to the MS-CG method, $\rho_\Sigma(\nu;\mathbf{r}^n) = p_\Sigma(\nu;\mathbf{r}^n)$. This has a notable similarity to the residual of MS-CG,[154, 153, 246] and it can be treated much the same way algorithmically, though the choice of weight is novel and has significant implications that will be commented on further in Section 7. Using a set of trial force fields that are linear combinations of state-dependent basis functions, it becomes a linear least-squares problem amenable to the same numerical techniques used in implementations of MS-CG. However, as mentioned in the previous section, it is important to remember that the variance of this expression may be infinite if the state assignment probabilities are not smooth functions of atomistic configuration.

The two error metrics $S_{rel,g}[W]$ and $\chi^2_\rho[W]$ are each uniquely minimized by the consistent UCG force field with no other global minima, and each therefore forms the basis for a rigorous variational method for deriving state-dependent UCG force fields from FG simulation data. Each can be seen as a weighted sum of metrics for the CG of individual state-specific FG models with potentials $u(\mathbf{r}^n) - \beta^{-1}\ln p_\Sigma(\nu;\mathbf{r}^n)$. Furthermore, for trial force fields that can be written as linear combinations of basis functions with the coefficients the only free parameters, each of these functionals is convex so that the global minimum is the only minimum. The two methods are complementary, with each based on a different concept of error with different implications for application and each requiring a different minimization

procedure with its own advantages and disadvantages. Relative entropy minimization is often more expensive but better able to fit structural properties, even when the trial potentials do not match the underlying model's essential physics, while force matching is often less expensive but will not fit structural properties well when the trial potentials cannot represent the physics of a model. Readers interested in choosing between the two approaches are directed to the excellent analysis provided by Rudzinski and Noid[285] and Chaimovich and Shell.[69]

11.7 Discussion

Systematic CG of atomistic models has become an increasingly important method for analyzing and simulating complex biological processes that are infeasible to study with state-of-the-art atomistic methods. Current approaches based on grouping atoms together into effective particles become unwieldy as many amino acid residues are collapsed into single particles, while approaches based on modeling conformational space as a network of discrete states represent smooth structural deformations jaggedly. This work has proposed a systematic approach to incorporating both continuous and discrete degrees of freedom in a single model for UCG. It has introduced definitions of maps, consistency and figures of merit for such CG models. Conveniently, the two methods we have introduced, one method based on relative entropy minimization and the other on force matching, retain many of the strengths of each respective method. Both lead to the same optimal, consistent, unique potential when a complete basis is used to represent the set of trial potentials, and both residuals are fully convex when the set of trial potentials is a space of linear combinations of a set of fixed-form basis functions. This work also described methods to develop models in which effective particle definitions overlap and in which some continuous degrees of freedom in the CG model transform as internal modes of a particle rather than particles; these can be applied whether or not the model incorporates discrete state transitions. Finally, the derivation of the force matching residual for this type of model makes it clear that while canonical sampling of the

FG model can be used for model development, certain types of biased sampling can also be used with equal algorithmic difficulty.

According to the development in this work, constructing a systematic UCG model should proceed according to the following steps:

1. Substantial analysis of the biomolecular peculiarities of the system under investigation, and typically a certain amount of preliminary MD simulation as well, must be performed to decide on the number and the definitions of discrete CG variables to be used.

2. Each configuration in the FG model must be assigned a set of normalized weights describing its similarity to each state present in the UCG model.

3. The groups of atoms that should be used to define each continuous degree of freedom in each state of the UCG model must be chosen. Contrary to previous work,28 it is not necessary that at least one atom be specific to each site in a given state.

4. The coefficients of each atom in each definition of a continuous degree of freedom in each state must be assigned so that the new degree of freedom transforms like a particle or an internal mode.

5. The sets of coefficients comprising each definition of each continuous degree of freedom must be linearly independent in each state.

6. If the instantaneous dynamics of the system is of interest, then kinetic energy must be assigned for each state according to the velocity consistency equation. To obtain a kinetic energy with no coupling between particles and modes, the mapped degrees of freedom must also satisfy orthogonality relations under the inverse-mass-weighted dot product.

7. An estimator for either figure of merit for UCG potentials, relative entropy or force residual, must be devised using data from the FG model.

8. The state and configuration dependent potential of the UCG model must be chosen by variationally minimizing the estimator for the chosen figure of merit over an appropriate set of trial potentials.

In principle, if these steps are followed using any set of trial functions that contains the exact potential and the estimators for the optimized figure of merit are sufficiently accurate, then the resulting UCG model will be consistent with the FG model in the sense defined in Section 3, eq. (21). However, obtaining accurate estimators for the variational minimization and choosing an appropriate set of trial force fields are major challenges for both the existing relative entropy and force matching methods.

While this method has been presented in terms of equilibrium statistics, we have been intentionally silent on the topic of finite time dynamics. The status of time in effective particle CG simulations is generally unclear except where memory functions are explicitly matched between the FG and CG models.[155, 25] Also, for the UCG models proposed in this paper molecular dynamics or even generalized Langevin dynamics will not be a feasible simulation scheme; to achieve a dynamic simulation of the full model a state-transition scheme such as kinetic Monte Carlo may perhaps be included in the integration of the equations of motion. Models that change in effective particle number will prove a further challenge to implement, since these may require dynamic particle insertion and deletion.

The behavior of the UCG method is precisely analogous to previous methods, but the set of feasible models in this UCG method is considerably larger than in its MS-CG precursor. Not only do UCG models contain both continuous and discrete degrees of freedom, the definitions of continuous degrees of freedom may overlap far more than in the development of Noid et al.,[246] the models may include both particle-like and mode-like combinations of atomic coordinates, and they can vary throughout configuration space through their state dependence. Each of these new features can be assigned to a paradigmatic application that currently resists study by existing CG models. State-dependent mappings can be motivated by the idea of using a single CG site to represent an ATP-hydrolyzing

domain, which is feasible only if a single bead can correspond to each combination of domain+ATP, domain+ADP+phosphate, domain+ADP, and the domain without nucleotide bound.[266] Overlapping site definitions arise naturally in coarse representations of proteins without clear separation between dynamic units, and have appeared in work on the optimal CG of elastic network models.[304] Internal modes, meanwhile, play an important role in the polarization of protein domains and their inclusion may be necessary to describe fluctuations in subdomains neighboring to redox active sites.[162, 25]

However, it is not necessarily the case that all of the above effects will be needed in every UCG model, and the scheme put forward here scales back gracefully so that models can be only as complex as they need to be. If desired, the configuration map for each state may be the same as every other, the maps may not have any overlapping effective particle definitions, or a model may have overlapping effective particle definitions and internal mode-like coordinates, but no states. Similarly, the state-assignment probabilities can be zero or one everywhere, with no uncertainty, though that may make force field development more challenging unless the states are chosen carefully. Each of those special cases are derived in this work implicitly as well, including rederivation of the original MS-CG and relative entropy methods. In fact, the rederivation of MS-CG contains a notable change that is worth discussing on its own.

In the foundational MS-CG theory,[246] the variational residual was defined as a squared difference between force fields at many individual points sampled from canonical simulation. In this work, by deriving a new method *ab initio* rather than trying to find a justification for an existing method, it has been shown that it is equally valid to use a squared difference between force fields at many individual points sampled from biased simulations as well. The only restrictions on this claim are that the bias must be a function of only CG variables and that it must be finite everywhere. Thus many biased-ensemble enhanced sampling methods, from grid and windowing schemes like adaptive biasing force[93] and umbrella sampling[172] to extended Lagrangian schemes such as temperature-accelerated molecular dynamics,[214]

are entirely compatible with MS-CG and can be used to generate the samples at which CG and FG models should be compared without reweighting. Because the cost of sampling the FG canonical ensemble has consistently been the highest cost of MS-CG to date,[247, 201] this has significant implications for the potential of MS-CG; investigations into its application will be described in a forthcoming paper.

Any UCG model with multiple states will define a specific number of effective CG potential surfaces. However, it is not necessary to define each surface separately in the variational minimization. For instance, it will often be interesting to define force fields so that one transition affects the force field for one protein domain while leaving the others unchanged, and another independent transition affects a different domain exclusively. In that specific case, the two transitions define four total states, but most of the trial force field will be the same in all four states and no portions of the trial force field will have more than two possible forms, so the complexity of the potential does not grow as quickly as the number of states. Since the complexity of the UCG potential is a major determinant of the cost of the variational minimization, it is important to design the state-dependence of the force field to take advantage of as many symmetries and invariances as possible.

As the examples in Section 5 demonstrate, the methods introduced in this paper provide elegant descriptions of systems that neither effective-particle-CG models nor effective-state-CG models can represent as intuitively. At present, it is not yet clear whether these techniques for systematic CG will be applicable to the study of all biological processes at the UCG scale. In particular, the approach in this paper may be unable to describe certain fractal-like structures hypothesized in many proteins,[303, 26, 277] since it is restricted to particle-, mode-, and state-like degrees of freedom. However, pure effective particle models and pure state models have each proven useful for studying substantial, complementary sets of problems, and the ways in which each loses efficacy in the UCG regime suggest that hybrid models will be able to address key open challenges in the field.

11.8 Conclusions

The UCG methods presented in this work describe means to systematically derive CG state-and-particle models from FG data, bridging two major existing research efforts into the systematic derivation of state-only and particle-only CG models. The UCG hybrid models can represent phenomena expected to play a major role in biomolecular, nanotechnological and other supramolecular systems related to the existence of multiple metastable FG wells underlying single CG particle configurations and the existence of different natural particle-CG in different regions of configuration space. Neither particle-only nor state-only models represent this physics intuitively. The preceding section has summarized the proposed methods and discussed them in depth. The sections preceding that summary motivate, describe and formally justify the new UCG methods, as well as and provide low dimensional examples as intuitive models for the failure of the state-only and particle-only models and the relative success of hybrid state-and-particle models. Furthermore, the derivation of the formal justifications for the new methods indicates the possibility of significant performance improvements for the MS-CG method based on a new, CG-configuration-weighted force-matching variational residual. In sum, this work has provided a hypothesis as to the type of physical model that can efficiently represent complex chemical and biochemical systems at the UCG level and derived formulae for developing the models necessary to test this hypothesis. In forthcoming papers we will demonstrate practical implementations of these methods, their numerical performance and their implications for the hypothesis that combining state and particle degrees of freedom together in a single CG model is necessary and sufficient to describe a wide range of key biological phenomena at the level of hundreds of atoms per CG degree of freedom.

11.9 Appendix: Example Derivations

Example 1

One of the major interactions leading to actin filament stability is suggested to be the interaction of the D-loop of monomer n with subdomain one (SD1) of the n+2 monomer. The conformation of the D-loop has been found to contain many different states including two limiting cases, folded and unfolded (see references for actin in main text). One can envision a drastic simplification to this system as the D-loop and bound subdomain two (SD2) interacting by a double well bonded potential (to mimic the folded and unfolded states) and a co-linear SD1 interacting with both the D-loop and SD2 via a Lennard-Jones potential. This system and energy functions are depicted in Figure 11.2.

The total energy of the system is

$$u(x,y) = u_{dw}(y) + u_{LJ}(x) + u_{LJ}(x-y) \tag{11.53}$$

where the individual energy functions are given by (in units of k_BT)

$$u_{dw}(y) = 5/2[(y-2)^2 - 1]^2 \tag{11.54}$$

$$u_{LJ}(x) = 4[(6/x)^1 2 - (6/x)^6] \tag{11.55}$$

Given that the total energy is a function of only two variables, one can compute the two-dimensional fine-grained (FG) Boltzmann weights of different conformations using the standard definition

$$p(x,y) \sim e^{-u(x,y)} \tag{11.56}$$

This can be computed numerically over a finite range of y for the specific energy functions given above, yielding the Boltzmann weights in Figure 11.3.

A natural coarse-graining of this system can be achieved by grouping the D-loop and SD1 of monomer n into a single particle. Integrating the FG probability function over y yields the desired one-dimensional single state CG probability distribution

$$P(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\delta(x-X)dxdy = \int_{-\infty}^{\infty} p(X,y)dy \tag{11.57}$$

$$U(X) = -k_B T \ln\left[P(X)\right] \tag{11.58}$$

$$F(X) = -dU(X)/dX \tag{11.59}$$

In practice, this integration is done over a finite range and yields the one-dimensional Boltzmann weights, energy and forces depicted as blue lines in Figure 11.3b, Figure 11.3c, and Figure 11.3d respectively. This procedure, however, leads to at least two issues. The first issue is the misleadingly low energy barrier of approximately 1 $k_B T$ between energy minima at $X = 8$ and $X = 10$. This barrier, though technically correct, is the barrier along a poorly chosen reaction coordinate and therefore may not be very informative as to the mechanism of the transition. In order to go from a minimum at $X = 8$ to a minimum at $X = 10$ the FG system would have to cross the double-well barrier of 2.5 $k_B T$. This is compensated for by the nonbonded interaction in the single-state model, conflating the effects of the bonded and nonbonded interactions obscurely. The second issue is the peak in the Single State Model Force (blue line in Figure 11.3d) at $X = 9$. This type of peak will be very difficult to resolve in a numeric force matching procedure for more elaborate systems.

A more physically intuitive CGing of this system is achieved by separating the system based on states and then integrating over y. Here a state separation at $y = 2$ seems like a natural choice due to the double well potential in y having minima at $y = 1$ and $y = 3$. State 1 will consist of all configurations with $y < 2$; state 2 will be $y > 2$. The latent state

probability functions for this choice are

$$p_\Sigma(1; X, y) = H(2 - y) \tag{11.60}$$

$$p_\Sigma(2; X, y) = H(y - 2) \tag{11.61}$$

where H is the Heaviside step function. Given these we can obtain state specific CG probabilities

$$P(1, X) = \int_{-\infty}^{\infty} p(X, y) p_\Sigma(1; X, y) dy = \int_{-\infty}^{2.0} p(X, y) dy \tag{11.62}$$

$$P(2, X) = \int_{-\infty}^{\infty} p(X, y) p_\Sigma(2; X, y) dy = \int_{2.0}^{\infty} p(X, y) dy \tag{11.63}$$

and, respectively, state specific free energies $U(\nu, X)$ and forces $F(\nu, X)$

$$U(\nu, X) = -\ln [P(\nu, X)] \text{ where } \nu \in \{1, 2\} \tag{11.64}$$

$$F(\nu, X) = -dU(\nu, X)/dX \text{ where } \nu \in \{1, 2\} \tag{11.65}$$

The CG energies for both states are plotted in Figure 11.3c. The energy difference between the minimum at $X = 8$ and $X = 10$ is approximately 2.0 $k_B T$ in this model, which is closer to the original double well barrier of 2.5 $k_B T$ and qualitatively much more understandable. Similarly the CG forces are plotted in Figure 11.3d and the peak in the single state model at $X = 9$ is not present in either state. Instead, the state specific CG model leads to more smoothly varying forces that should converge more quickly in a force matching procedure.

Example 2

As a second example, consider a two-dimensional probability distribution in which the state separation is not independent of the CG variable as it was in Example 1. One such distri-

bution is composed of two similarly shaped but orthogonally oriented anisotropic Gaussians (Figure 11.4a); in our example it is described analytically as

$$p(x,y) = N\exp[-0.5((x-0.5)^2/0.2^2 + (y-0.1)^2/0.06^2)]\ldots$$
$$\ldots + N\exp[-0.5((x-0.1)^2/0.06^2 + (y-0.5)^2/0.2^2)] \tag{11.66}$$

where N is a normalization constant ($N = 0.15$). As in the previous example, it may be useful to coarse-grain these distributions by integrating out the fast degree of freedom. Due to the orthogonal nature of the two Gaussians, however, using only a single CG dimension does not allow us to remove the high frequency motion from both Gaussians. We can, however, split the distribution by introducing state probabilities of the form

$$p_\Sigma(1;x,y) = 1/(e^{100(x-y)} + 1) \tag{11.67}$$
$$p_\Sigma(2;x,y) = 1/(e^{100(y-x)} + 1) \tag{11.68}$$

On the right-hand side of eq. (A15) there is a smoothly varying function that goes from a value of 1 in the upper triangle made by the dashed line in Figure 11.4a to zero in the lower triangle, as shown in Figure 11.4b. Multiplying eqs. (A14) and (A15) yields a state-specific probability that is pictured in Figure 11.4c. We can then integrate out the fast degree of freedom by integrating along the x direction. Similarly we can introduce a latent state probability distribution that selects only the lower Gaussian in Figure 11.4a but due to the symmetric nature of this example we will not consider this second state; the results are identical.

Now consider three ways of coarse-graining this probability distribution using force matching. The first step in each will be to compute the state independent FG force along

the y direction

$$f(y) = \frac{\partial}{\partial y} \ln p(x, y) \tag{11.69}$$

For a single state CG model, one can simply integrate this force over all x values to give the CG force

$$F_{1-\text{state}}(Y) = \frac{\int_{-\infty}^{\infty} f(y) p(x, y) \delta(y - Y) dx dy}{\int_{-\infty}^{\infty} p(x, y) \delta(y - Y) dx dy)} \tag{11.70}$$

This approach leads to the force plotted in blue in Figure 11.4d. An improvement to the single state model is to determine the expectation value of the force in the y direction for the UCG (state-dependent) probability distributions

$$F_{\text{naive}-2-\text{state}}(1, Y) = \frac{\int_{-\infty}^{\infty} f(y) p(x, y) p_{\Sigma}(1; x, y) \delta(y - Y) dx dy}{\int_{-\infty}^{\infty} p(x, y) p_{\Sigma}(1; x, y) \delta(y - Y) dx dy} \tag{11.71}$$

The resulting force is depicted as a green curve in Figure 11.4d. However, this approach does not lead to a consistent CG model. Instead, this nave approach leads to an artificial attraction to the overlap region that can be seen clearly as a shoulder around $Y = 0$ of the yellow line entitled "Naive State Specific Model" in Figure 11.4e. The reason is that eq. (A19) neglects the entropic force contribution due to the dependence of the state probability on the CG variable. The correct state-specific CG force can be determined using the UCG approach

$$F_{\text{UCG}-2-\text{state}}(1, Y) = \frac{\int_{-\infty}^{\infty} (f(y) - (\frac{d}{dy} \ln p_{\Sigma}(1; x, y)) p(x, y) p_{\Sigma}(1; x, y) \delta(y - Y) dx dy}{\int_{-\infty}^{\infty} p(x, y) p_{\Sigma}(1; x, y) \delta(y - Y) dx dy} \tag{11.72}$$

From Figure 11.4d it can be seen that the three resulting forces are drastically different. Converting back to probabilities helps to make these differences clear; these probabilities are

given in Figure 11.4e. From the one-dimensional distributions one can see that the single state model predicts a large narrow peak center at $Y = 0.1$ and a broad peak centered at $Y = 0.5$. The naive two-state model removes most of the narrow peak centered at $Y = 0.1$, but has a small shoulder in this region corresponding to attraction to the latent state boundary. The UCG method compensates for this as can be seen by the removal of the shoulder at $Y = 0.1$ in the red curve of Figure 11.4e, yielding a fully consistent state dependent model.

CHAPTER 12

ULTRA COARSE-GRAINING WHEN STATES CHANGE RARELY AND DISCRETELY

This chapter substantially reproduces the material of the paper Davtyan, A., Dama, J. F., Sinitskiy, A. V, & Voth, G. A. (2014). The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. Journal of Chemical Theory and Computation, 10(12), 52655275. Aram Davtyan implemented the theory in this paper; my contribution was its plan, from the specific Markov state transition rule and the optimization of the kinetic parameters to the form of the implementation in LAMMPS and the Voth group force matching code, and much of the writing and overall paper structure and analysis.

The previous chapter introduced models in which coarse-grained particles and bonds could change discretely, which was related to a social model in which households could be occupied by families with discretely changing members and visitors. Modeling the arrival and departure of visitors in general would be a tough problem. Moves can take weeks or months; new friends can visit often one week and then visit less frequently for a month.

We might start by considering only the visitors that arrive by plane from afar. They cannot come and go often but they don't take more than a day to make their trips. Whether or not they visit will depend on general considerations like the cost of plane tickets, but it will also depend on their preference for one home or another. Those preferences in turn are a combination of preferences for the family itself and feelings for the neighborhoods it lives in: a college student far from home might be more likely to return if she still has a crush on the boy next door.

A simple model where the plane ticket costs are random but uniformly distributed and independently sampled for all visitors to all households corresponds roughly to the rule developed in this chapter, where the rates of state transitions are given by a uniform base rate multiplied by factors accounting for the effects of interactions with neighboring particles.

12.1 Introduction

The underlying theory of ultra-coarse-graining (UCG) was introduced in a previous paper[90] and described again in the previous chapter. It defines a systematic ultra-coarse-graining procedure and obtains necessary and sufficient conditions for consistency between UCG and fine-grained (FG) representations. The formulation of this exact consistency condition, then, makes it possible to establish frameworks for deriving approximately consistent UCG models. The last chapter introduced two variational approaches by analogy to force-matching[154, 153, 246, 247] and relative entropy,[296, 297] respectively, either of which can be used to derive optimally consistent many-body UCG interaction potentials and state energies given a choice of how to assign FG structures to UCG states and configurations. The UCG modeling approach is agnostic as to whether the FG model is atomistic or itself coarse-grained at a higher resolution, making it possible to construct a hierarchy of CG models in which each model is derived from a higher-resolution model, starting from all-atom simulations and proceeding through a series of successively lower-resolution CG models.

While providing the theoretical framework, the last chapter does not provide numerical algorithms for simulating UCG models for a specific system. The current paper will fill this gap in a particular limit of time scale separation by providing the algorithmic details of a UCG implementation based on the force-matching approach. We will pay particular attention to the methodology for internal CG state sampling and to the systematic procedures used for the derivation of the UCG interaction potentials and state transition rates.

To demonstrate the features of the UCG approach we have chosen a simple system which includes both spatial and state-like motion, liquid 1,2-dichloroethane (DCE). DCE is a small molecule with well-defined long-lived dihedral conformational states and a large time scale separation between transition rates between those conformational states and the rates of other intramolecular motions and rotations. The characteristic time scale for a conformational transition between the gauche and anti states is 10 ns, while the other motions are on the scale of femtoseconds to picoseconds. As we will show, the resulting UCG model

accurately reproduces the local microstructure of the centers-of-mass of the DCE molecules as well as the key characteristics of the spontaneous conformational state dynamics that exist within the CG sites. The one-bead UCG model of DCE liquid, complemented with one internal variable per molecule, accurately captures equilibrium properties of the system that cannot be captured with a single CG site without internal states.

The remainder of this chapter is organized as follows: the Methods section starts with short recapitulation of the key UCG ideas from the previous chapter that will be needed again, discussion of state dynamics fitting and simulation methodology, and a detailed description of a UCG force matching implementation. This section also summarizes the results of preliminary all-atom simulation which motivate a particular definition for the DCE liquid UCG model. The next section presents and discusses the main results of UCG modeling of DCE liquid and compares the consistency of the results to the results of all-atom simulation. A final section concludes the paper.

12.2 Methods

12.2.1 Methodology of Ultra-Coarse-Graining

The theory of Ultra-Coarse-Graining is discussed in detail in the last chapter. Here we will briefly overview the key ideas relevant to this work. The state of a UCG model is defined by collective spatial coordinates $\mathbf{R}^{N_\nu} \equiv \{R_I\}$ and discrete variables $\nu \in \Sigma$. These discrete variables can characterize certain implicit (hidden) properties of the CG system, which cannot otherwise be deduced from the CG site representation directly. For example, ν can describe a conformational state, a ligand binding, or a chemical state. However, to avoid a possibly arbitrary definition, the value of ν is defined in a probabilistic manner. Given the FG coordinates $\mathbf{r}^n$, and a set of order parameters defined with them, we say that there is a certain probability for the system to be in each of the possible states. Those probabilities are called "membership functions", and are denoted by $p_\Sigma(\nu; \mathbf{r}^n)$. In the most general case,

the CG representation (including the number of CG sites) can vary with ν. In this paper we will only deal with the case when the number of CG sites is the same across all states of the system. However, following the notation of previous paper we will still use N_ν to denote the number of CG sites.

The last chapter defined necessary and sufficient criteria for CG and FG models to be statistically consistent with each other. Equation 12.1 below represents this criterion in terms of probability distributions for each state of the system, where $P_{R,\nu}(\mathbf{R}^{N_\nu})$ and $p_r(\mathbf{r}^n)$ are the CG and FG configurational probabilities, respectively, and $M_{R,\nu}^{N_\nu}$ are mapping operators for coordinates in each state:

$$P_{R,\nu}(\mathbf{R}^{N_\nu}) = \int d\mathbf{r}^n \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu; \mathbf{r}^n) p_r(\mathbf{r}^n) \tag{12.1}$$

This consistency condition can also be written in terms of forces acting in CG and FG representations. In Equation 12.2 below $U(\nu, \mathbf{R}^{N_\nu})$ is the CG many-body potential of mean force (PMF), $u(\mathbf{r}^n)$ is the FG interaction potential, and the $M_{R,\nu}^{N_\nu +}$ are mapping operators for forces for each state. The left-hand side represents the CG forces for a particular state of the system. Those forces should be equal to the expectations of effective forces, which consist of the FG forces minus terms accounting for the entropy of the UCG state distribution, i.e.,

$$\frac{\partial}{\partial \mathbf{R}^{N_\nu}} U(\nu, \mathbf{R}^{N_\nu}) = \left\langle M_{R,\nu}^{N_\nu +} \left(\frac{\partial}{\partial \mathbf{r}^n} (u(\mathbf{r}^n) - \beta^{-1} \ln [p_\Sigma(\nu; \mathbf{r}^n)]) \right) \right\rangle_{\mathbf{R}^{N_\nu}, \nu} \tag{12.2}$$

Besides having fundamental significance, this consistency condition provides a practical modeling scheme in the form of a variational method that can be used to derive CG interactions from FG simulation data, which are approximation to the many-body PMF $U(\nu, \mathbf{R}^{N_\nu})$. The basic idea behind this method is to minimize the mean-squared error between forces acting in CG and FG representations. In other words, we must minimize the following

expression among trial CG potentials W:

$$\chi^2_\rho[W] = \sum_{\nu\in\Sigma} \int d\mathbf{R}^{N_\nu} \rho(\nu, \mathbf{R}^{N_\nu}) ||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) - \langle M^{N_\nu +}_{R,\nu}(\mathbf{f}^n(\nu, \mathbf{r}^n))\rangle_{\mathbf{R}^{N_\nu},\nu}||^2 \tag{12.3}$$

$$\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}) \equiv \frac{\partial}{\partial \mathbf{R}^{N_\nu}} W(\nu, \mathbf{R}^{N_\nu}) \tag{12.4}$$

$$\mathbf{f}^n(\nu, \mathbf{r}^n) \equiv \frac{\partial}{\partial \mathbf{r}^n} \left(u(\mathbf{r}^n) - \beta^{-1} \ln \left[p_\Sigma(\nu; \mathbf{r}^n) \right] \right) \tag{12.5}$$

Here, $\rho(\nu, \mathbf{R}^{N_\nu})$ is an arbitrary everywhere-positive normalized probability distribution. For the canonical ensemble and for the choice of $\rho(\nu, \mathbf{R}^{N_\nu}) = P_{R,\nu}(\nu, \mathbf{R}^{N_\nu})$ we can rewrite the above equation in terms of the time expectation of an unbiased simulation in the following way:

$$\chi^2_\rho[W] = \lim_{T\to\infty} \frac{1}{3NT} \sum_{t=0}^{T} \sum_{\nu\in\Sigma} p_\Sigma(\nu; r^n_t) ||\mathbf{G}^{N_\nu}(\nu, \mathbf{R}^{N_\nu}_t) - M^{N_\nu +}_{R,\nu} \mathbf{f}^n(\nu, \mathbf{r}^n_t)||^2 \tag{12.6}$$

Equation 12.6 defines a least-squares problem that can be solved using various numerical methods given an anticipated form of the trial potential W. Similarly to the previous work on multiscale coarse-graining (MS-CG),[154, 153, 246, 247] one very important case is when W can be represented as a linear combination of basis functions for each given state of the system. Equation 12.7 below shows such a representation of potential W_ζ of type ζ, where the type of the potential can correspond to a particular type of interatomic interaction (i.e. bonded, angle, dihedral, or non-bonded). The outer sum is over a set of sites participating in the type of interaction ζ. The inner sum goes over the basis functions; each basis function is weighted by state-dependent coefficients λ. The w_d denote the basis functions, which are defined as functions of scalar variables x_γ corresponding to the particular interaction. These can correspond to bond lengths, angle values, or inter-bead distances for non-bonded

interactions, such that

$$W_\zeta(\nu, \mathbf{R}^{N_\nu}) = \sum_\gamma \sum_d \lambda_d(\nu) w_d(x_\gamma(\mathbf{R}^{N_\nu})) \tag{12.7}$$

Typical choices for the basis functions include step functions, B-splines, Gaussians, and polynomials of second order and higher. In that case we can use standard linear least-squares numerical techniques to find the λ coefficients and thus the interaction potential.

Another important aspect of UCG implementation involves the formulation of the state sampling procedure (i.e., state dynamics). This was not discussed in the last chapter, which dealt exclusively with the definition of UCG potentials. However, one can formulate the main criteria that the state dynamics must obey. First, the UCG state dynamics should reproduce FG equilibrium dynamics. This means they must satisfy the detailed balance criterion, and therefore the state transitions must depend on the current CG configuration. Second, the overall rate of transition in both the UCG and FG models should be the same. Other, more sophisticated requirements, for instance matching transition rates conditioning on CG configurations, can be added. However, here we will consider only these.

Taking into account what has been said above, one possible way to represent the instantaneous transition rates is as follows:

$$K_{\nu_1 \to \nu_2} = k_{\nu_1 \to \nu_2} \min\left[e^{-\beta(U(\nu_2, \mathbf{R}^{N_\nu}) - U(\nu_1, \mathbf{R}^{N_\nu})) - h}, 1\right] \tag{12.8}$$

$$K_{\nu_2 \to \nu_1} = k_{\nu_2 \to \nu_1} \min\left[e^{-\beta(U(\nu_1, \mathbf{R}^{N_\nu}) - U(\nu_2, \mathbf{R}^{N_\nu})) + h}, 1\right] \tag{12.9}$$

In Equations 12.8 and 12.9, $k_{\nu_1 \to \nu_2}$ and $k_{\nu_2 \to \nu_1}$ are constant prefactors that need to be determined in some way. The second parts of those expressions are Metropolis-Hastings-like terms that depend on the difference between the many-body PMFs in the current and candidate states. The term h is a third free constant which must also be determined. Here, the ν_1 and ν_2 represent the overall states of the system that are one transition apart. For example, if there is one discrete variable for each of the N particles in a UCG system and the

molecules make transitions between these states independently, then ν_1 and ν_2 must differ by the state of only one of the particles.

By construction Equations 12.8 and 12.9 can preserve the equilibrium distributions for each state because they depend on the energy difference between the states. The choice of the constant prefactors and the free constant h will allow us to satisfy the conditions fully. Equation 12.10 below represents the detailed balance criterion required to ensure consistency, given accurate state-specific models:

$$K_{\nu_1 \to \nu_2} / K_{\nu_2 \to \nu_1} = \frac{P(\nu_2, \mathbf{R}^{N_\nu})}{P(\nu_1, \mathbf{R}^{N_\nu})} \tag{12.10}$$

where the probability distributions $P(\nu, \mathbf{R}^{N_\nu})$ are defined as follows:

$$P(\nu, \mathbf{R}^{N_\nu}) = e^{-\beta U(\nu, \mathbf{R}^{N_\nu})} / \sum_{\nu \in \Sigma} \int d\mathbf{R}^{N_\nu} e^{-\beta U(\nu, \mathbf{R}^{N_\nu})} \tag{12.11}$$

Plugging Equations 12.8 and 12.9 into Equation 12.10 we then arrive at the following expression:

$$k_{\nu_1 \to \nu_2} / k_{\nu_2 \to \nu_1} e^{-\beta(U(\nu_2, \mathbf{R}^{N_\nu}) - U(\nu_1, \mathbf{R}^{N_\nu})) - h} = e^{-\beta(U(\nu_2, \mathbf{R}^{N_\nu}) - U(\nu_1, \mathbf{R}^{N_\nu}))} \tag{12.12}$$

Using this equation we see that $h = \ln [k_{\nu_1 \to \nu_2} / k_{\nu_2 \to \nu_1}]$, which leads to the following expressions for the transition rates:

$$K_{\nu_1 \to \nu_2} = k_{\nu_1 \to \nu_2} \min \left[k_{\nu_2 \to \nu_1} / k_{\nu_1 \to \nu_2} e^{-\beta(U(\nu_2, \mathbf{R}^{N_\nu}) - U(\nu_1, \mathbf{R}^{N_\nu}))}, 1 \right] \tag{12.13}$$

$$K_{\nu_2 \to \nu_1} = k_{\nu_2 \to \nu_1} \min \left[k_{\nu_1 \to \nu_2} / k_{\nu_2 \to \nu_1} e^{-\beta(U(\nu_1, \mathbf{R}^{N_\nu}) - U(\nu_2, \mathbf{R}^{N_\nu}))}, 1 \right] \tag{12.14}$$

The constant prefactors k can be estimated from all-atom simulations by matching the average (over the all-atom trajectory) instantaneous rates $\langle K \rangle_{AA}$ with the effective transition rates $R^{AA}_{\nu_1 \to \nu_2}$ and $R^{AA}_{\nu_2 \to \nu_1}$ calculated independently based on the same all-atom simulations

only. In other words, neglecting re-crossing effects, we need to solve the following equations, where the averaging is over the realizations of the corresponding states in the all-atom simulation:

$$R^{AA}_{\nu_1\to\nu_2} = \langle K_{\nu_1\to\nu_2}\rangle_{AA,\nu_1} \tag{12.15}$$

$$R^{AA}_{\nu_2\to\nu_1} = \langle K_{\nu_2\to\nu_1}\rangle_{AA,\nu_2} \tag{12.16}$$

where averages $\langle\ldots\rangle_{AA,\nu}$ can be calculated in the most general case using the following expression:

$$\langle A\rangle_{AA,\nu} = \frac{\int d\mathbf{r}^n A(\mathbf{r}^n) p_r(\mathbf{r}^n) p_\Sigma(\nu;\mathbf{r}^n)}{\int d\mathbf{r}^n p_r(\mathbf{r}^n) p_\Sigma(\nu;\mathbf{r}^n)} \tag{12.17}$$

At a first glance this approach may seem impractical for large composite systems (e.g. liquids), because if the state is defined by composition of a large number of individual discrete variables, properly sampling all of the possible states may become very hard. However, if those variables describe identical and statistically independent properties of different molecules or subunits, then the average can also be taken over the discrete variables themselves. This type of approximation can also be made in presence of weak dependence if the effect of surrounding is well averaged. Thus, this approach will work well for most electroneutral liquids, but will not work for highly charged non-ideal gas.

Equation 12.6 provides a rigorous method to find the state-specific interaction potentials up to a constant per state-specific potential. This makes it possible to find the forces acting in the system at any given moment in a given state, and to propagate it in space over time as long as the state remains fixed. However, for the state dynamics governed by Equations 12.13 and 12.14 we need to know the relative energies before and after the transition. Thus, in order to be able to formulate the transition dynamics in terms of the energies $U_R(\nu,\mathbf{R}^{N_\nu})$ derived through the force-matching technique, we will introduce an additional correction factor in the Metropolis-Hastings-like terms of Equations 12.13 and 12.14. We will denote

this factor by ϵ, such that

$$K_{\nu_1\to\nu_2} = k_{\nu_1\to\nu_2} \min\left[\frac{k_{\nu_2\to\nu_1}}{k_{\nu_1\to\nu_2}} e^{-\beta(U_R(\nu_2,\mathbf{R}^{N_\nu})-U_R(\nu_1,\mathbf{R}^{N_\nu})-\epsilon)}, 1\right] \tag{12.18}$$

$$K_{\nu_2\to\nu_1} = k_{\nu_2\to\nu_1} \min\left[\frac{k_{\nu_1\to\nu_2}}{k_{\nu_2\to\nu_1}} e^{-\beta(U_R(\nu_1,\mathbf{R}^{N_\nu})-U_R(\nu_2,\mathbf{R}^{N_\nu})+\epsilon)}, 1\right] \tag{12.19}$$

One possible way to define ϵ is through the following criterion, which, besides accounting for the ambiguity in zero energy level, insures the maximal surface hopping acceptance rate and least effort calculating the Metropolis-Hastings-like terms.

$$\left\langle \frac{K_{\nu_1\to\nu_2}}{k_{\nu_1\to\nu_2}} \right\rangle_{AA,\nu_1} = \left\langle \frac{K_{\nu_2\to\nu_1}}{k_{\nu_2\to\nu_1}} \right\rangle_{AA,\nu_2} \tag{12.20}$$

The most straightforward approach to find the prefactors k and the relative energies ϵ is to carry out an extensive all-atom simulation to get a system of three equations with three unknowns from these equations. However, in practice the re-crossing effects in the FG state transitions, neglected in these equations, result in non-negligible differences between the FG and CG models, and, consequently, the solutions of the prefactors k and ϵ will not immediately reproduce the equilibrium distribution of the FG system. Thus, as we will discuss later, one typically needs to perform a self-consistent optimization by running UCG simulations and comparing the effective transition rates calculated from those simulations with the ones obtained from the FG all-atom model, adjusting the parameters each time. This kind of systematic procedure insures that the equilibrium distribution of UCG system matches the all-atom one.

12.2.2 All-atom Simulations of DCE

The all-atom simulations of DCE in liquid and gas phases were carried out in GROMACS[144] using the OPLS-AA[168, 170] (Optimized Potentials for Liquid Simulations, All-Atom) force field, with long range electrostatics calculated through Ewald summation. Liquid simulations

contained 1000 DCE molecules, placed in a cubic box with periodic boundary conditions. The system was energy-minimized and then equilibrated for 2 ns, with equilibration under constant temperature (NVT), and under constant temperature and pressure (NPT) for 1 ns each. During the constant NPT run the box size and the density changed from 5.11 nm to 5.04 nm and from 1.23 g/cm^3 to 1.28 g/cm^3, respectively. After, a constant NVT production simulation was run for 10 ns. Gas phase simulations contained one molecule placed in a non-periodic box. The system was minimized and equilibrated under constant temperature for 1 ns, following which a constant temperature production simulation was run for 1 μs.

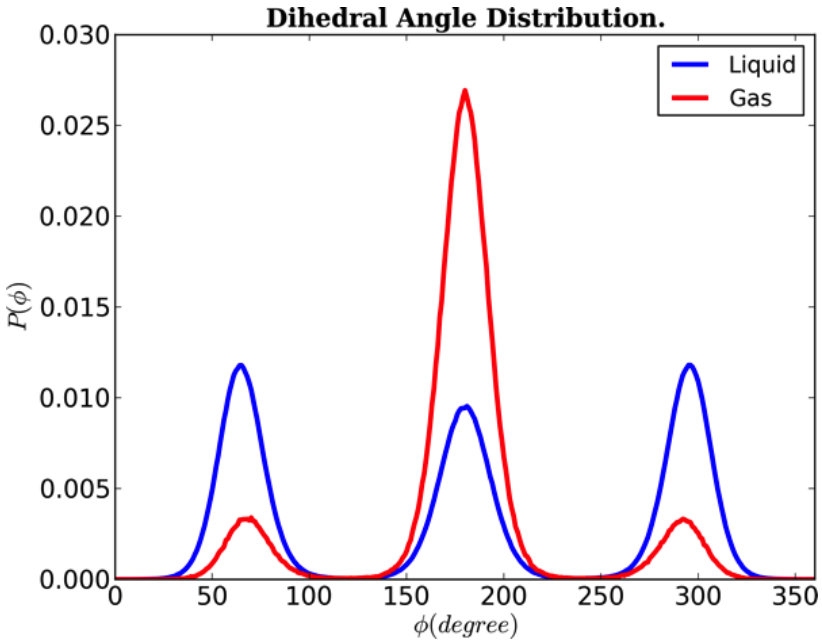

Figure 12.1: The normalized distribution of the carbon and chlorine dihedral angle of 1,2-dichloroethane for liquid and gas phases. The middle peaks correspond to the anti state, while the two peaks on the side correspond to gauche states.

For temperature coupling we used the V-Rescale thermostat, which is similar to Berendsen thermostat but has an additional stochastic term that ensures a proper canonical ensemble will be sampled.[55] For the NPT simulations a Parrinello-Rahman pressure coupling was also used. All simulations were carried out under the temperature of 298 K and the constant pressure 1 atm when applicable. A time step of 1 fs was used.Using the all-atom

simulations described above, we obtained the distribution of dihedral angle ϕ (formed by carbons and chlorines) for liquid and gas phases. Those results are summarized in Figure 12.1. As was expected, for both liquid and gas we found three peaks corresponding to the two gauche states and one anti state (see Figure 12.2a). The two important features to note here are that the peaks are well-separated and that the heights of those peaks are very different for liquid and gas phases. The former means that in this particular case there is no ambiguity in how to assign the state of the molecule, given the value of the dihedral angle ϕ. Thus, the definition of the UCG membership functions is not utterly critical as far as they distinguish between peaks corresponding to each of the states. The latter means that state distribution is quite different for liquid and gas phases. In the gas phase the anti state is dominant and occurs 81% of the time, while in the liquid state it occurs in only 32% of times. This difference illustrates the importance of many-body correlations in state dynamics.

We also calculated the transition rates between gauche and anti states. We found that for the liquid phase the transition rates from gauche to anti was 9.34 ns^{-1}, and the rate from anti to gauche was 20.12 ns^{-1}. For the gas phase those rates were found to 14.69 ns^{-1} and 3.53 ns^{-1}, respectively. The obtained rates show the clear time scale separation between the state transitions and the other molecular motions.

12.2.3 UCG model of DCE

1,2-dicholoethane is a small molecule with chemical formula C2H4Cl2. It can be found in three well-separated states defined by the value of the dihedral angle (ϕ) formed by carbons and chlorines. An anti state corresponds to ϕ values of approximately 180° and two gauche states correspond to ϕ values of approximately $\pm 65^\circ$ (see Figure 12.2a). For simplicity we will consider only two states: the anti state ($\phi \sim 180^\circ$) and a joint gauche state with $|\phi| \sim 65^\circ$. This kind of simplification is based on two assumptions: 1) the transitions between the two gauche states and the anti state occur at the same rate, 2) interactions between gauche states can be described by one potential (instead of three). The numerical calculations (based on

all-atom simulations) show that the first assumption holds almost exactly. The second one is more ambiguous, but as the results will later show, it does not greatly affect the accuracy of the derived UCG model, shown schematically in Figure 12.2b.

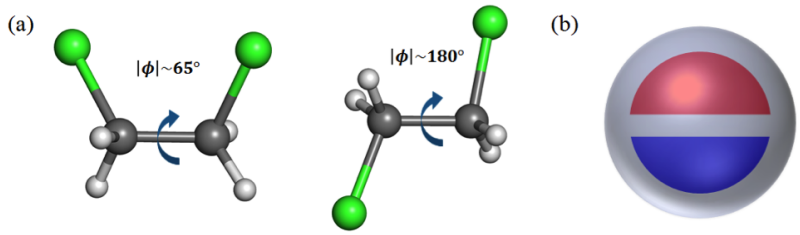

Figure 12.2: (a) 1,2-dichloroethane molecule in gauche (left) and anti (right) states; (b) schematic of the 1-site UCG DCE model. The two shades of color denote the hidden states in the single UCG site.

Given the considerations above and the results of the all-atom simulations, we have chosen to represent each DCE molecule with one bead placed at its center of mass and one internal variable corresponding to either gauche or anti states. The state of each molecule was defined in a deterministic manner using step membership functions, given by

$$p_{\Sigma,mol,I}(g;\mathbf{r}^n) = \begin{cases} 0, \text{ if } |\phi_I(\mathbf{r}^n)| > \phi_0 \\ 1, \text{ if } |\phi_I(\mathbf{r}^n)| > \phi_0 \end{cases} \tag{12.21}$$

$$p_{\Sigma,mol,I}(a;\mathbf{r}^n) = \begin{cases} 1, \text{ if } |\phi_I(\mathbf{r}^n)| > \phi_0 \\ 0, \text{ if } |\phi_I(\mathbf{r}^n)| > \phi_0 \end{cases} \tag{12.22}$$

Equation 12.21 describes the probability of the particular molecule to be in gauche state, and Equation 12.22 in anti state, respectively. Here, ϕ_I is the dihedral angle corresponding to the molecule I, and ϕ_0 is the dihedral angle value at the transition point. Based on all-atom simulations we defined ϕ_0 to be 120°. Using the per-molecule membership functions defined in Equations 12.21 and 12.22 we now define the membership functions for the whole system.

The full DCE system is composed of N_ν identical molecules, and each can be in either gauche or anti state. Therefore, as system state labels we can choose to use sets of N_ν single-molecule labels; $\nu_{sys} = \{\nu_(mol, I)\}$, where $\nu_{mol,I}$ is either g or a. Thus, we can define

the system membership function as a product of per-molecule membership functions, given by

$$p_\Sigma(\{\nu_(mol, I)\}; \mathbf{r}^n) = \prod_{I=1}^{N_\nu} p_{\Sigma,mol,I}(\nu_{mol,I}; \mathbf{r}^n) \tag{12.23}$$

According to this definition there will be 2^{N_ν} membership functions, with only one of them being equal to 1 and the rest to zero, for any given configuration of the system. So, for example, in a system of three DCE molecules where the first one is gauche and the two others are in the anti state, only $p_\Sigma(g, a, a; \mathbf{r}^n)$ will have a non-zero value of 1. In a system of 1000 identical molecules, using this sort of divide and conquer definitional scheme is essential to taming the complexity of the 2^{1000} state space.

The potential function can be written as a sum of pairwise interactions that depend on the state of the interacting CG beads and the distances between them, such that

$$U(\nu_{sys}, \mathbf{R}^{N_\nu}) = \sum_{I<J} U_{IJ}(\nu_{pair,IJ}, R_{IJ}), \tag{12.24}$$

where $\nu_{pair,IJ}$ denote the pair of per-molecule state labels $\{\nu_{mol,I}, \nu_{mol,J}\}$. Again, using state symmetries is essential to taming the 2^{1000} dimensional state space; deriving independent force fields for each would be unfeasible.

We chose to represent the pairwise potential as a linear combination of sixth order B-splines, given by

$$U(\nu_{sys}, \mathbf{R}^{N_\nu}) = \sum_{I<J} \sum_{d} \lambda_d(\nu_{pair,IJ}) u_d(R_{IJ}) \tag{12.25}$$

The $\lambda_d(\nu_{pair,IJ})$ coefficients are obtained through force-matching for every possible combination of the states of the interacting sites.

The state transitions are governed by Equations 12.13 and 12.14, where the prefactors $k_{\nu_1 \to \nu_2}$ and $k_{\nu_2 \to \nu_1}$ set the overall time scales with which the transitions between gauche

and anti states were attempted, while the Metropolis-Hastings-like terms define the acceptance rates of the corresponding transitions. An important thing to note here is that each accepted transition was immediately applied to the system and thus affected any subsequent transitions.

12.2.4 Force-matching procedure

Here we will discuss the force-matching procedure that was used to derive the interaction potential for DCE liquid. In this regards we will focus only on non-electrostatic, pairwise CG interactions. However, everything we will discuss can easily be extended for interaction potentials of the most general linear form.

Equation 12.6 defines the linear least-square problem in terms of the trial potential W. The negative gradients of this potential with regards to the positions of the UCG sites give us the forces $\mathbf{G}^{N_\nu}$ acting on those sites. In this case the W follow the form of Equation 12.25 above with sixth order B-splines basis functions as the w_d.

The outer sum in this potential goes over all pairs of sites, while the inner sum goes over the N_d basis functions. The coefficients $\lambda_d(\nu_{pair,IJ})$ depend on the states of the interacting molecules. Thus, with two possible states per molecule, we will in reality have three $\lambda_d(\nu_{pair,IJ})$ parameters for each index d. To make this more explicit, and to simplify the equations, we can replace the sum over d by a sum over all parameters λ. Therefore, we can write the interaction potential as

$$W(\nu_s ys, \mathbf{R}^{N_\nu}) = \sum_D \sum_\gamma \lambda_D w_D(R_\gamma) \tag{12.26}$$

where the inner sum goes over only the pairs in corresponding states. Also, one should note that the set of new basis functions $\{w_D\}$ will simply contain the three copies of the old basis functions w_d, specialized by state similar to how pair interactions are usually specialized by type. In this regards, it can be useful to consider the DCE liquid as a system of molecules

of two different types, with the types of the molecules dynamically changing over time.

The forces acting on UCG site I will be

$$\mathbf{G}_I(\nu, \mathbf{R}^{N_\nu}) = \sum_D \sum_\gamma \lambda_D g_D(R_\gamma)(\frac{\partial R_\gamma}{\partial \mathbf{R}_I} \tag{12.27}$$

where we define

$$g_D(R_\gamma) = -\frac{dw_D(R_\gamma)}{dR_\gamma}. \tag{12.28}$$

This can further be transformed to

$$\mathbf{G}_I(\nu, \mathbf{R}^{N_\nu}) = \sum_D \lambda_D \mathcal{G}_{I;D}(\mathbf{R}^{N_\nu}) \tag{12.29}$$

where

$$\mathcal{G}_{I;D}(\mathbf{R}^{N_\nu}) \sum_\gamma g_D(R_\gamma) \frac{\partial R_\gamma}{\partial \mathbf{R}_I} \tag{12.30}$$

Using this, we can rewrite Equation 12.6 for a finite number of sampled configurations n_t. The sum over ν reduces to a single component as only one membership function $p_\Sigma(\nu_{sys}; r_t^n)$ has a non-zero value of one for any given configuration, such that

$$\chi^2_\rho[W] = \frac{1}{3N_\nu T} \sum_{t=0}^{T} \sum_{I=1}^{N_\nu} ||M_{R,\nu}^{N_\nu +} f_I^n(\nu_{sys;t}, \mathbf{r}_t^n) - \sum_D \lambda_D \mathcal{G}_{I;D}(\nu_{sys;t}, M_{R,\nu}^{N_\nu}(\mathbf{r}_t^n))||^2 \tag{12.31}$$

The above linear least-square problem can be solved with a variety of numerical techniques. Analogously to earlier work on MS-CG,[247] we solved this least-square problem using a singular value decomposition of the corresponding normal equations. The primary difference from an MS-CG calculation was that the types of UCG sites changed dynamically from one system configuration to another.

In order to estimate the error in the UCG force field, we divided the all-atom trajectory into five equal parts and derived interaction potentials for each of those subtrajectories by solving the least-square problems defined in Equation 12.31 for each. We found that the error in the pairwise potentials did not exceed 0.1 kJ/mol for distance separations larger than 0.35 nm. For distances smaller than 0.35 nm we found the error to be much larger, presumably due to lack of sampling.[98] Thus, we fitted that portion of the pairwise potentials with functions of the form A/r^L. The final force field and potentials were constructed from the averages over the obtained five potentials for distances larger than 0.35 nm, and from fitted functions for distances smaller than 0.35 nm.

We described this force-matching procedure for a system with only non-electrostatic pairwise CG interactions. However, everything discussed above can easily be generalized for more complex interaction potential basis sets. This procedure was implemented in the Multiscale Coarse-Graining Force-Matching (MSCGFM) package[201] and can be used to derive general molecular interaction potentials including bonded potentials (distance, angle, and dihedral) and three-body non-bonded potentials.

12.2.5 Transition rate calculations and optimization

The transition rates for the all atom simulations were calculated using state correlation functions (see Figure 12.3) that are defined as the average fractions of molecules found in a particular state α at time $t+t'$ given that they were in that state at time t. In the definition below, N_α is the number of molecules in state α at time t and $\theta_{\alpha,I,t}$ is an indicator function that is unity if the molecule I is in state α at time t and zero otherwise. The sum goes over all the molecules in the system, i.e.,

$$Y_\alpha(t') \equiv \left\langle \frac{1}{N_\alpha} \sum_I \theta_{\alpha,I,t} \theta_{\alpha,I,t+t'} \right\rangle_t \tag{12.32}$$

Figure 12.3 shows the state correlation functions obtained for DCE liquid. Starting from a value of unity, these functions exponentially decay to stationary values within a few 100 ps. Knowing those stationary values and the rates of decay, one can then estimate the transition rates between the corresponding states. For this system the dynamics are simple and each of the correlation functions can be fit with a function of a form

$$y = a + (1 - a)e^{-bx}. \tag{12.33}$$

The parameters a correspond to ratios of the rates, while the parameters b correspond to their sum.

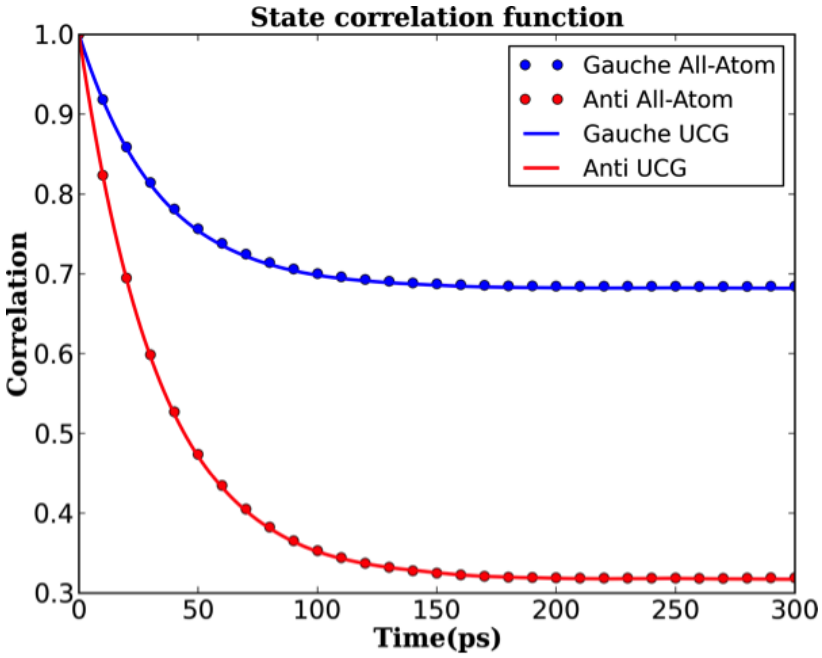

Figure 12.3: The figure shows the state correlation functions obtained from the all-atom (circles) and UCG (solid lines) simulations.

For the UCG model, the rates can be calculated simply by taking the ratio of the number

of transitions to the number of realizations of the initial state, such that

$$R^{UCG}_{\nu_1 \to \nu_2} = \frac{N_{\nu_1 \to \nu_2}}{N_{\nu_1}} \tag{12.34}$$

$$R^{UCG}_{\nu_2 \to \nu_1} = \frac{N_{\nu_2 \to \nu_1}}{N_{\nu_2}} \tag{12.35}$$

Due to the impossibility of re-crossing in the UCG simulations, the rates calculated using Equations 12.34 and 12.35 will be equal to the rates calculated using the state correlation functions as described above. This seemingly unimportant detail actually makes it straightforward to establish a systematic rate optimization procedure.

Using these ideal-case equations, one can define initial estimates of $k_{\nu_1 \to \nu_2}$, $k_{\nu_2 \to \nu_1}$, and ϵ. However, as mentioned before, those values will not exactly reproduce the equilibrium distribution of the all-atom system in the UCG simulations, so we adapted a self-consistent Robbins-Monro[278] optimization scheme which consists of the following: 1) run a 10 ns long UCG simulation using the parameters obtained from the all-atom trajectory, 2) use the last 9 ns of the simulation to calculate the effective transition rates using Equations 12.34 and 12.35, 3) compare the calculated transition rates with the ones estimated from the all-atom simulations, and then calculating the corrections to the prefactors k as described in below in Equation 12.37, 4) obtain new prefactors by mixing the corrections with the old values using a decaying mixing parameter σ, 5) use Equation 12.20 to obtain a new ϵ, 6) run new UCG simulations using the newly adjusted parameters, 7) repeat steps from 2 to 6 until the optimization converges. The relevant equations are given by

$$k^{(i+1)} = k^{(i)} + \sigma_i \Delta k^{(i)} \tag{12.36}$$

$$\Delta k^{(i)} = k^{(i)} \left(\frac{R^{AA}}{R^{UCG}_i} - 1 \right) \tag{12.37}$$

$$\sigma_i = 1/(1 + q(i-1)) \tag{12.38}$$

The last of these equations describes the mixing parameters σ_i that, starting from the

value of one at the initial iteration $i = 1$, gradually decay according to a chosen positive rate q. On the first iteration we completely accept the new values without mixing them with the previous values, but starting from the second iteration the new parameters are composed of the old parameters together with fractions of the new estimates. The decay rate q must be small enough to allow the parameters to reach their equilibrium values before saturating, but, at the same time, it must be large enough for the optimization procedure to converge in a reasonable time. In the current work we chose $q = 0.5$. Figure 12.4 shows the effective transition rate from the anti to gauche state calculated according to the above procedure over ten iterations. In the first iteration the parameters are estimated solely based on all-atom simulations. As illustrated, the optimization converges in 5 or 6 steps, reaching the desired value measured from the all-atom simulations marked on the plot by the horizontal dashed line.

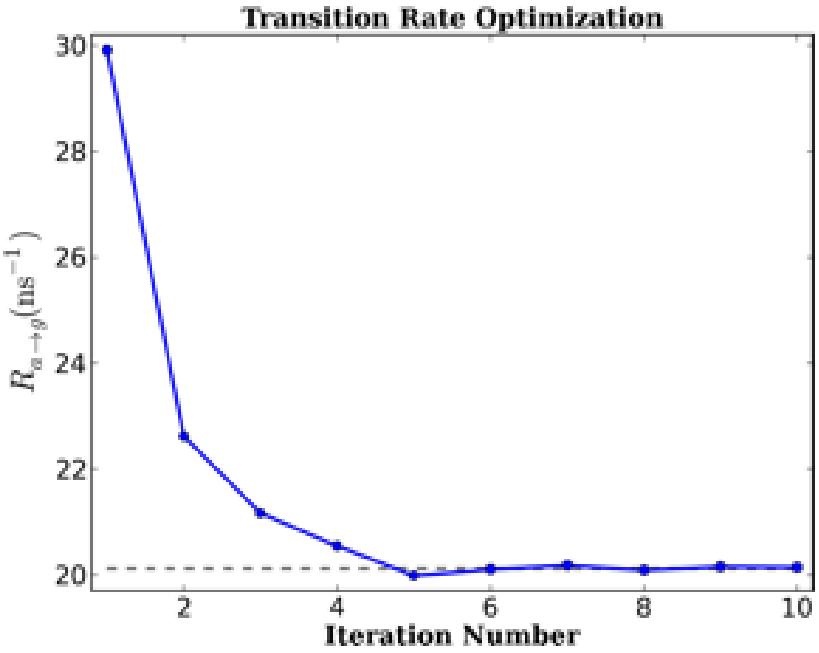

Figure 12.4: An illustration of the rate optimization procedure for the UCG anti to gauche transition rate over the course of 10 iterations. The dashed gray line indicates the rate value obtained from all-atom simulations towards which the UCG rates are optimized.

12.2.6 UCG Simulations of DCE

The UCG simulations of liquid 1,2-dicholoethane were carried out in LAMMPS[265] (http://lammps.sandia. using a newly implemented module for UCG state dynamics (UCG-SDM). This module works by changing the LAMMPS atom types of stochastically selected CG beads at each time step; changing type changes the interactions. The CG sites interact through the tabular pair potentials obtained by the force-matching procedure described above. We used a Nosé-Hoover[249, 148] thermostat to run constant NVT simulations of liquid DCE in a periodic box with the same dimensions (and number of molecules) as in the all-atom production runs. UCG runs were 10 ns long, with the first 1 ns used for equilibration and remaining 9 ns were used for production. A time step of 1 fs was used, though care should be taken in relating CG time scales to FG time scales.

Due to the reduction in the number of degrees of freedom, the UCG simulations were significantly more efficient than the FG ones. Namely, using 16 processors, FG simulations were running with the rate of 1.93 ns/hour, while for UCG simulations the rate was 16.93 ns/hour (also using 16 processors). The cost of the UCG-SDM module was found to be about 10% from the total computational time.

Caution should be taken to ensure that the jumps in energy in UCG integration do not break the thermostat and do not bring to the creation of "hot spots". We have not encountered this issue here, but this could conceivably be a problem for a different system.

12.3 Results and Discussion

To illustrate the UCG approach described above, we applied it to the molecular liquid consisting of 1000 1,2-dichloroethane (DCE) molecules described earlier. Using one CG bead per molecule to describe the spatial dynamics and one internal variable per molecule to describe the conformational dynamics, we derived a UCG model that reproduces key equilibrium properties of the DCE liquid.

Table 12.1: Parameter values for state transition equations. The $k_{g\to a}$ and $k_{a\to g}$ are the rate prefactors for gauche to anti and anti to gauche transitions, respectively. ϵ is the self-energy factor.

$k_{g\to a}$ (ns^{-1})	$k_{a\to g}$ (ns^{-1})	ϵ (kJ/mol)
83.73	20.36	0.0787

Using the force-matching procedure described above we obtained pairwise interaction potentials between UCG DCE molecules for the three possible combinations of states of the CG sites involved (see Figure 12.5). Knowing the interaction potential allowed us to then find the prefactor k and the energy correction ϵ (see Methods and Table 1). Therefore, having everything needed, we ran UCG simulations of DCE liquid using UCG-SDM in LAMMPS.

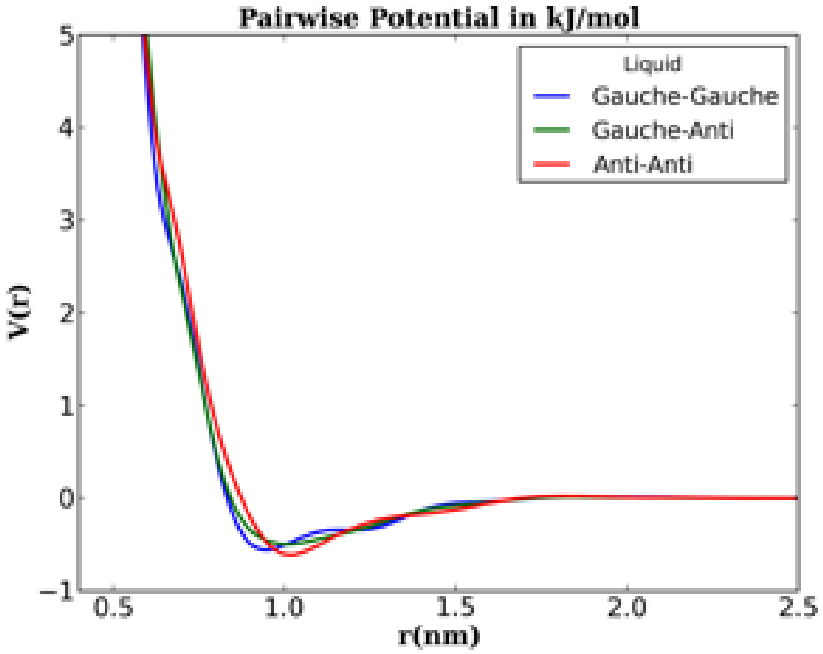

Figure 12.5: The pairwise interaction potentials between DCE beads, given the conformational state of the pair.

Figure 12.6 shows the state dynamics of liquid DCE starting from an all-anti configuration. It can be seen that the system reaches equilibrium within only a few 100 ps (shown by the vertical red bar). After that the numbers of CG molecules in gauche and anti states fluctuate around their average values. To evaluate the quality of the UCG model we also calculated the radial distribution functions (RDFs) between centers-of-mass (COM) of DCE

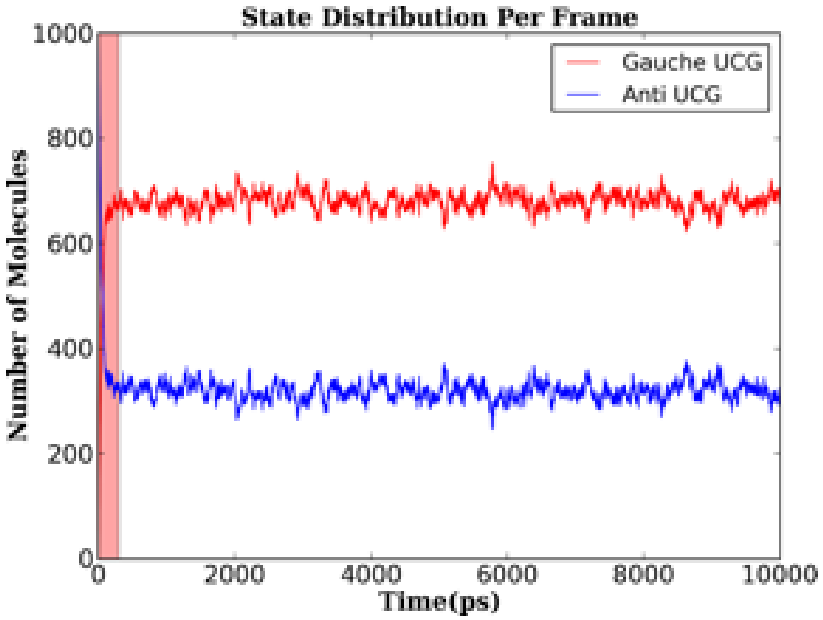

Figure 12.6: State dynamics of liquid DCE obtained from UCG simulations. All molecules were initialized in the anti state.

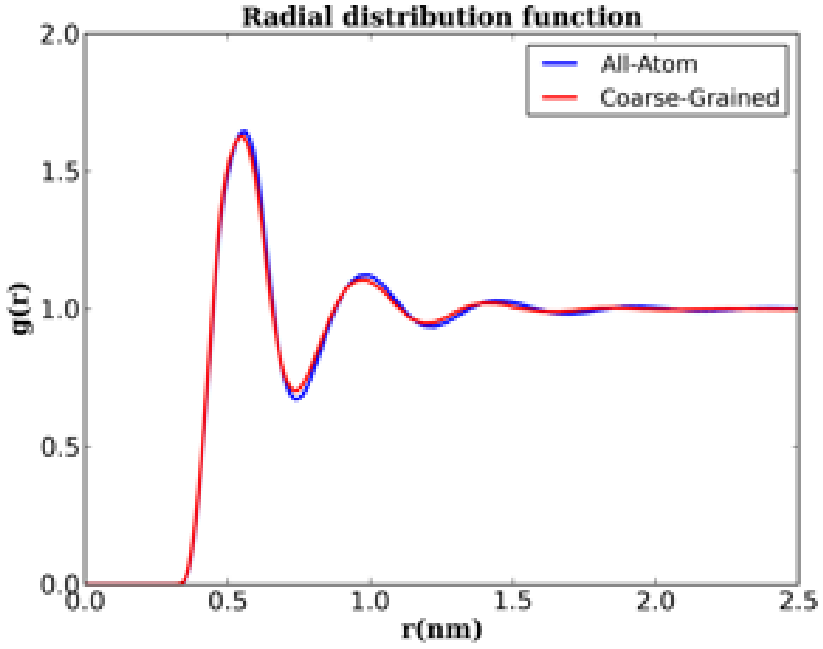

Figure 12.7: Comparison of overall center-of-mass (COM) radial distribution functions (RDFs) obtained from all-atom and UCG simulations.

molecules. Figure 12.7 compares the overall RDFs between UCG and all-atom simulations, while Figure 12.8 shows the state differentiated RDFs side by side. As one can see, the RDFs obtained from the UCG simulations capture all the major details of the all-atom

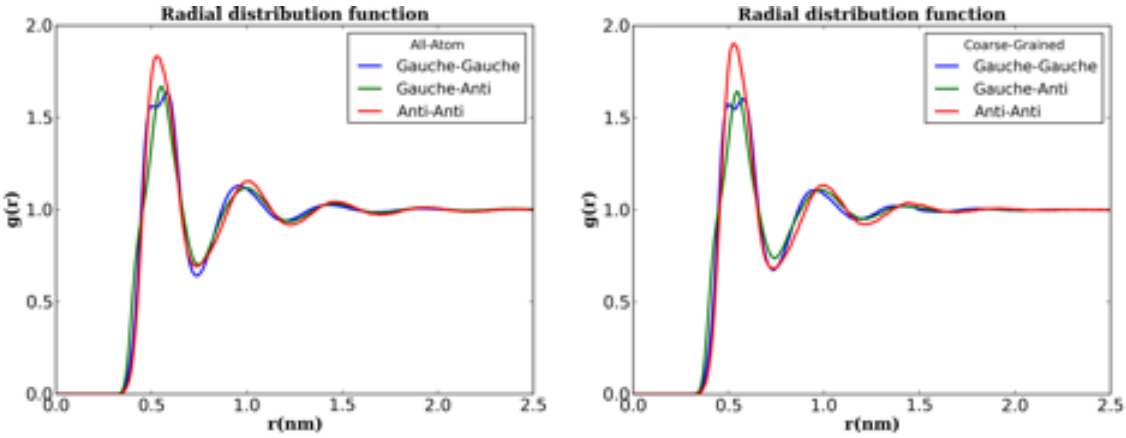

Figure 12.8: State-differentiated COM RDFs of liquid DCE obtained from all-atom (left) and UCG (right) simulations.

RDFs, including the splitting of the first peak in the gauche-gauche correlation. However, small differences exist in the locations and the magnitudes of the minimums and maximums. One possible reason for these discrepancies is neglect of small many-body effects in the UCG potential. Inclusion of three-body interactions may help to improve the agreement between the RDFs.[185] Another possibility would be to treat the DCE UCG molecules as a three state system (see Methods). Either of those two additions, however, will complicate the model while only marginally improving it.

Figure 12.9 shows the quality of agreement in the equilibrium state dynamics. As one can judge from the plots and the histograms the agreement is indeed good and the UCG simulations reproduce the average and the standard deviation of the state distribution. We also return here to the state correlation functions (see Figure 12.3) that were used to calculate the rates of transitions. Using them we found the gauche to anti and the anti to gauche transition rates in UCG simulations to be equal to 9.36 ns^{-1} and 20.12 ns^{-1}, respectively, which agree well with the rates found from the all-atom simulations (9.34 ns^{-1} and 20.12 ns^{-1}, respectively).

The results discussed above show that a one-bead UCG model that includes only pairwise interactions can adequately reproduce the main equilibrium properties of DCE liquid despite its long-lived conformational multistability and the dependence of local liquid structure on

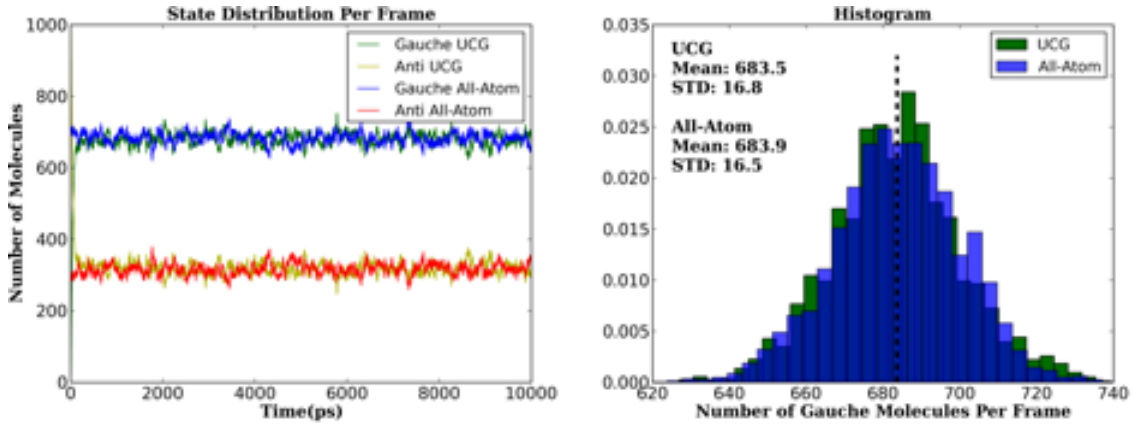

Figure 12.9: Comparison of UCG equilibrium dynamics to all-atom dynamics. The figure on the left shows the fluctuations of number of molecules in gauche and anti states over time. The figure on the right shows the histogram of the number of the molecules in the gauche state from the equilibrated part of the trajectory. In the histogram the green and blue colors correspond to the UCG and the all-atom distributions, and the dark blue corresponds to the overlap in the two distributions.

the conformation of each molecule. Nevertheless, the dihedral angle distributions obtained from all-atom simulations for the gas and liquid phases (see Figure 12.1) may imply that capturing many-body effects will be important for developing transferable models that could describe the system well over a range of physical conditions (such as density). Such a model may need to include many-body interaction terms or many-body corrections in the state dynamics equations. However, further discussion of these effects is outside the scope of this paper and will be considered in future publications.

12.4 Conclusions

In the current work we developed numerical methods to implement the methodology of ultra-coarse-graining as originally discussed in the last chapter. The addition of state dynamics made it possible to simulate a 1-site UCG model for 1,2-dichloroethane liquid. The resulting model reproduces the key equilibrium properties of the all-atom system. Namely, the COM RDFs reproduced all the major features of the corresponding all-atom functions while only slightly differing from them in the positions and magnitudes of minima and maxima. The

equilibrium state distribution also agreed well with the all-atom results, reproducing the averages of the state occupancies, their standard deviations, and the time correlation functions of the state. This is particularly important given that only the macroscopic transition rates were optimized towards the ones measured from all-atom simulations, i.e., the standard deviations were not targeted, and suggests that this modeling strategy is indeed a robust one.

In future work we plan to apply the UCG methodology to charged systems, e.g., ionic liquids and simple polymers, as well as to elaborate more on many-body effects. Inclusion of many-body correlations may also allow us to formulate more transferable models that could be applied over a range of physical conditions. Also, as large-scale conformational changes are common in biomolecular systems, it will be natural to consider modeling them using the UCG approach as well.

CHAPTER 13

ULTRA COARSE-GRAINING WHEN STATES CHANGE RAPIDLY IN QUASI-EQUILIBRIUM

This chapter comprises the meat of a paper in preparation for publication, to be submitted under the authorship of Dama, J. F., Jin, J., & Voth, G. A.; the plan and theory and a fair amount of implementation have been mine, but the simulations and many details of implementation are the work of Jaehyeok Jin, whose aid I am grateful to have.

The previous chapter considered a neighborhood model in which visitors arrive by plane from afar so they do not come and go often but their trips are short. This chapter will consider the opposite extreme, where visitors are always basically around as much as they want to be. This could be a model that tracks how young children run around a suburb to visit each other's houses every day. Actually keeping track of how all the kids prefer to distribute themselves may be an impossible task, so the parents often just let it happen.

However, they're generally able to guess where the kids will be by the concentration of families with big yards, frequent fresh cookies, or the best toys. And in fact, families considering moving into a neighborhood do frequently pay more attention to these proxies for where kids are happiest rather than spying out the local kids' actual complex and shifting preferences. This corresponds perfectly to the molecular models described below.

13.1 Motivation

Ultra coarse-grained (UCG) models could have very complex dynamics in the general case,[155, 146, 115] but systems with separation of time scales are more tractable to model. The previous chapter described an implementation of the dynamics for models with discrete states in the limit of infrequent but prompt state changes. This paper will describe a polar opposite limit of frequent and prompt state changes, i.e., quasi-equilibrium state occupation.

This limit applies whenever a molecule's interactions with its neighbors should equilibrate

rapidly to changes in environment, and for that reason will be referred to as rapidly equilibrating UCG, or RE-UCG for short. An ideal simple motivating example from the theory of liquids would be interpolation between mean-field models of a quickly-rotating polarizable molecule in vapor and liquid phases, where different effective dipoles average out to different effective pair interactions in the two phases.

This interpolation is closely related to defining a force field dependent on collective variables or order parameters, which has seen various previous use in the literature.[197, 114, 9, 322, 150] Trying to emphasize its similarities to such methods as much as possible, one might look at the probabilities of the states as the collective variables or order parameters that enter into the force field. However, the UCG approach tightly constrains the form in which these enter into the force field—through mixing parameters for various state-specific force fields of more standard form—and these constraints guarantee especially convenient implementations for both simulation and model parameterization.

Not all quasi-equilibrium limit UCG models are convenient for simulation. In particular, to compute the instantaneous quasi-equilibrium state distribution at fixed CG configuration for an arbitrary model, one would need to contruct a many-body Boltzmann average linking every particle in the system, destroying strict force locality and therefore computational efficiency. These models are most efficient, rather, when the state distributions have a locally separable form. A major concern of this paper's theoretical section will therefore be to discuss those separable forms and their modular implementation in both 1) simulation code and 2) bottom-up parameterization code, with maximal reuse of existing technology and maximal conceptual modularity.

After discussing these theoretical concerns, this paper will then discuss three suggestive examples of new UCG force fields using locally separable pair interactions. The first is a model in which interactions are modulated according to position, as if by an externally imposed screening. In the second, two force fields for two different phases are mixed via a local order parameter that discriminates the phases. In the final model, solutes aggregate

cooperatively by becoming more solvophobic upon coordinating to other solutes.

13.2 Theory

Before delving into the theory of quasi-equilibrium RE-UCG, certain definitions will be helpful to establish beforehand. First, the configuration of a CG system with N particles will be denoted $\mathbf{R}^N$. For notational simplicity, in this article we will assume a single N for all UCG states. The UCG state will be denoted by σ, and further, σ will typically be broken down into a set of substate assignments $\{s_g\}$ where g indexes the specific substates. For instance, the s_g is each particle's state when particles have states as in the previous chapter. When particles have two states each, σ could be considered to take on binary string values where each bit corresponds to the state s_I of a particular particle I. In general, substates might also inhere in bonds (pairs of CG sites) or larger groups of particles, hence the general subscript g. The representation is not usually unique; it may be entirely consistent to say, for instance, that each bond has a substate given by the combination of the substates of the two particles in the bond.

Given this, one defines the UCG configurational free energy function $U(\sigma, \mathbf{R}^N)$. In each σ, we will define this to have the form of a sum over classes c of individual energy terms between groups of particles g that interact via that class,

$$U(\sigma, \mathbf{R}^N) = \sum_{c \in C} \sum_{g \in G(c)} U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) \tag{13.1}$$

where C is the set of classes, for instance "pair nonbonded interactions, pair bonded interactions, and bonded angular interactions," $G(c)$ is the set of particle groups interacting via class c, for instance the set of neighbors or the set of bonded particles, ζ_g picks out the configuration of the particle group g from the full configuration $\mathbf{R}^N$, for instance two neighboring particle positions, ψ_c calculates the parameters of the interaction class c given a group configuration, for instance a distance or angle, and U_{c,s_g} describes the functional

form of the interaction class c when the interacting group is in substate s_g.

An unusual feature of this notation is that it assigns every group a substate, even if the UCG model does not define multiple substates for the group. In those cases, we say the number of substates is one and the substate index always takes the same value. This might seem to add unnecessary complexity at first, but in the long run, this actually simplifies analysis by putting everything on an equal footing.

In this paper, we will be particularly concerned with the effective potential that arises when one assumes that the states rapidly reach quasi-equilibrium whenever a new CG configuration is sampled; call this $U_{mix}(\mathbf{R}^N)$. It is defined up to an arbitrary constant by the free energy

$$U_{mix}(\mathbf{R}^N) = -kT \ln \sum_{\sigma} \exp(-\beta U(\sigma, \mathbf{R}^N)) \tag{13.2}$$

where $\beta = 1/kT$ as usual. This will be the primary object of interest in this investigation.

13.2.1 Separability

As stated in the introduction to this section, propagating a UCG simulation using quasi-equilibrium state occupation probabilities will not always be feasible in the general case. To make this concrete, consider the rapid mixing limit of a case like in the previous paper in this series: use two substates per bead and a different pair potential between each pair of substates. In this case

$$\begin{aligned} U(\sigma, \mathbf{R}^N) &= \sum_{c \in C} \sum_{g \in G(c)} U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) \\ &= \sum_I U_{1,s_I} + \sum_{I,J \text{ neighs}} U_{2,s_I,s_J}(R_{IJ}) \end{aligned} \tag{13.3}$$

with R_{IJ} the interparticle distance between CG sites I and J. Assuming a quasi-equilibrium distribution for the states creates a new effective potential that is a function of CG configu-

ration alone:

$$U_{mix}(\mathbf{R}^N) = -kT \ln \sum_{\sigma} \exp(-\beta U(\sigma, \mathbf{R}^N))$$
$$= -kT \ln \sum_{\sigma} \exp\left[-\beta \sum_{I} U_{1,s_I} - \beta \sum_{I,J \text{ neighs}} U_{2,s_I,s_J}(R_{IJ}) \right] \tag{13.4}$$

which is an ungainly expression to calculate. It leads to a sum of 2^N terms, each of which will lead to many-body forces with an unknown range.

Some locality could presumably be discovered in this expression and used to approximate it. However, it is simpler to instead design that locality by defining the substate assignments differently, so that the locality comes about obviously and automatically.

Specifically, consider the case that the substate assignments are statistically independent and do not depend on the UCG energy. For instance, the states might correspond to characteristic structures in two different phases so that a local parameter for the phase transition around each particle completely determines the state probabilities for those particles. Instead, they automatically shift unambiguously based on UCG configuration: there are functions $p(\sigma|\mathbf{R}^N)$ describing the probability of each σ given $\mathbf{R}^N$. In that case, the state-mixed free energy can be relatively simply written as the sum of the expected internal energy given the state distribution and the entropy of the state distribution:

$$U_{mix}(\mathbf{R}^N) = \sum_{\sigma} p(\sigma|\mathbf{R}^N) U(\sigma, \mathbf{R}^N) - kT \sum_{\sigma} p(\sigma|\mathbf{R}^N) \ln p(\sigma|\mathbf{R}^N) \tag{13.5}$$
$$= \sum_{\sigma} \left[\prod_{g \in G_{\text{base}}} p(s_g|\mathbf{R}^N) \right] \sum_{g \in G(c)} U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) \ldots$$
$$\ldots - kT \sum_{\sigma} \left[\prod_{g \in G_{\text{base}}} p(s_g|\mathbf{R}^N) \right] \ln \left[\prod_{g \in G_{\text{base}}} p(s_g|\mathbf{R}^N) \right] \tag{13.6}$$

where G_{base} is the base set of independent group substate probabilities from which all other group substate probabilities can be calculated, for instance the particle substates determine

all pair substates in this case so the set of base groups is the set of particles. Then, using the form Equation 13.3 with per-substate probabilities, this becomes

$$U_{mix}(\mathbf{R}^N) = \sum_{\sigma} \left[\prod_I p(s_I|\mathbf{R}^N) \right] \left(\sum_I U_{1,s_I} + \sum_{I,J \text{ neighs}} U_{2,s_I,s_J}(R_{IJ}) \right) \dots$$
$$\dots - kT \sum_{\sigma} \left[\prod_I p(s_I|\mathbf{R}^N) \right] \ln \left[\prod_I p(s_I|\mathbf{R}^N) \right] \tag{13.7}$$

which can be rearranged into

$$U_{mix}(\mathbf{R}^N) = \sum_I \sum_{s_I} p(s_I|\mathbf{R}^N) \left(U_{1,s_I} - kT \ln p(s_I|\mathbf{R}^N) \right) \dots$$
$$\dots + \sum_{I,J \text{ neighs}} \sum_{s_I,s_J} p(s_I|\mathbf{R}^N) p(s_J|\mathbf{R}^N) U_{2,s_I,s_J}(R_{IJ}) \tag{13.8}$$

which is now a much more tractable form. In particular, taking the derivative with respect to any particular particle position $\mathbf{R}_I$ will lead to a fully local force so long as the $p(s_I|\mathbf{R}^N)$ are local; the final expression after some simplification based on normalization of probability in the first term is

$$\nabla U_{mix}(\mathbf{R}^N) = \sum_I \sum_{s_I} \left(U_{1,s_I} - kT \ln p(s_I|\mathbf{R}^N) \right) \nabla p(s_I|\mathbf{R}^N) \dots$$
$$\dots + \sum_{I,J \text{ neighs}} \sum_{s_I,s_J} p(s_I|\mathbf{R}^N) p(s_J|\mathbf{R}^N) \nabla U_{2,s_I,s_J}(R_{IJ}) \dots$$
$$\dots + \sum_{I,J \text{ neighs}} \sum_{s_I,s_J} U_{2,s_I,s_J}(R_{IJ}) \Big[p(s_I|\mathbf{R}^N) \nabla p(s_J|\mathbf{R}^N) \ \dots$$
$$\dots + p(s_J|\mathbf{R}^N)(R_{IJ}) \nabla p(s_I|\mathbf{R}^N) \Big] \tag{13.9}$$

These simplifications are specific to the case of per-particle probabilities, but the strategy

is by no means specific to that case. In general, one can always simplify Equation 13.6 to

$$U_{mix}(\mathbf{R}^N) = \sum_{c \in C} \sum_{g \in G(c)} p(s_g|\mathbf{R}^N) U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) - kT \sum_{g \in G_{\text{base}}} p(s_g|\mathbf{R}^N) \ln p(s_g|\mathbf{R}^N) \tag{13.10}$$

and the corresponding derivatives to

$$\begin{aligned} \nabla U_{mix}(\mathbf{R}^N) = & \sum_{c \in C} \sum_{g \in G(c)} U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) \nabla p(s_g|\mathbf{R}^N) \dots \\ & \dots + \sum_{c \in C} \sum_{g \in G(c)} p(s_g|\mathbf{R}^N) \nabla U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N))) \dots \\ & \dots - kT \sum_{g \in G_{\text{base}}} \nabla p(s_g|\mathbf{R}^N) \ln p(s_g|\mathbf{R}^N) \end{aligned} \tag{13.11}$$

where the $p(s_g|\mathbf{R}^N)$ for g not in G_{base} are products of $p(s_g|\mathbf{R}^N)$ in G_{base} and their forces simplify via application of the product rule, as for the pair case. Calling the set of base groups g that define the substate probabilities of another group h $G(h)$, the specific rule is that

$$p(s_h|\mathbf{R}^N) = \prod_{g \in G(h)} p(s_g|\mathbf{R}^N) \tag{13.12}$$

$$\nabla p(s_h|\mathbf{R}^N) = \sum_{g \in G(h)} \frac{p(s_h|\mathbf{R}^N)}{p(s_g|\mathbf{R}^N)} \nabla p(s_g|\mathbf{R}^N). \tag{13.13}$$

An interesting use of this more generalized form might be to use per-pair states for non-bonded interactions to track strong hydrogen bonding or other specific interaction effects, for instance, or to use large group substates to mix elastic network models for proteins that can take on a few distinct conformational states.

Finally, before moving on, when the state distributions become extremely peaked around particular substates, the entropic term becomes very large. However, if the substate is

strongly peaked then it is natural to assume that the driving force towards that substate should be equally large. For that reason, we commonly consider force fields in which the UCG potential for the individual states counteracts the entropic contribution for that state, leaving no entropic terms in the final force field. This is equally physically meaningful so long as it is understood that one has implicitly added a strong driving force towards state demixing by canceling the entropic term.

13.2.2 Implementation

While it is interesting to write down new models for force fields for purely theoretical and mathematical reasons, in practice new models are tightly constrained by existing simulation software. The huge amount of pre-existing capital investment into common force fields is a powerful incentive to develop methodology that can reuse this capital to the greatest extent possible. In the language of evolution, generative entrenchment plays a significant role in determining which new force fields succeed or fail.

Happily, the form of Equation 13.10 is highly amenable to modular coding and substantial re-use of existing algorithms. Specifically, the form of the forces

$$\begin{aligned}\nabla U_{mix}(\mathbf{R}^N) = &\sum_{c\in C}\sum_{g\in G(c)} U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N)))\nabla p(s_g|\mathbf{R}^N)\ldots \\ &\ldots + \sum_{c\in C}\sum_{g\in G(c)} p(s_g|\mathbf{R}^N)\nabla U_{c,s_g}(\psi_c(\zeta_g(\mathbf{R}^N)))\ldots \\ &\ldots - kT\sum_{g\in G_{\text{base}}} \nabla p(s_g|\mathbf{R}^N)\ln p(s_g|\mathbf{R}^N)\end{aligned} \tag{13.14}$$

can always be calculated in three steps that essentially decouple the state assignment from force field evaluation so that the force field evaluation code requires minimal change.

In particular, the code can always follow three steps. First, one calculates all base probabilities $p(s_g|\mathbf{R}^N)$ for the g in G_{base}. Next, one calculates all forces due to the standard force fields, but scaling the terms by the probabilities found above, at the same time as

calculating the energies due to the standard force fields in order to calculate the coefficient of every $\nabla p(s_g|\mathbf{R}^N)$ for the g in G_{base} that appear after simplification of the composite (non-base) g into components according to Equation 13.13, i.e.,

$$\sum_{c \in C} \sum_{h \in G(c): g \in G(h)} U_{c,s_h}(\psi_c(\zeta_h(\mathbf{R}^N))) \frac{p(s_h|\mathbf{R}^N)}{p(s_g|\mathbf{R}^N)} \tag{13.15}$$

After those coefficients are stored, one can then calculate the $\nabla p(s_g|\mathbf{R}^N)$ for every base group and apply forces from the corresponding terms by multiplying by the stored coefficients defined just above from the previous loop.

Then the required additional memory scales as two times the sum of the number of substates per base group after the first (one array for probabilities, one for coefficients of the probability derivatives), and the calculation cost scales as the sum of the cost of calculating each of the base probabilities once, the cost of calculating each of their derivatives once, and the sum of the costs of evaluating the forcefields for each combination of substates.

For a pair potential using per-particle state assignments based on a local pair order parameter, the cost is 3 traversals of a neighbor list: the work done inside the first and last traversals scaling linearly with the number of substates and the work done inside the center one scaling as the square of the number of substates (the number of distinct pair interactions given distinct substates).

For instance, in our applications we describe cases in which the pair substate probabilities depend on a local order parameter variable defined as $W_I(\mathbf{R}^N) = \sum_{J \text{ neigh } I} w(R_{IJ})$ where $w(R)$ is a local proximity function that monotonically decreases to zero with increasing R. In this case, to calculate the forces and energy in this model, one first calculates $p(s_I) \equiv p(s_I|W_I(\mathbf{R}^N))$ by going through a neighbor loop. Next, in a second neighbor loop one accumulates the force $\sum_{I,J} \sum_{s_I,s_J} p(s_I)p(s_J)\nabla U_{2,s_I,s_J}(R_{IJ})$ simultaneously with the coefficients $\sum_{J \text{ neigh } I} \sum_{s_J} p(s_J)U_{2,s_I,s_J}(R_{IJ})$ for each I and s_I. Finally, one performs a third and last neighbor loop, calculating the $\nabla p(s_I)$ inside the loop and finding the associ-

ated forces by multiplying the resulting derivatives by the coefficients found in the previous pair loop.

We implemented this pairwise, per-particle-substates force field in LAMMPS[265, 53, 52] for any number of species with any number of substates per species interacting through nonbonded tabulated or LJ interactions, with or without entropic contributions or the one-body chemical potential term. In this code the pair interactions are specified completely as normal, expect that each LAMMPS type corresponds to a specific substate of a particle particle type instead of corresponding directly to a UCG particle type. Our code allows for states based on position within the box and local coordination of particles as sketched above. The particular definition of coordination number in the implementation uses

$$w(R) = \frac{1}{2}\left(1 + \tanh\left(\frac{R - R_{th}}{0.1R_{th}}\right)\right) \tag{13.16}$$

for some input R_{th}, so that W_I corresponds closely to the coordination number in the region $R < R_{th}$.

13.2.3 Parameterization

Lastly, the complexity of parameterizing force fields is also a major concern. However, in this case the task is relatively simple to accomplish with existing software methods using the bottom-up variational framework of the first paper in this series. It is very similar to our approach in the paper immediately preceding this one.

Specifically, consider the case that there is a fine-grained system with configuration variables $\mathbf{r}^n$ that can be mapped via some linear transformation M onto the CG configuration variables $\mathbf{R}^N$; the fine-grained potential $u(\mathbf{r}^n)$ will give rise to forces $\nabla_r u(\mathbf{r}^n)$ that will be mapped onto CG forces by an M^+ satisfying that $M(M^+)^T$ is an identity on the CG space. Then to estimate the UCG force-matching residual given state assignment probabili-

ties $p_\Sigma(\sigma|\mathbf{r}^n)$,

$$\chi^2[U(\sigma,\mathbf{R}^N)] = \sum_\sigma \langle p_\Sigma(\sigma|\mathbf{r}^n)||\nabla U(\sigma,M(\mathbf{r}^n)) - M^+(\nabla(u(\mathbf{r}^n) - kT\ln p_\Sigma(\sigma|\mathbf{r}^n))||^2\rangle_{r^n}, \tag{13.17}$$

one will approximate the expectation by a sum of T samples indexed t over the course of a sufficiently long trajectory, i.e.,

$$\chi^2[U(\sigma,\mathbf{R}^N)] = \lim_{T->\infty}\frac{1}{T}\sum_t\sum_\sigma p_\Sigma(\sigma|\mathbf{r}_t^n)||\nabla U(\sigma,M(\mathbf{r}_t^n)) - M^+(\nabla(u(\mathbf{r}_t^n) - kT\ln p_\Sigma(\sigma|\mathbf{r}_t^n))||^2. \tag{13.18}$$

However, when the number of states with nonzero probability of assignment is very large, as it will be for rapidly mixing systems, the calculation of $p_\Sigma(\sigma|\mathbf{r}_t^n)$ will be unwieldy. Therefore, instead of calculating the exact sum, we can estimate expectations over the states by randomly sampling the state some number L times per state, indexed by l, so that

$$\lim_{T->\infty}\frac{1}{T}\sum_t\frac{1}{L}\sum_l ||\nabla U(\sigma_l,M(\mathbf{r}_t^n)) - M^+(\nabla(u(\mathbf{r}_t^n) - kT\ln p_\Sigma(\sigma_l|\mathbf{r}_t^n))||^2. \tag{13.19}$$

This allows for computationally efficient force matching. We always recommend using this estimator for the residual whenever the number of nonzero state assignment probabilities is very large but not very many of those probabilities are large.

Moreover, however, when the state assignment probabilities depend on the CG variables only, i.e., when $p_\Sigma(\sigma|\mathbf{r}_t^n) = p_\Sigma(\sigma|M(\mathbf{r}_t^n))$, then the optimization becomes even simpler. The extra log probability terms will be identical to the entropic term defined above, so it is already fit analytically and can be left out of the optimization. Then the final residual estimator for

the mixed-state UCG free energy surface minus its state entropy component will be

$$\lim_{T->\infty} \frac{1}{T} \sum_t \frac{1}{L} \sum_l ||\nabla(U(\sigma_l, M(\mathbf{r}_t^n)) + kT \ln p_\Sigma(\sigma_l | M(\mathbf{r}_t^n))) - M^+(\nabla u(\mathbf{r}_t^n))||^2 \quad (13.20)$$

meaning that to fit the UCG force field, no new code at all is required for the force matching code. All one must do is feed in a trajectory in which each sampled frame of CG-mapped site positions and forces are duplicated several times and different states have been assigned to each CG site according to the correct probabilities. Those probabilities can be calculated by the simulation code, as in a LAMMPS[265, 53, 52] rerun, so that the force-matching procedure is computationally entirely independent of the state assignment scheme and the state definitions never need to be computationally implemented more than once per simulation engine.

13.3 Results

13.3.1 Interactions governed by an external field

To demonstrate the most basic features of the new RE-UCG force fields, we first developed a model in which the per-particle UCG substates were assigned based on position, as if they were controlled by an external field. Each particle was given two substates and the particles interacted via 6-12 Lennard-Jones (LJ) interactions in each, but with LJ parameters dependent on the substate. The cross-interaction followed an arithmetic mixing rule. The particle states were assigned based on the x coordinate of the particle's position in the simulation box according to

$$p(s_I = 1, x_I) = \frac{\exp(-c\cos(2\pi x_I/L))}{\exp(-c\cos(2\pi x_I/L)) + \exp(c\cos(2\pi x_I/L))} \quad (13.21)$$

$$p(s_I = 2, x_I) = 1 - p(s_I = 1, x_I) \quad (13.22)$$

where the adjustable parameter c describes the strength with which the external field modulates the particle interactions. In state 1, the particles had LJ parameters $\epsilon = 0.09$ kcal/mol, $\sigma = 2.4$ Å, and in state 2 the particles were given $\epsilon = 0.36$ kcal/mol, $\sigma = 2.0$ Å.

Each model was used to construct a system composed of 216 coarse-grained particles in a cubic box under periodic boundary conditions. The length of the simulation boxes was 18.618 (Å) and CG particles were initialized from an homogeneously box-filling Face Centered Cubic (FCC) lattice state. The simulations followed NVT dynamics at 300 K maintained via Nosé-Hoover[249, 148] thermostat. During 1 ns simulation, CG particles were simulated with a time step of 1 fs with a Verlet time integrator.[334] The following parameters c, representing state segregation strength, were tested in this way: 0.01, 0.1, 0.25, 0.5, 1.0, 2.0, 5.0, and 10.0.

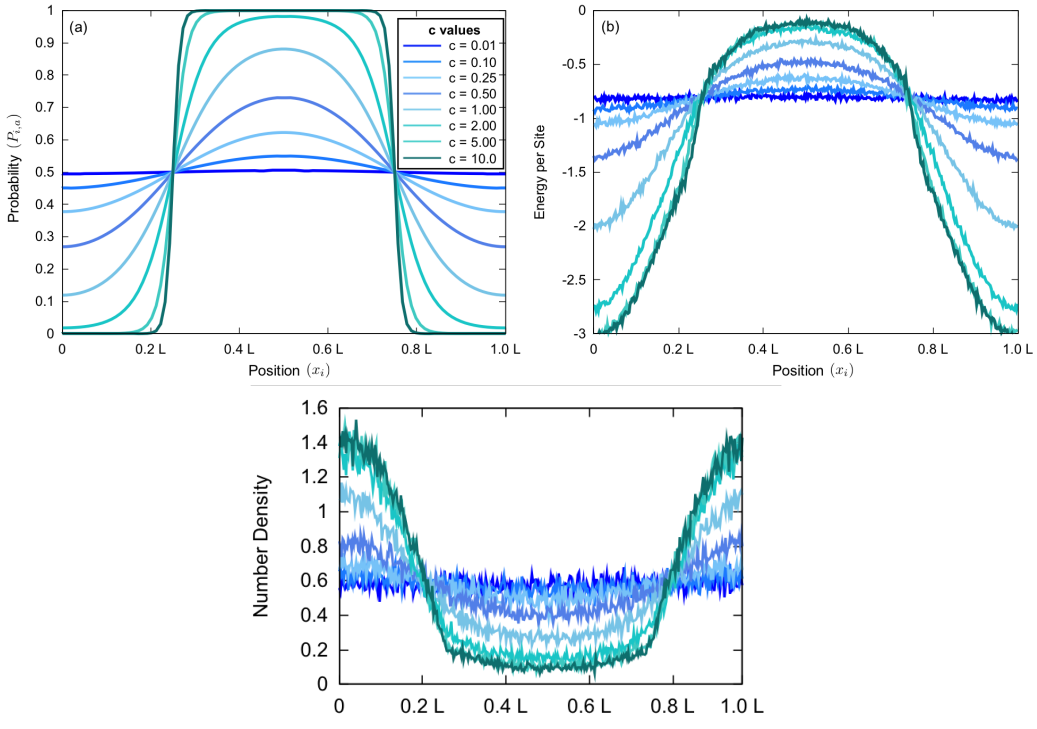

Figure 13.1: Probability of the more weakly interacting state (top left), per particle energy (top right), and particle number density (lower) as a function of x position for the position-modulated RE-UCG system described in the text given a variety of state demixing strength values (c). As c increases, the particles migrate more to the regions where they are more likely to interact more strongly.

The results are shown in Figure 13.1. The lower panel shows the density of the system as a function of x at various values of c. For low c, the particles are uniformly distributed throughout the box along the x direction, but for high c they migrate to the areas in which they have the most favorable interactions with one another. The top right panel shows the x-dependent average energy per particle as a function of c. Once again, for low c the distribution is uniform but for high c a clear split has occured. The substates do not directly affect the energy of the particles. However, by affecting their interactions, they have the effect of drawing the particles into the regions where strong favorable interactions can take place most frequently.

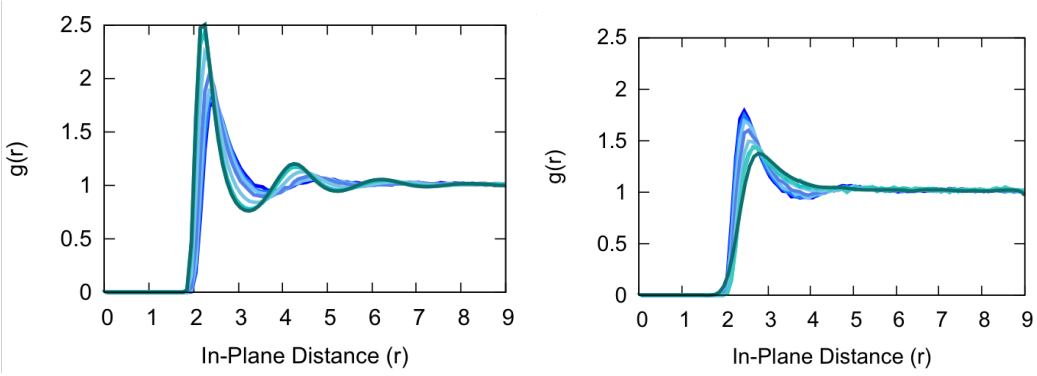

Figure 13.2: yz-plane radial distribution functions of the RE-UCG system with state probabilites determineed by x position in the most dense plane (left, $x = 0$) and least dense plane (right, $x = 0.5L$) given a variety of state segragation strengths c following the legend in Figure 13.1; as c increases, the planes become more different, with the denser plane becoming more liquid structured and the sparser plane becoming more gas like.

The states affect the interparticle interactions, and thus affect their local structure, shown in Figure 13.2. Comparing the local structure of the various systems in the yz plane at the density maximum, we see that as the interactions become stronger in this area with increasing c, and as the density rises with c, the in-plane pair correlations become more structured and more liquid-like. In contrast, at the density minimum, rising c decreases density and strength of interaction, so that the local yz-plane pair correlations become more and more vapor like.

13.3.2 Interactions changing across phase boundaries

For the second set of simulations, we chose once again to use two states per particle, but this time to modulate the particle interactions based on local coordination number instead of position in the box. It is common that pair potentials derived via bottom-up CG are not transferable across phase boundaries; this investigation is meant to be a proof of principle that this RE-UCG scheme can be used to interpolate between two force fields, each one a good description of the interactions in only one phase, in order to achieve richer phase coexistence properties than a single pair potential would be capable of generating.

The simulation box was composed of 1125 CG particles in a rectangular parallelepiped with sizes of 10.7 by 10.7 by 21.4 Åwith periodic boundary conditions applied to the x, y, and z directions. The initial configuration of CG particles was set up by constructing one half of the box (in the region $z < L_z/2$) filled with a dense FCC lattice with lattice length 1.07 Åand the other (in the region $z > L_z/2$) with a sparse fcc lattice with lattice length 2.14 Åto seed phase coexistence. This system size and these initial position lattice parameters were chosen so that the system would contain at least 1000 particles while dimensions L_x, L_y would not be smaller than 10σ. The system temperature was set at 90.6096 K, the liquid/vapor coexistence temperature $T^* = 1.0$ ($= kT/epsilon$) for an ϵ value of 0.18 kcal/mol.

In order to examine the effect of phase-dependent substates controlled via a local coordination order parameter, we first simulated a reference set of pure LJ systems with parameters of $\epsilon = 0.15$, 0.18, and 0.36 (kcal/mol) and $\sigma = 1.0$ (Å). Following this, we have designed the RE-UCG models with two substates per particle with assignment probabilities governed by the local coordination number, calculated as described in the Implementation subsection of the Theory section. The probability of each state was assigned according to a threshold rule

$$p(s_I = 1, W_I) = \frac{1}{2}(1 + \tanh(\frac{W_I - W_{th}}{0.1W_{th}})) \tag{13.23}$$

$$p(s_I = 2, W_I) = 1 - p(s_I = 1, W_I) \tag{13.24}$$

with R_{th} for the definition of W_I chosen to be 2.0 Åto fall in between the two lattice spacings used in initialization and W_{th} set to 10.0 in order to land in between the typical coordination numbers of the liquid and vapor phases for the system setup described below. Three cases of this RE-UCG setup were modeled with different mixed force fields. In each, the lower-coordination state (state 2) interacted with $\epsilon = 0.18$ kcal/mol, and σ was invariably 1.0 Åin all cases and substates.

In case 1, the more coordinated state (state 1) was given $\epsilon = 0.18$ kcal/mol, a dummy test. In case 2, state 1 was given $\epsilon = 0.64$ kcal/mol, to promote the stability of the more dense phase. In case 3, state 1 was given $\epsilon = 0.04$ kcal/mol, to suppress the stability of the more dense phase.

For each model, MD simulation was conducted for 500 ps with a time step of 1.0 fs using a Verlet time integrator[334] using a Nosé-Hoover thermostat[249, 148] (NVT) to enforce constant temperature. Furthermore, to eliminate the effect of the translation of phase boundary, we also fixed the center of mass during the simulation. The density profiles of the system shown in each figure were calculated only after the system reached equilibrium.

Figure 13.3 shows a plot of the density in the supercell as a function of slab-normal position in the box for a variety of simulation parameters. Three of the curves are pure LJ reference systems, and the other three are mixed-state systems. In the reference curves, the middle keeps the phase coexistence ϵ for this temperature, the higher is a homogenous fluid, and the lower is phase segregated. The first of the UCG systems uses the coexistence ϵ at low coordination and a lower ϵ at higher coordination, which has the effect of destabilizing the high-density liquid phase and suppressing phase coexistence. The second of the UCG systems is a dummy, in which the states have no effect, to show that the behavior is negligibly different from the reference. The final of the UCG systems, uses the coexistence ϵ at low coordination and a higher ϵ at high coordination, which has the effect of stabilizing the liquid phase and exaggerating phase segregation. However, importantly, the second UCG system enhances phase segregation very differently than simply raising ϵ would: it creates a sharper

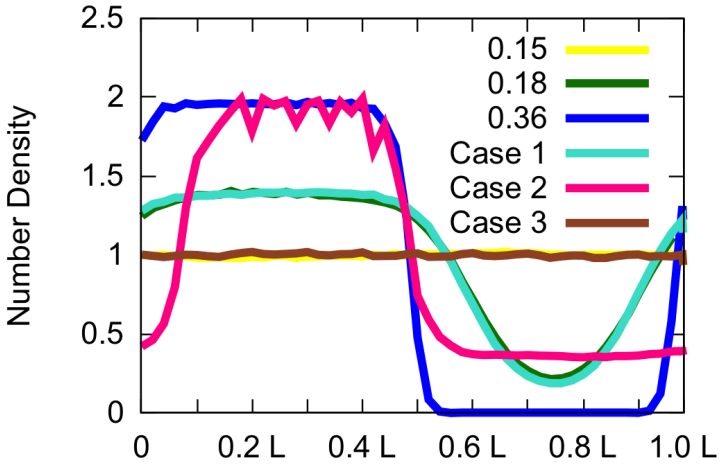

Figure 13.3: Density profiles as a function of x-position for three pure LJ systems with variable ϵ (numerical values in kcal/mol) and the three RE-UCG systems (Cases 1-3) described in the text. Each RE-UCG model gives the particles two substates controlled by local coordination number, with the low-coordination substate having an ϵ of 0.18 kcal/mol. In case 1, the two substates have identical interactions. In case 2, the high-coordination substate interacts more strongly, promoting phase separation ($\epsilon = 0.64$ kcal/mol). In case 3, the high-coordination substate interacts more weakly ($\epsilon = 0.04$ kcal/mol), suppressing phase separation.

interface given the same number of particles in the vapor phase. In other words, this UCG force field is fitting phase behavior more complex than the force field functional form for any individual state would be capable of capturing alone.

We can see the effect's origins directly in Figure 13.4, which shows the distributions of local coordination values in the two UCG systems and the LJ phase coexistence. The figure also colors the distributions by the fraction of time a particle of that local coordination would be assigned to each of the two substates. In the pure LJ system, there is a weak bimodality to the coordination number distribution: a long shoulder. In the liquid-destabilizing mixed state UCG system, that bimodality is entirely suppressed, but in the liquid-stabilizing mixed state UCG system, the bimodality is strongly enhanced, with two clear peaks.

Using states allows for finer control of the state coexistence, allowing the interactions between particles in each phase to change to suppress or enhance coexistence with different tradeoffs than the plain LJ system faces. Each phase can have a different $g(r)$ set by

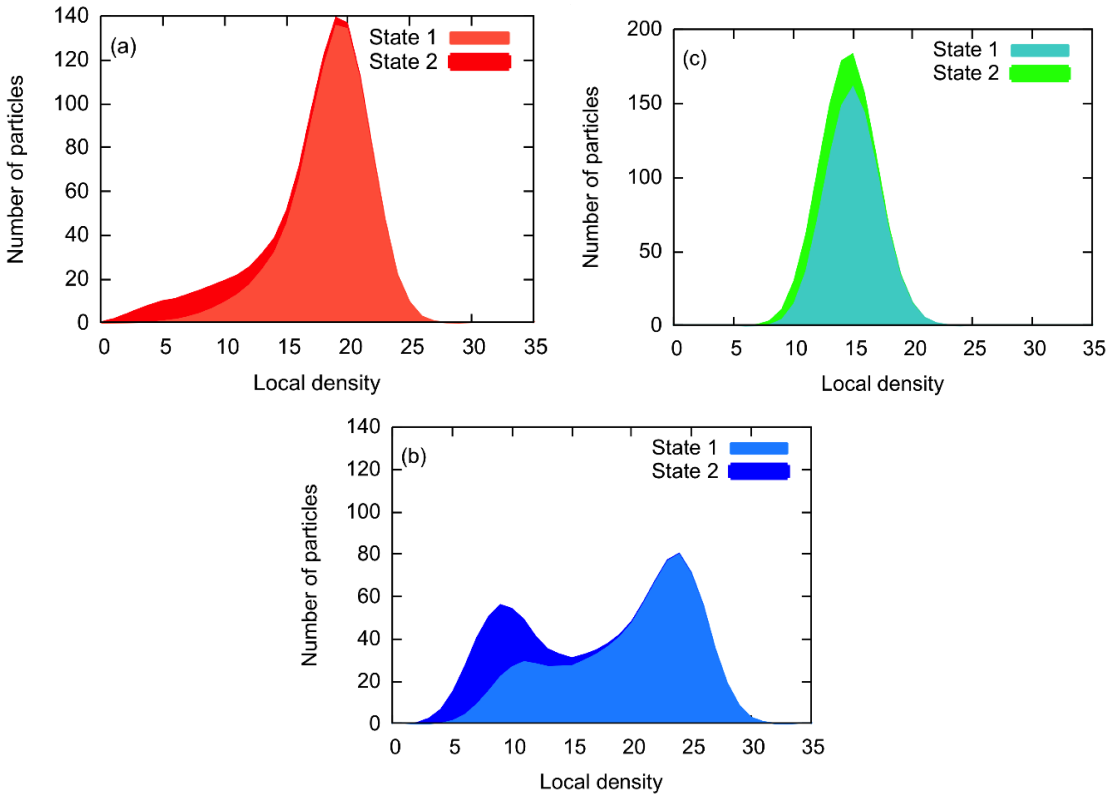

Figure 13.4: Histograms of particle coordination number in three systems, a pure LJ coexistence system (top left, $\epsilon = 0.18$ kcal/mol), a RE-UCG model in which state change upon high coordination makes neighbor interaction weaker (top right) and an RE-UCG model in which the high-coordination state has stronger neighbor interactions (bottom left), with bars colored by the likelihood a particle with that local coordination is in either the high-coordination state (state 1, light) or the low-coordination state (state 2, dark).

different pair potentials, rather than requiring a single CG pair potential to fit the $g(r)$s for both phases simultaneously.

13.3.3 Interactions changing through cooperative aggregation

As a final modelistic investigation, we simulated a minimal RE-UCG model of cooperative aggregation. This time, the systems are composed of particles with two types, one majority component (solvent) with only one substate and the other minority component (solute) with two substates. Once again all the particles interact through LJ interactions with different LJ ϵ parameters in each type and substate.

In this case, the solute substate probabilities are once again assigned based on a threshold of local coordination number around each particle as in the last example, but only solute coordination number; the states change depending on how many solutes are in one another's immediate vicinity. The parameter R_{th} is kept fixed at 1.0 Å, while W_{th} is varied. W_{th} accounts for the critical number of particles that must assemble around a particle before it begins to exhibit significant cooperativity.

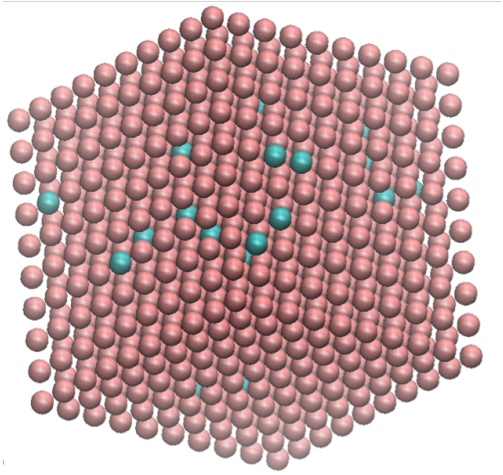

Figure 13.5: Representative random initial condition for the RE-UCG cooperative aggregation model simulations prior to equilibration. Teal particles represent solutes with two internal substates; pink particles represent solvent.

Each system contained 1000 CG particles, 27 solute particles and 973 solvent particles in a cubic box 10.7 by 10.7 by 10.7 Åwith periodic boundary conditions. CG particles were initialized from a equally distributed homogeneous face center cubic (FCC) lattice state, then the minority components (solute particles) were assigned randomly; a typical initial state is shown in 13.5. NVT dynamics with a Nosé-Hoover thermostat[249, 148] were performed at 90.6096 K, the same temperature as the previous phase coexistence study. A Verlet time integrator[334] was used with a time step of 1 fs as in previous cases. The simulation time was extended to 5 ns to examine the clustering behavior among solute particles.

Solvent CG particles were given LJ parameters of $\epsilon = 0.10$ (kcal/mol) and $\sigma = 1.0$ Å. To model the RE-UCG interactions between CG solutes, the two UCG states were first, a

high-coordination state with $\epsilon = 0.26$ (kcal/mol), $\sigma = 0.7$ Åand second, the low-coordination state with $\epsilon = 0.04$ (kcal/mol), $\sigma = 1.0$ Å; thus solutes get smaller σ and stickier ϵ as they assemble. These models were investigated with W_{th} (coordination number threshold) of 1.0, 1.5, and 2.0.

Furthermore, three additional two-component LJ fluid simulations without UCG substates for the solute were also conducted as references, describing the solute particle as only the stickier, more compact state ($\epsilon = 0.26$ kcal/mol, $\sigma = 0.7$ Å) or only the less sticky, bigger state ($\epsilon = 0.10$ kcal/mol, $\sigma = 1.0$ Å), or the arithmetic mean between them, for comparison.

The purpose of the states in this model was to induce cooperative aggregation behavior, to enhance nucleation. The success is evident in Figure 13.6, which shows the size of the largest cluster in each simulation as a function of time. The lower the number of neighboring particles required to induce cooperativity (W_{th}), the greater the chance of larger aggregates forming.

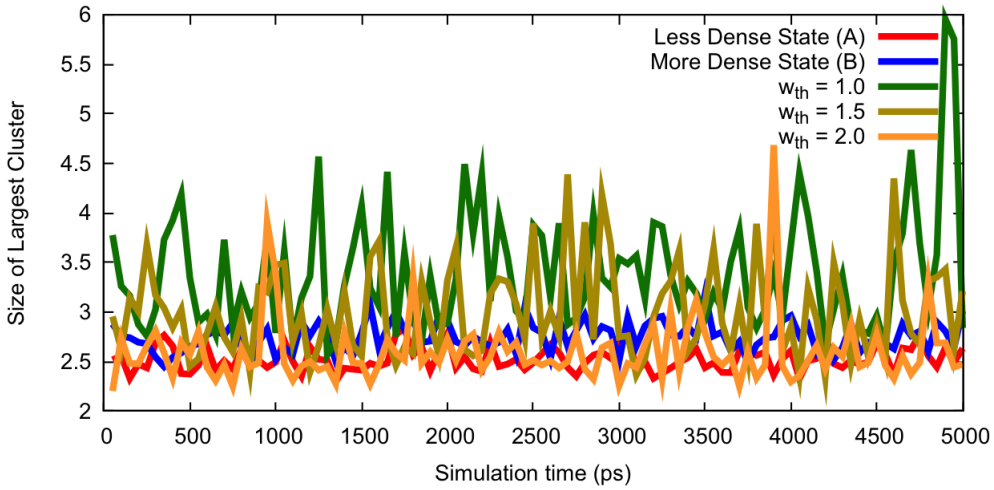

Figure 13.6: Trajectories of the largest cluster size in the system as a function of time for two LJ models with high and low LJ solute interaction strength (red and blue) and three RE-UCG models (green, gold, orange) that change from low interaction to high interaction strength after being coordinated by w_{th} other solute particles. Clusters are defined as networks of particles within 2.0 Åof one another. The effect of state change in the RE-UCG system increases clustering far more than direct increase of interaction strength.

The log histogram of aggregate sizes over the course of the runs is shown in Figure 13.6,

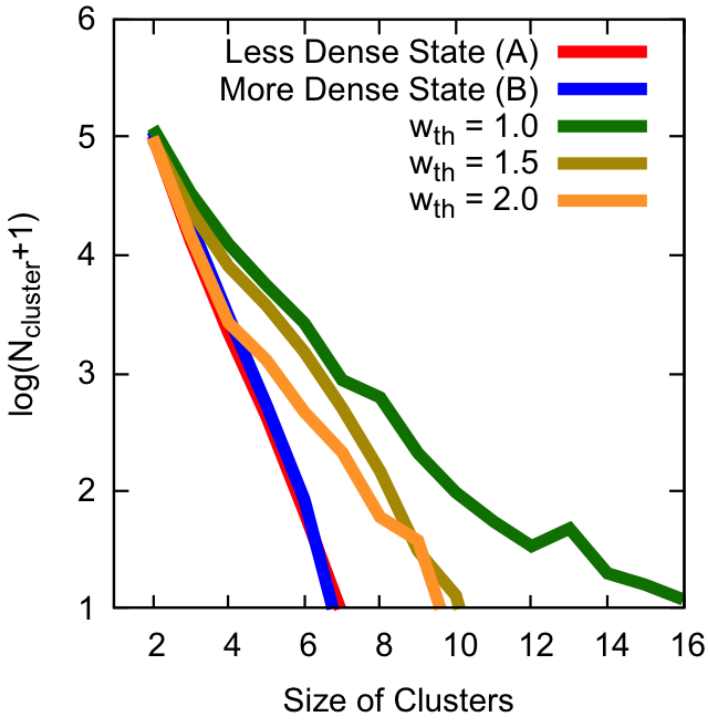

Figure 13.7: Log histograms of cluster sizes for two LJ models with high and low LJ solute interaction strength (red and blue) and three RE-UCG models (green, gold, orange) that change from low interaction to high interaction strength after being coordinated by w_{th} other solute particles. Clusters are defined as networks of particles within 2.0 nm of one another. The effect of state change in the RE-UCG system increases clustering far more than direct increase of interaction strength.

and confirms the story. Moreover, Figures 13.6 and 13.7 provide comparisons with the solutes using fixed LJ parameters (no substates) that are high and low, showing that the enhanced aggregation cannot be explained solely in terms of an increased ϵ parameter. It is critical to increase ϵ only when close to other solutes in order to enhance aggregation.

Using state switching has a qualitatively different effect on aggregation than simply increasing ϵ, and the same is also of the effect on pair correlations. Figure 13.8 shows that the solute-solute pair correlations for the mixed-state UCG systems differ from either limit of the systems with fixed-parameter LJ solutes. The first peaks of the UCG systems' pair correlations grow more and more like the higher-ϵ LJ solute as W_{th} is decreased, yet the second peak remains in essentially the same place. The mixture of the force fields based on

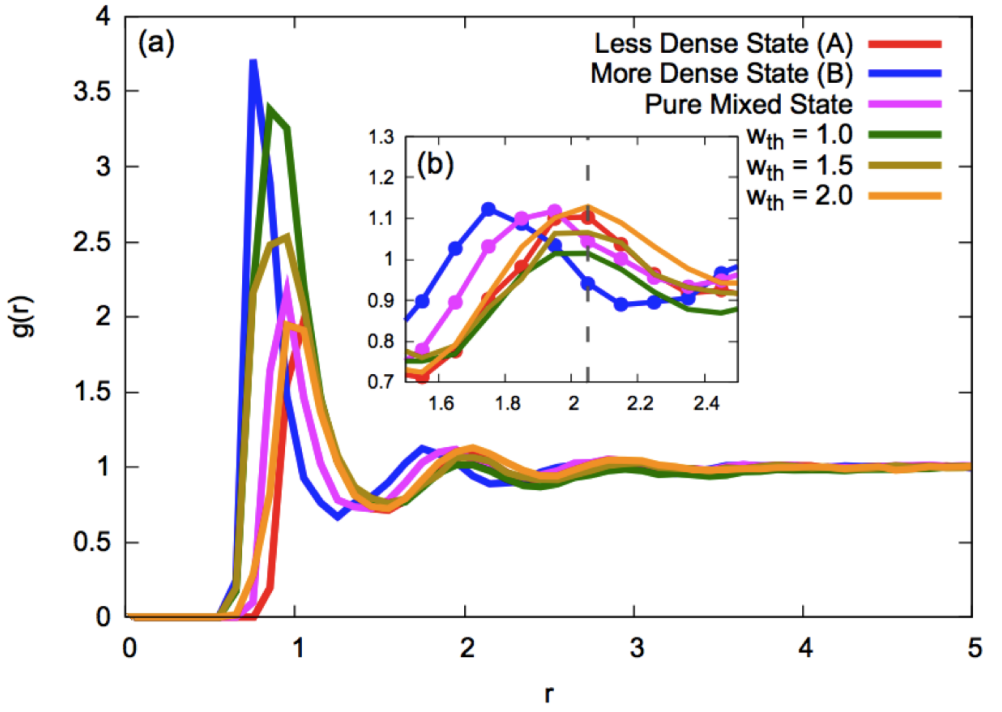

Figure 13.8: Solute-solute pair correlation functions $g(r)$ for two LJ models with high, low, and intermediate LJ solute interaction strength (red, blue, and pink) and three RE-UCG models (green, gold, orange) that change from low interaction to high interaction strength after being coordinated by w_{th} other solute particles. Inset shows a closeup of the second peak, which shifts steadily with LJ parameter if it is changed directly, but does not shift across RE-UCG model parameters.

coordination is not equivalent to a simple intermediate system.

13.4 Conclusion

The first paper in this series described a variational principle for the equilibrium statistics of a UCG model to an atomistic reference, but left the dynamics to future work. The second paper considered the case that states change so rarely and promptly that they can be treated as Markov states, giving rise to a state hopping dynamics. This third paper has now considered the case that the states equilibrate so rapidly that the state changes are entirely governed by their quasi-equilibrium distribution. Both limits have a natural domains of applicability, and though the intermediate cases also appear interesting, these two elegant

limits provide important experience and low-hanging fruit for the UCG approach.

Calculating a global quasi-equilibrium of states using substates defined like in the previous chapter would be infeasible. Like in Hartree-Fock or multiproton MS-EVB,[342] designing an efficient RE-UCG force field depends crucially on the separability and locality of the UCG substate probabilities. For UCG, a natural choice is to assign states through substates based on strictly CG-observable neighborhood measurements instead of implicit optimization. This gives a relatively simple mathematical form for the interactions that can be programmed in a way that makes excellent use of existing simulation engine implementations and can be fit using existing force-matching or relative entropy codes with minimal new coding–no new coding inside the fitting program, in fact.

The new force fields will have a number of potential applications. It is very common to have one force field to describe one set of conditions and a second for another, for instance in cases of phase coexistence, and this allows a smooth mixing based on the CG order parameters of the phases. Cooperative interactions short of phase segregation are also common, from mutual polarization of simple molecules to complex allosteric effects on protein interactions. Attempting to model this fully generally with arbitrary force field forms is infeasible, but UCG's tight focus means it does not have the incredible complexity and cost that entirely generic force fields do. UCG uses standard and well-understood force fields as components and simply adds a mixing layer around those.

Nonetheless, a great deal remains to be determined. Any dynamical regimes between the two complete separations of time scale described so far in this series could be considered valid modeling targets. For example, there may be cases where states equilibrate at a limited rate so that the probabilities change according to a simple differential equation. Also, though it has not been explored here outside of the theory, the UCG approach is not limited to adjusting interactions through particle types. One could assign states to the particular pair interactions instead, perhaps to mimic exclusive hydrogen bonding or protein allostery.

There remain major open opportunities for finding additional relatively simple limiting

cases that can be written on top of existing simulation and modeling infrastructure naturally, like both the second paper in this series and this paper have now done, and for discovering how best to apply them to real problems in molecular science.

CHAPTER 14

CONCLUSION

Current bottom-up coarse-graining methods cannot account for the full physics of mesoscale models, but the work presented in thesis enriches those methods. First, it brings us closer to understanding transferability and associated issues of model interpretation for comparison to both experiment and more detailed chemical models. Second, it provides new understanding of and greater power for the enhanced sampling tools required to characterize those internal degrees of freedom in complex molecules. Third, it provides new tools for bottom-up modeling of complex molecules with internal degrees of freedom that aren't well-accounted for by particle motion at the desired level of coarse-graining.

Chapters 3-5 reviewed the multiscale coarse-graining (MS-CG) method,[154, 246, 247] then introduced new tools for calculating the connection between fine-grained models and MS-CG coarse-grainings of them, and then discussed the interpretation of these models in terms of their possible comparison to experiment or fine-grained models. These both provide new means to use coarse-grained modeling to guide chemical design at the atomic level and to use coarse-grained models to make different classes of prediction with cleaner conceptual connections to experiment.

The theory in Chapters 4 and 5 has strong precedents when it comes to thermodynamics. In normal cases in soft matter molecular physics, coarse-grained force fields correspond to free energies, not internal energies, and similarly the pressure of a coarse-grained model should not be simply related to its virial unless the structural distributions do not correspond between the fine-grained and coarse-grained models. This is not often taken into account in the coarse-graining literature, but these chapters also describe a happy cancellation of misinterpretations that has helped ensure that things do not go too wrong when such questionable choices are made: if one looks for optimally transferable models, i.e., ones in which the potential is least sensitive to thermodynamic state variables, then the thermodynamic observables described in Chapter 5 will be as close to the normal atomistic model observable

forms as possible.

Chapters 6-10 reviewed metadynamics, provided proof that it converges exactly, and then described three new methodological improvements for metadynamics. Most directly, this means that metadynamics has become much more powerful in our hands and we are using it with higher success rates than ever before, applying it to membrane permeation, actin mechanics, quantum mechanics/molecular mechanics simulation of adenosine triphosphate hydrolysis in actin, protein-ligand binding, proton transfer, and more. The former problems of tuning parameters based on missing pre-knowledge of the transition barrier height, driving forces that cause irreversible damage to sensitive protein structures, and false convergence of the bias that cannot be reproduced in statistically independent samples are all significantly reduced and often eliminated completely.

The proof and all of the method development relied on a stochastic approximation analysis via the stability differential equation method,[179] which is remarkably general. Before this work was carried out, researchers believed that metadynamics required much stronger separation of time scales to be exact. However, as in the previous 3 chapters, we uncovered an unexpected cancellation of errors: even if the finite hill height does lead to inaccuracies early on, these answers cancel out almost regardless of the associated collective variable dynamics. If the sampling autocorrelation decays at least as fast as a stretched exponential, the free energy profiles can converge exactly. Of course, practical convergence remains an issue, but examining the steps in the proof showed that the most common convergence problems have effective remedies (described in Chapters 7-10).

Finally, Chapters 11-13 provided a new Ultra Coarse-Graining (UCG) theoretical framework that marries Markov state models with MS-CG and relative entropy-based bottom-up coarse-graining. Chapter 11 defined the framework and these marriages, while Chapters 12 and 13 described particular implementations of the framework with particular discrete state dynamics. In Chapter 12, the discrete states are Markov states that transition with environment-dependent rates systematically optimized to match underlying model rates. In

Chapter 13, the discrete state populations are always in quasi-equilibrium, and the force field appears as a mixture of force fields with smoothly changing weights.

In Chapter 12, including states in the simple molecular fluid revealed differences in radial distribution functions between molecules in each pair of the two molecular conformational states which were completely obscured in the one-state-per-molecule model that does not distinguish between conformations. Yet MS-CG actually force-matches the one-state-per-molecule model extremely well, indicating that somehow the force matching is robust to the multimodal pair structure distribution. This contradicts expectations that force matching should fail when the underlying distribution does not in fact match the assumed form of the force field, and indicates the possibility of deeper cancellation of error that even force matching is prone to, undermining its diagnostic power. However, this cancellation of error seems likely to reflect deep underlying physics and will be interesting to characterize and investigate for further insight.

Conditional cancellations like the ones identified in each major section above are one hallmark of a successful singular perturbation theory: "improving" any one of the errors is likely to make the model worse, not better, so adding corrections ad infinitum might as easily lead to ever poorer approximation rather than ever better. Indeed, in Chapter 4 we saw that the sensitivities of these models to underlying parameters can be strikingly poor predictors of the actual change in the models when the parameters are changed by a finite amount. Even more tellingly, Chapter 4 showed that the sensitivities of the few-body approximations to the ideal many-body coarse-grained force field often looked very different from few body approximations of the sensitivity of the ideal many-body coarse-grained force field: the basis set restriction is doing both positive and negative work when it comes to transferability. In Chapter 7, several singularities were immediately apparent in the proof and integral to its success: first, that the bias level must go to infinity precisely logarithmically, second, that the restrictions on the collective variable dynamics correspond to convergence criteria for particular sums. The quasi-equilibrium differential equation describes a particular limit of a

family of stochastic processes and the proof described the radius of convergence around it.

However, short of the grand and speculative goal of discovering new singular perturbation theories from these cancellations of errors, this research will also be continued on in a number of other ways more immediately and certainly.

First, starting from the back, Jaehyeok Jin and I are in the process of actually parameterizing models as defined in Chapter 13 from the bottom up using MS-CG code. The first application system is one of the hydrophobic aggregation of neopentane in methanol, using two states for each methanol to distinguish between an isotropic state far from neopentane and a relatively oriented state when immediately next to neopentane. Once this is successful, our next steps will likely be to tackle either lipid phase transitions or lipid tail splay near regions of high Gaussian curvature in coarse-grained implicit solvent phospholipid membranes.

Second, John Grime and I have collaborated on top-down models of the self-assembly of HIV capsid using two discrete conformational states for the coarse-grained capsid proteins: one assembly competent and one not assembly competent. This rather successful model was recently accepted for publication and, for his next target, he is planning on studying the stages of virion maturation before capsid self-assembly begins: the action of certain proteases on large multidomain membrane proteins. For this model we plan to finally use a UCG state that corresponds to breaking bonds, and break the pattern of seeing every UCG state as something that lives in a particle and is equivalent to changing a particle type. Once this is implemented in simulation code, it should be a natural next step to work on a bottom-up model of the same class, building off of Chapter 12.

Third, the problem of defining state membership functions for UCG is a tricky one that Alek Durumeric and I are attacking together at the moment. We have devised a two-step machine learning approach that first learns whether the local environments of particles seen through a coarse-grained lens on a fine-grained system show state-like autocorrelation. If they do, we then fit a hidden Markov model to explain the autocorrelation, then train a classifier

for those state assignments that takes only internal variables of the apparent coarse-grained entity as input, both to ensure that the state assignment corresponds to actual slow internal motions and to construct an explicit state assignment probability.

Next, a great deal of metadynamics work is now ongoing in the Voth group: eminently practical stuff like calculations of membrane permeation, ligand binding affinities, and reaction barriers in enzymes. With the three methodological improvements described in Chapters 8-10, we find metadynamics far more robust and reliable. Rui Sun in particular has become an expert in application who has been helping innumerable Voth group members learn to use the software and interpret the results.

Insight from the metadynamics proof in Chapter 7 has also inspired a flurry of other activity, including two works I coauthored, one Andrew White's experiment-directed metadynamics[353] that uses metadynamics to achieve any sampling distribution, not just flat sampling, and the second a collaboration with Pratyush Tiwary and Michele Parrinello that uses a perturbative solution to the quasi-equilibrium metadynamics differential equation to estimate the severity of errors associated with using metadynamics hills that are broader than the features of a free energy surface.[319] Several more are in preparation or currently submitted.

There are countless small new ideas floating around my head from the convergence work at this point, and it's almost impossible these days that any conversation about adaptive enhanced sampling does not inspire me to write a page of notes and sketches. I implemented driven metadynamics[232] with Mahmoud Moradi, made a transition-tempered variant of the orthogonal space random walk[369, 198] with Mike Schnieder, and wrote notes on a spatially regularized adaptive biasing force[93, 145] method with Jerome Henin. There are many incredible opportunities for further development in this space, big and small.

The work in Chapters 4 and 5, meanwhile, is being continued by Jacob Wagner; some unpublished results on these problems have us convinced that there are deep, strange phenomena hiding in the thermodynamics of even the simplest force-matched methanol model. One project that deserves particular mention, though it is his and I lay no claim to it, uses

the primary observable definition of Chapter 5 as the basis for systematically matching general few-body approximations to fine-grained per-particle observables, like force matching matches forces.

All of these are very direct continuations, which is in one sense a pleasant confirmation that the work has been worthwhile. However, it is also rather predictable. I have left one further more ambitious and unusual continuation for last.

With Yining Han, I am currently implementing what is to my knowledge the first rigorous supramolecular coarse-graining of Lagrangian fluid elements that does not rely either on assignment based on tracer particles or periodic re-initialization.[152] The scheme uses UCG ideas to account for mass transfer from coarse-grained entity to coarse-grained entity, constructing a smooth dynamic mapping propagated via a well-defined matrix differential equation given an input fine-grained trajectory. Our goal is to provide the first rigorous, bottom-up derivation of smoothed particle dissipative particle dynamics,[114] and we have high hopes given our initial findings.

If this supramolecular coarse-graining works, it will enable connections I have long dreamed of being able make from molecular physics to mesoscopic phenomenological models: most importantly, it will enable the bottom-up definition of de Gennes blobs[103, 283] that include both fixed polymer mass and solvent mass that comes and goes dynamically.

De Gennes blob models are known to be rather natural to map up to tractable mesoscopic field theories.[129, 283, 130] Given simple geometries like a lamellar microphase separation, those in turn map up to continuum field theories via homogenization.[258] Therefore this could provide the last link in a complete chain from atomistic scales to continuum scales for a wide class of polymer models. This would complete one of my long-term projects coming into this graduate program unexpectedly early. Time will tell.

REFERENCES

[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden. The heterogeneous multiscale method. *Acta Numerica*, 21:1–87, 2012.

[2] K. M. Aberg, A. P. Lyubartsev, S. P. Jacobsson, and A. Laaksonen. Determination of solvation free energies by adaptive expanded ensemble molecular dynamics. *Journal of Chemical Physics*, 120:3770–3776, 2004.

[3] C. Abrams and G. Bussi. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy*, 16:163–199, 2013.

[4] C. F. Abrams and E. Vanden-Eijnden. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 107:4961–4966, 2010.

[5] J. B. Abrams and M. E. Tuckerman. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *Journal of Physical Chemistry B*, 112:15742–15757, 2008.

[6] A. Accardi and C. Miller. Secondary active transport mediated by a prokaryotic homologue of ClC Cl- channels. *Nature*, 427:803–807, 2004.

[7] S. A. Adcock and J. A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106:1589–1615, 2006.

[8] R. L. C. Akkermans and W. J. Briels. A structure-based coarse-grained model for polymer melts. *Journal of Chemical Physics*, 114:1020–1031, 2001.

[9] E. C. Allen and G. C. Rutledge. A novel algorithm for creating coarse-grained, density dependent implicit solvent models. *Journal of Chemical Physics*, 128:154115, 2008.

[10] E. C. Allen and G. C. Rutledge. Coarse-grained, density dependent implicit solvent model reliably reproduces behavior of a model surfactant system. *Journal of Chemical Physics*, 130:204903, 2009.

[11] E. C. Allen and G. C. Rutledge. Evaluating the transferability of coarse-grained, density-dependent implicit solvent models in mixtures and chains. *Journal of Chemical Physics*, 130:034904, 2009.

[12] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford UP, Oxford, 1989.

[13] H. C. Andersen and D. Chandler. Optimized cluster expansions for classical fluids. I. General theory and variational formulation of the mean spherical model and hard sphere Percus-Yevick equations. *Journal of Chemical Physics*, 57:1918–1929, 1972.

[14] C. A. Angell, R. D. Bressel, M. Hemmati, E. J. Sare, and J. C. Tucker. Water and its anomalies in perspective: Tetrahedral liquids with and without liquid-liquid phase transitions. *Physical Chemistry Chemical Physics*, 2:1559–11566, 2000.

[15] A. Arkhipov, P. L. Freddolino, and K. Schulten. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure*, 14:1767–1777, 2006.

[16] P. Ascarelli and R. J. Harrison. Density-dependent potentials and the hard sphere model for liquid metals. *Physical Review Letters*, 22:385–388, 1969.

[17] H. S. Ashbaugh, H. A. Patel, S. K. Kumar, and S. Garde. Mesoscale model of polymer melt structure: self-consistent mapping of molecular correlations to coarse-grained potentials. *Journal of Chemical Physics*, 122:104908, 2005.

[18] Y. F. Atchadé and J. S. Liu. The Wang-Landau algorithm in general state spaces: Applications and convergence analysis. *Statistica Sinica*, 20:209–233, 2010.

[19] G. S. Ayton, E. Lyman, and G. A. Voth. Hierarchical coarse-graining strategy for protein-membrane systems to access mesoscopic scales. *Faraday Discussions*, 144:347–357, 2010.

[20] G. S. Ayton, W. G. Noid, and G. A. Voth. Multiscale modeling of biomolecular systems: in serial and in parallel. *Current Opinion in Structural Biology*, 17:192–198, 2007.

[21] G. S. Ayton and G. A. Voth. Hybrid coarse-graining approach for lipid bilayers at large length and time scales. *Journal of Physical Chemistry B*, 113:4413–4424, 2009.

[22] F. Baftizadeh, X. Biarnes, F. Pietrucci, F. Affinito, and A. Laio. Multidimensional view of amyloid fibril nucleation in atomistic detail. *Journal of the American Chemical Society*, 134:3886–3894, 2012.

[23] F. Baftizadeh, P. Cossio, F. Pietrucci, and A. Laio. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Current Physical Chemistry*, 2:79–91, 2012.

[24] I. Bahar, T. R. Lezon, L.-W. Yang, and E. Eyal. Global dynamics of proteins: Bridging between structure and function. *Annual Review of Biophysics*, 39:23–42, 2010.

[25] J. N. Bandaria, S. Dutta, M. W. Nydegger, W. Rock, A. Kohen, and C. M. Cheatum. Characterizing the dynamics of functionally relevant complexes of formate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 107:17974–17979, 2010.

[26] A. Banerji and I. Ghosh. Fractal symmetry of protein interior: what have we learned? *Cellular and Molecular Life Sciences*, 68:2711–2737, 2011.

[27] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 1:826–843, 2011.

[28] A. Barducci, G. Bussi, and M. Parrinello. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, 100:020603, 2008.

[29] C. Bartels and M. Karplus. Multidimensional adaptive umbrella sampling: applications to main chain and side chain peptide conformations. *Journal of Computational Chemistry*, 18:1450–1462, 1997.

[30] B. A. J. Baumann, D. W. Taylor, Z. Huang, F. Tama, P. M. Fagnant, K. M. Trybus, and K. a. Taylor. Phosphorylated smooth muscle heavy meromyosin shows an open conformation linked to activation. *Journal of Molecular Biology*, 415:274–287, 2012.

[31] D. Beglov and B. Roux. An integral equation to describe the solvation of polar molecules in liquid water. *Journal of Physical Chemistry B*, 101:7821–7826, 1997.

[32] R. Belardinelli, S. Manzi, and V. Pereyra. Analysis of the convergence of the 1t and Wang-Landau algorithms in the calculation of multidimensional integrals. *Physical Review E*, 78:067701, 2008.

[33] R. Belardinelli and V. Pereyra. Fast algorithm to calculate density of states. *Physical Review E*, 75:046701, 2007.

[34] R. E. Belardinelli and V. D. Pereyra. Wang-Landau algorithm: a theoretical analysis of the saturation of the error. *Journal of Chemical Physics*, 127:184105, 2007.

[35] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory*. Springer, 1999.

[36] N. C. Benson and V. Daggett. A Comparison of Multiscale Methods for the Analysis of Molecular Dynamics Simulations. *Journal of Physical Chemistry B*, 116:8722–8731, 2012.

[37] A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures*. American Mathematical Society Chelsea Publishing, Providence, RI, 1978.

[38] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman, editor, *Intermolecular Forces*, pages 331–342. Reidel, Dordrecht, The Netherlands, 1981.

[39] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91:6269–6271, 1987.

[40] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, a. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81:3684–3690, 1984.

[41] H. J. C. Berendsen, D. Vanderspoel, and R. Vandrunen. Gromacs - a message-passing parallel molecular-dynamics implementation. *Computer Physics Communications*, 91:43–56, 1995.

[42] B. Berg and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68:9–12, 1992.

[43] I. Bilionis and P. Koutsourelakis. Free energy computations by minimization of KullbackLeibler divergence: An efficient adaptive biasing potential method for sparse representations. *Journal of Computational Physics*, 231:3849–3870, 2012.

[44] I. Bilionis and N. Zabaras. A stochastic optimization approach to coarse-graining using a relative entropy framework. *Journal of Chemical Physics*, 138:044313, 2013.

[45] D. C. Boisvert, J. Wang, Z. Otwinowski, A. L. Horwich, and P. B. Sigler. The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nature Structural and Molecular Biology*, 3:170–177, 1996.

[46] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *Journal of Computational Chemistry*, 30:1615–1621, 2009.

[47] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180:1961–1972, 2009.

[48] K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, and P. B. Sigler. The crystal structure of the bacterial chaperonin GroEL at 2.8 Angstrom. *Nature*, 371:578–586, 1994.

[49] D. Branduardi, G. Bussi, and M. Parrinello. Metadynamics with adaptive Gaussians. *Journal of Chemical Theory and Computation*, 8:2247–2254, 2012.

[50] D. Branduardi, F. L. Gervasio, and M. Parrinello. From A to B in free energy space. *Journal of Chemical Physics*, 126:054103, 2007.

[51] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodriguez-Ropero, and N. F. A. van der Vegt. Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter*, 9:2108–2119, 2013.

[52] W. M. Brown, A. Kohlmeyer, S. J. Plimpton, and A. N. Tharrington. Implementing molecular dynamics on hybrid high performance computers - particle-particle particle-mesh. *Computer Physics Communications*, 183:449–459, 2012.

[53] W. M. Brown, P. Wang, S. J. Plimpton, and A. N. Tharrington. Implementing molecular dynamics on hybrid high performance computers - short range forces. *Computer Physics Communications*, 182:898–911, 2011.

[54] M. Brunori, D. Bourgeois, and B. Vallone. The structural dynamics of myoglobin. *Journal of Molecular Biology*, 147:223–234, 2004.

[55] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, 126:014101, 2007.

[56] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society*, 128:13435–13441, 2006.

[57] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Physical Review Letters*, 96:090601, 2006.

[58] C. Caccamo and G. Pellicane. Microscopic theories of model macromolecular fluids and fullerenes: The role of thermodynamic consistency. *Journal of Chemical Physics*, 117:5072–5085, 2002.

[59] A. C. Cameron, J. B. Gelbach, and D. L. Miller. Boostrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90:414–427, 2008.

[60] Z. Cao and G. A. Voth. The multiscale coarse-graining method. XI. Accurate interactions based on the centers of charge of coarse-grained sites . *Journal of Chemical Physics*, 143:243116, 2015.

[61] Z. Cao, J. F. Dama, L. Lu, and G. A. Voth. Solvent free ionic solution models from Multiscale Coarse-Graining. *Journal of Chemical Theory and Computation*, 9:172–178, 2013.

[62] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55:2471–2474, 1985.

[63] S. P. Carmichael and M. S. Shell. A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly. *Journal of Physical Chemistry B*, 116:8383–8393, 2012.

[64] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chemical Physics Letters*, 156:472–477, 1989.

[65] M. Ceriotti, G. Bussi, and M. Parrinello. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Physical Review Letters*, 102:020601, 2009.

[66] M. Ceriotti, G. Bussi, and M. Parrinello. Colored-noise thermostats à la carte. *Journal of Chemical Theory and Computation*, 6:1170–1180, 2010.

[67] M. Ceriotti, G. A. Tribello, and M. Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences of the United States of America*, 108:13023–13028, 2011.

[68] A. Chaimovich and M. S. Shell. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Physical Chemistry Chemical Physics*, 11:1901–1915, 2009.

[69] A. Chaimovich and M. S. Shell. Coarse-graining errors and numerical optimization using a relative entropy framework. *Journal of Chemical Physics*, 134:094112, 2011.

[70] D. Y. C. Chan, P. Linse, and S. N. Petris. Phase separation in deionized colloidal systems: Extended Debye-Hückel theory. *Langmuir*, 17:4202–4210, 2001.

[71] D. Chandler and H. Andersen. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. *Journal of Chemical Physics*, 57:1930–1937, 1972.

[72] L. Chen, N. Goldenfeld, and Y. Oono. Renormalization group and singular perturbations: Multiple scales, boundary layers, and reductive perturbation theory. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 54:376–394, 1996.

[73] C. Chennubhotla and I. Bahar. Markov methods for hierarchical coarse-graining of large protein dynamics. *Journal of Molecular Biology*, 14:765–776, 2007.

[74] H. M. Cho and J.-W. Chu. Inversion of radial distribution functions to pair forces by solving the Yvon-Born-Green equation iteratively. *Journal of Chemical Physics*, 131:134107, 2009.

[75] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics*, 126:155101, 2007.

[76] N. Chopin, T. Lelièvre, and G. Stoltz. Free energy methods for Bayesian inference: efficient exploration of univariate Gaussian mixture posteriors. *Statistics and Computing*, 22:897–916, 2012.

[77] C. D. Christ and W. F. van Gunsteren. Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *Journal of Chemical Physics*, 126:184110, 2007.

[78] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, 31:1569–1582, 2009.

[79] J. W. Chu and G. A. Voth. Allostery of actin filaments: molecular dynamics simulations and coarse-grained analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 102:13111–13116, 2005.

[80] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *Chemphyschem*, 6:1809–1814, 2005.

[81] P. Cieplak, D. A. Pearlman, and P. A. Kollman. Walking on the free energy hypersurface of the 18-crown-6 ion system using free energy derivatives. *Journal of Chemical Physics*, 101:627–633, 1994.

[82] T. E. Colla, A. P. dos Santos, and Y. Levin. Equation of state of charged colloidal suspensions and its dependence on the thermodynamic route. *Journal of Chemical Physics*, 136:194103, 2012.

[83] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Press, Cambridge, MA, 1st edition, 1990.

[84] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117:5179–5197, 1995.

[85] Y. Crespo, F. Marinelli, F. Pietrucci, and A. Laio. Metadynamics convergence law in a multidimensional system. *Physical Review E*, 81:055701, 2010.

[86] G. D'Adamo, A. Pelissetto, and C. Pierleoni. Predicting the thermodynamics by using state-dependent interactions. *Journal of Chemical Physics*, 138:234107, 2013.

[87] J. F. Dama, G. M. Hocky, R. Sun, and G. A. Voth. Exploring valleys without climbing every peak: More efficient and forgiving metabasin metadynamics via robust on-the-fly bias domain restriction. *Journal of Chemical Theory and Computation*, 11:5638–5650, 2015.

[88] J. F. Dama, M. Parrinello, and G. A. Voth. Well-tempered metadynamics converges asymptotically. *Physical Review Letters*, 112:240602, 2014.

[89] J. F. Dama, G. Rotskoff, M. Parrinello, and G. A. Voth. Transition-tempered metadynamics: Robust, convergent metadynamics via on-the-fly transition barrier estimation. *Journal of Chemical Theory and Computation*, 10:3626–3633, 2014.

[90] J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth. The theory of Ultra-Coarse-Graining. 1. General principles. *Journal of Chemical Theory and Computation*, 9:2466–2480, 2013.

[91] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, 98:010089, 1993.

[92] E. Darve and A. Pohorille. Calculating free energies using average force. *Journal of Chemical Physics*, 115:9169–9183, 2001.

[93] E. Darve, D. Rodríguez-Gómez, and A. Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *Journal of Chemical Physics*, 128:144120, 2008.

[94] A. Das and H. C. Andersen. The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. *Journal of Chemical Physics*, 131:034102, 2009.

[95] A. Das and H. C. Andersen. The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. *Journal of Chemical Physics*, 132:164106, 2010.

[96] A. Das and H. C. Andersen. The multiscale coarse-graining method. IX. A general method for construction of three body coarse-grained force fields. *Journal of Chemical Physics*, 136:194114, 2012.

[97] A. Das and H. C. Andersen. The multiscale coarse-graining method. VIII. Multiresolution hierarchical basis functions and basis function selection in the construction of coarse-grained force fields. *Journal of Chemical Physics*, 136:194113, 2012.

[98] A. Das, L. Lu, H. C. Andersen, and G. A. Voth. The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems. *Journal of Chemical Physics*, 136:194115, 2012.

[99] D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink. Improved parameters for the MARTINI coarse-grained protein force field. *Journal of Chemical Theory and Computation*, 9:687–697, 2013.

[100] B. M. Dickson. Approaching a parameter-free metadynamics. *Physical Review E*, 84:037701, 2011.

[101] B. M. Dickson, F. Legoll, T. Lelièvre, G. Stoltz, and P. Fleurat-Lessard. Free energy calculations: an efficient adaptive biasing potential method. *Journal of Physical Chemistry B*, 114:5823–5830, 2010.

[102] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[103] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Oxford UP, Oxford, 1986.

[104] P. Doruker, R. L. Jernigan, and I. Bahar. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *Journal of Computational Chemistry*, 23:119–127, 2002.

[105] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24:1999–2012, 2003.

[106] N. H. J. Dunn and W. G. Noid. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *Journal of Chemical Physics*, 143:243148, 2015.

[107] Z. A. Durer, D. S. Kudryashov, M. R. Sawaya, C. Altenbach, W. Hubbell, and E. Reisler. Structural states and dynamics of the d-loop in actin. *Biophysical Journal*, 103:930–939, 2012.

[108] R. Dutzler, E. B. Campbell, M. Cadene, B. T. Chait, and R. MacKinnon. X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. *Nature*, 415:287–294, 2002.

[109] R. Dutzler, E. B. Campbell, and R. MacKinnon. Gating the selectivity filter in ClC chloride channels. *Science*, 300:108–112, 2003.

[110] W. E, B. Engquist, X. Li, and W. Ren. The heterogeneous multiscale method: A review. *Communications in Computational Physics*, 2:367–450, 2007.

[111] B. Efron. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68:589–599, 1981.

[112] F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *Europhysics Letters*, 26:583–588, 1994.

[113] F. A. Escobedo and F. J. Martinez-Veracoechea. Optimization of expanded ensemble methods. *Journal of Chemical Physics*, 129:154107, 2008.

[114] P. Español and M. Revenga. Smoothed dissipative particle dynamics. *Physical Review E*, 67:026705, 2003.

[115] P. Español and I. Zúñiga. Obtaining fully dynamic coarse-grained models from MD. *Physical Chemistry Chemical Physics*, 13:10538–10545, 2011.

[116] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *Journal of Chemical Physics*, 103:8577–8593, 1995.

[117] J. Fan, M. G. Saunders, and G. A. Voth. Coarse-graining provides insights on the essential nature of heterogeneity in actin filaments. *Biophysical Journal*, 103:1334–1342, 2012.

[118] J. D. Faraldo-Gómez and B. Roux. Electrostatics of ion stabilization in a ClC chloride channel homologue from Escherichia coli. *Journal of Molecular Biology*, 339:981–1000, 2004.

[119] S. E. Feller and A. D. MacKerell Jr. An improved empirical potential energy function for molecular simulations of phospholipids. *Journal of Physical Chemistry B*, 104:7510–7515, 2000.

[120] L. Feng, E. B. Campbell, Y. Hsiung, and R. MacKinnon. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science*, 330:635–641, 2010.

[121] M. J. Ferrarotti, S. Bottaro, A. Pérez-Villa, and G. Bussi. Accurate multiple time step in biased molecular simulations. *Journal of Chemical Theory and Computation*, 11:139–146, 2015.

[122] G. Fiorin, M. L. Klein, and J. Hénin. Using collective variables to drive molecular dynamics simulations. *Molecular Physics*, 111:3345–3362, 2013.

[123] M. Fixman. Classical statistical mechanics of constraints: A theorem and application to polymers. *Proceedings of the National Academy of Sciences of the United States of America*, 71:3050–3053, 1974.

[124] S. H. Fleishman and C. L. Brooks. Thermodynamics of aqueous solvation: Solution properties of alcohols and alkanes. *Journal of Chemical Physics*, 87:3029–3037, 1987.

[125] E. W. Forgy. Cluster analysis of multivariate data - efficiency vs interpretability of classifications. *Biometrics*, 21:768–769, 1965.

[126] G. Fort, B. Jourdain, E. Kuhn, T. Lelièvre, and G. Stoltz. Convergence of the Wang-Landau algorithm. 2012.

[127] H. Frauenfelder, B. H. McMahon, R. H. Austin, K. Chu, and J. T. Groves. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 98:2370–2374, 2001.

[128] H. Frauenfelder, B. H. McMahon, and P. W. Fenimore. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 100:8615–8617, 2003.

[129] G. Fredrickson, V. Ganesan, and F. Drolet. Field-theoretic computer simulation methods for polymers and complex fluids. *Macromolecules*, 35:16–39, 2002.

[130] G. Fredrickson. *The Equilibrium Theory of Inhomogeneous Polymers*. Oxford University Press, 2005.

[131] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, Orlando, 2nd edition, 2001.

[132] S. Genheden. Coarse-grained bond and angle distributions from atomistic simulations: On the systematic parameterization of lipid models. *Journal of Molecular Graphics and Modeling*, 63:57–64, 2016.

[133] F. L. Gervasio, A. Laio, and M. Parrinello. Flexible docking in solution using metadynamics. *Journal of the American Chemical Society*, 127:2600–2607, 2005.

[134] Z. Ghaemi, M. Minozzi, P. Carloni, and A. Laio. A novel approach to the investigation of passive molecular permeation through lipid bilayers from atomistic simulations. *Journal of Physical Chemistry B*, 116:8714–8721, 2012.

[135] A. Gil-Ley and G. Bussi. Enhanced conformational sampling using replica exchange with collective-variable tempering. *Journal of Chemical Theory and Computation*, 11:1077–1085, 2015.

[136] S. C. Glotzer and W. Paul. Molecular and Mesoscale Simulation Methods for Polymer Materials. *Annual Review of Materials Research*, 32:401–436, 2002.

[137] F. Glover. Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13:533–549, 1986.

[138] N. Go. Theoretical studies of protein folding. *Annual Review of Biophysics and Bioengineering*, 12:183–210, 1983.

[139] N. Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. Addison-Wesley, 1992.

[140] C. Hansen and D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computation*, 14:1487–1503, 1993.

[141] R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Physics Letters A*, 49:197–198, 1974.

[142] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.

[143] B. Hess. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4:116–122, 2008.

[144] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4:435–447, 2008.

[145] J. Henin, G. Fiorin, C. Chipot, and M. L. Klein. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *Journal of Chemical Theory and Computation*, 6:35–47, 2010.

[146] C. Hijon, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscaloni. Mori-Zwanzig formalism as a practical computational tool. *Faraday Discussions*, 144:301–322, 2010.

[147] R. D. Hills, L. Lu, and G. A. Voth. Multiscale coarse-graining of the protein energy landscape. *PLoS Computational Biology*, 6:e1000827, 2010.

[148] W. G. Hoover. Canonical dynamics - equilibrium phase-space distributions. *Physical Review A*, 31:1695–1697, 1985.

[149] T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation : A method for improving the searching properties of molecular dynamics simulation. *Journal of Computer-Aided Molecular Design*, 8:695–708, 1994.

[150] I. M. Ilie, W. K. den Otter, and W. J. Briels. A coarse grained protein model with internal degrees of freedom. Application to α-synuclein aggregation. *Journal of Chemical Physics*, 144:085103, 2016.

[151] S. Izvekov, P. W. Chung, and B. M. Rice. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grained potentials. *Journal of Chemical Physics*, 133:064109, 2010.

[152] S. Izvekov and B. M. Rice. Multi-scale coarse-graining of non-conservative interactions in molecular liquids. *Journal of Chemical Physics*, 140:104104, 2014.

[153] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *Journal of Physical Chemistry B*, 109:2469–2473, 2005.

[154] S. Izvekov and G. A. Voth. Multiscale coarse graining of liquid-state systems. *Journal of Chemical Physics*, 123:134105, 2005.

[155] S. Izvekov and G. A. Voth. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *Journal of Chemical Physics*, 125:151101, 2006.

[156] S. Izvekov and G. A. Voth. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *Journal of Chemical Theory and Computation*, 2:637–648, 2006.

[157] S. Izvekov and G. A. Voth. Mixed resolution modeling of interactions in condensed-phase systems. *Journal of Chemical Theory and Computation*, 5:3232–3244, 2009.

[158] S. Izvekov and G. A. Voth. Solvent-free lipid bilayer model using multiscale coarse-graining. *Journal of Physical Chemistry B*, 113:4443–4455, 2009.

[159] J. P. M. Jämbeck and A. P. Lyubartsev. Exploring the free energy landscape of solutes embedded in lipid bilayers. *The Journal of Physical Chemistry Letters*, 4:1781–1787, 2013.

[160] G. Jayachandran, V. Vishal, and V. S. Pande. Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *Journal of Chemical Physics*, 124:164902, 2006.

[161] T. J. Jentsch. CLC chloride channels and transporters: From genes to protein structure, pathology and physiology. *Critical Reviews in Biochemistry and Molecular Biology*, 43:3–36, 2008.

[162] S. K. Jha, M. Ji, K. J. Gaffney, and S. G. Boxer. Direct measurement of the protein response to an electrostatic perturbation that mimics the catalytic cycle in ketosteroid isomerase. *Proceedings of the National Academy of Sciences of the United States of America*, 108:16612–16617, 2011.

[163] W. Jiang, Y. Wang, T. Yan, and G. A. Voth. A multiscale coarse-graining study of the liquid/vacuum interface of room-temperature ionic liquids with alkyl substituents of different lengths. *Journal of Physical Chemistry C*, 112:1132–1139, 2008.

[164] S. Jo, T. Kim, V. G. Iyer, and W. Im. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29:1859–1865, 2008.

[165] M. E. Johnson, T. Head-Gordon, and A. a. Louis. Representability problems for coarse-grained water potentials. *Journal of Chemical Physics*, 126:144509, 2007.

[166] H. L. Jones. Jacknife estimation of functions of stratum means. *Biometrika*, 61:343–348, 1974.

[167] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79:926–935, 1983.

[168] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118:11225–11236, 1996.

[169] I. S. Joung and T. E. I. I. I. Cheatham. Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *Journal of Physical Chemistry B*, 113:13279–13290, 2009.

[170] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*, 105:6474–6487, 2001.

[171] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9:646–652, 2002.

[172] J. Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1:932–942, 2011.

[173] E. H. Kellogg, O. F. Lange, and D. Baker. Evaluation and optimization of discrete state models of protein folding. *Journal of Physical Chemistry B*, 116:11405–11413, 2012.

[174] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell, and R. W. Pastor. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *Journal of Physical Chemistry B*, 114:7830–7843, 2010.

[175] P. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93:2395–2417, 1993.

[176] J.-P. Korb, Y. Goddard, J. Pajski, G. Diakova, and R. G. Bryant. Extreme-values statistics and dynamics of water at protein interfaces. *Journal of Physical Chemistry B*, 115:12845–12858, 2011.

[177] V. Krishna, W. G. Noid, and G. a. Voth. The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *Journal of Chemical Physics*, 131:024103, 2009.

[178] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

[179] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, New York, 2 edition, 1997.

[180] D. Laërtius. *Lives of the Eminent Philosophers. Book II. 108.*

[181] A. Laio and M. Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, 99:12562–12566, 2002.

[182] A. Laio and F. L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71:126601, 2008.

[183] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the accuracy of metadynamics. *Journal of Physical Chemistry B*, 109:6714–6721, 2005.

[184] T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories. *Journal of the American Chemical Society*, 133:18413–18419, 2011.

[185] L. Larini, L. Lu, and G. A. Voth. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *Journal of Chemical Physics*, 132:164107, 2010.

[186] L. Larini and J.-E. Shea. Coarse-Grained Modeling of Simple Molecules at Different Resolutions in the Absence of Good Sampling. *Journal of Physical Chemistry B*, 116:8337–8349, 2012.

[187] A. Y. Lau and B. Roux. Structural biology: One domain, multiple conformations. *Nature Chemical Biology*, 7:130–131, 2011.

[188] C. Le Bris, T. Lelièvre, M. Luskin, and D. Perez. A mathematical formalization of the parallel replica dynamics. *Monte Carlo Methods and Applications*, 18:119–146, 2012.

[189] T. Lelièvre and F. Nier. Low temperature asymptotics for quasistationary distributions in a bounded domain. *Analysis & PDE*, 8:561–628, 2015.

[190] T. Lelièvre, M. Rousset, and G. Stoltz. Long-time convergence of an adaptive biasing force method. *Nonlinearity*, 21:1155–1181, 2008.

[191] V. Leone, F. Marinelli, P. Carloni, and M. Parrinello. Targeting biomolecular flexibility with metadynamics. *Current Opinion in Structural Biology*, 20:148–154, 2010.

[192] F. Liang, C. Liu, and R. J. Carroll. Stochastic approximation in Monte Carlo computation. *Journal of the American Statistical Association*, 102:305–320, 2007.

[193] V. Limongelli, M. Bonomi, L. Marinelli, F. L. Gervasio, A. Cavalli, E. Novellino, and M. Parrinello. Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 107:5411–5416, 2010.

[194] E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7:306–317, 2001.

[195] P. Liu, S. Izvekov, and G. A. Voth. Multiscale coarse-graining of monosaccharides. *Journal of Physical Chemistry B*, 111:11566–11575, 2007.

[196] P. Liu, Q. Shi, H. Daumé, and G. A. Voth. A Bayesian statistics approach to multiscale coarse graining. *Journal of Chemical Physics*, 129:214114, 2008.

[197] A. A. Louis. Beware of density dependent pair potentials. *Journal of Physics: Condensed Matter*, 9187:9187–9206, 2002.

[198] C. Lu, X. Li, D. Wu, L. Zheng, and W. Yang. Predictive sampling of rare conformational events in aqueous solution: Designing a generalized orthogonal space tempering method. *Journal of Chemical Theory and Computation*, 12:41–52, 2016.

[199] J. Lu, Y. Qiu, R. Baron, and V. Molinero. Coarse graining of TIP4P/2005, TIP4P-Ew, SPC/E and TIP3P to monatomic anisotropic water models using relative entropy minimization. *Journal of Chemical Theory and Computation*, 10:4104–4120, 2014.

[200] L. Lu, J. F. Dama, and G. A. Voth. Fitting coarse-grained distribution functions through an iterative force-matching method. *Journal of Chemical Physics*, 139:121906, 2013.

[201] L. Lu, S. Izvekov, A. Das, H. C. Anderson, and G. A. Voth. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *Journal of Chemical Theory and Computation*, 6:954–965, 2010.

[202] L. Lu and G. A. Voth. The multiscale coarse-graining method. In S. A. Rice and A. R. Dinner, editors, *Advances in Chemical Physics*, volume 149, chapter 2. Wiley-Interscience, New York, 2012.

[203] L. Y. Lu and G. A. Voth. Systematic coarse-graining of a multicomponent lipid bilayer. *Journal of Physical Chemistry B*, 113:1501–1510, 2009.

[204] L. Lu and G. A. Voth. The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. *Journal of Chemical Physics*, 134:224107, 2011.

[205] K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at small and large length scales. *Journal of Physical Chemistry B*, 103:4570–4577, 1999.

[206] E. Lyman, J. Pfaendtner, and G. A. Voth. Systematic multiscale parameterization of heterogeneous elastic network models of proteins. *Biophysical Journal*, 95:4183–4192, 2008.

[207] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review*, 52:3730–3737, 1995.

[208] A. D. J. MacKerell, D. Bashford, M. Bellott, R. L. Bunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. I. I. I. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wlórkiewicz-Kuczera, D. Yin, M. Karplus, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. I. I. I. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wio, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102:3586–3616, 1998.

[209] A. D. MacKerell, M. Feig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 25:1400–1415, 2004.

[210] M. Maduke, C. Miller, and J. A. Mindell. A decade of ClC chloride channels: structure, mechanism, and many unsettled questions. *Annual Review of Biophysics and Biomolecular Structure*, 29:411–438, 2000.

[211] L. Makowski, D. J. Rodi, S. Mandava, D. D. L. Minh, D. B. Gore, and R. F. Fischetti. Molecular crowding inhibits intramolecular breathing motions in proteins. *Journal of Molecular Biology*, 375:529–546, 2008.

[212] L. Maragliano, G. Cottone, G. Ciccotti, and E. Vanden-Eijnden. Mapping the network of pathways of CO diffusion in myoglobin. *Journal of the American Chemical Society*, 132:1010–1017, 2010.

[213] L. Maragliano and E. Vanden-Eijnden. Single-sweep methods for free energy calculations. *Journal of Chemical Physics*, 128:184110, 2008.

[214] L. Maragliano and E. Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426:168–175, 2006.

[215] F. Marinelli and J. D. Faraldo-Gómez. Ensemble-biased metadynamics: A molecular simulation method to sample experimental distributions. *Biophysical Journal*, 108:2779–2782, 2015.

[216] S. J. Marrink and D. P. Tieleman. Perspective on the MARTINI model. *Chemical Society Reviews*, 42:6801–6822, 2013.

[217] S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse grained model for semiquantitative lipid simulations. *Journal of Physical Chemistry B*, 108:750–760, 2004.

[218] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 111:7812–7824, 2007.

[219] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino. Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations. *Journal of Physical Chemistry B*, 110:14011–14013, 2006.

[220] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30:2157–2164, 2009.

[221] R. Martoák, A. Laio, and M. Parrinello. Predicting crystal structures: The Parrinello-Rahman method revisited. *Physical Review Letters*, 90:075503, 2003.

[222] M. P. Mayer. Gymnastics of molecular chaperones. *Molecular Cell*, 39:321–331, 2010.

[223] M. McCullagh, M. G. Saunders, and G. A. Voth. Unraveling the mystery of ATP hydrolysis in actin filaments. *Journal of the American Chemical Society*, 136:13053–13058, 2014.

[224] M. McGovern and J. de Pablo. A boundary correction algorithm for metadynamics in multiple dimensions. *Journal of Chemical Physics*, 139:084102, 2013.

[225] P. Mehta and David J. Schwab. An exact mapping between the Variational Renormalization Group and Deep Learning. *arXiv*, 2014.

[226] S. Menard. *Logistic Regression: From Introductory to Advanced Concepts and Applications*. SAGE Publications, Inc., 2010.

[227] M. Mezei. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. *Journal of Computational Physics*, 68:237–248, 1987.

[228] D. Min, Y. Liu, I. Carbone, and W. Yang. On the convergence improvement in the metadynamics simulations: a Wang-Landau recursion approach. *Journal of Chemical Physics*, 126:194104, 2007.

[229] S. Miyamoto and P. a. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13:952–962, 1992.

[230] V. Molinero and E. B. Moore. Water modeled as an intermediate element between carbon and silicon. *Journal of Physical Chemistry B*, 113:4008–4016, 2009.

[231] M. Monteferrante, S. Bonella, S. Meloni, and G. Ciccotti. Modified single sweep method for reconstructing free-energy landscapes. *Molecular Simulation*, 35:1116–1129, 2009.

[232] M. Moradi and E. Tajkhorshid. Driven metadynamics: Reconstructing equilibrium free energies from driven adaptive-bias simulations. *Journal of Physical Chemistry Letters*, 4:1882–1887, 2013.

[233] F. Müller-Plathe. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *ChemPhysChem*, 9:754–769, 2002.

[234] J. W. Mullinax and W. G. Noid. Extended ensemble approach for deriving transferable coarse-grained potentials. *Journal of Chemical Physics*, 131:104110, 2009.

[235] J. W. Mullinax and W. G. Noid. Generalized Yvon-Born-Green Theory for Molecular Systems. *Physical Review Letters*, 103:198104, 2009.

[236] J. W. Mullinax and W. G. Noid. A generalized Yvon-Born-Green theory for determining coarse-grained interaction potentials. *Journal of Physical Chemistry C*, 114:5661–5674, 2010.

[237] J. W. Mullinax and W. G. Noid. Recovering physical potentials from a model protein databank. *Proceedings of the National Academy of Sciences of the United States of America*, 107:19867–19872, 2010.

[238] J. W. Mullinax and W. G. Noid. Reference state for the generalized Yvon-Born-Green theory: application for coarse-grained model of hydrophobic hydration. *Journal of Chemical Physics*, 133:124107, 2010.

[239] T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *Journal of Chemical Physics*, 126:075101, 2007.

[240] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen. Multiscale modeling of emergent materials: biological and soft matter. *Physical Chemistry Chemical Physics*, 11:1869–1892, 2009.

[241] T. Murtola, M. Karttunen, and I. Vattulainen. Systematic coarse graining from structure using internal states: Application to phospholipid/cholesterol bilayer. *Journal of Chemical Physics*, 131:055101, 2009.

[242] M. Ndao, J. Devémy, A. Ghoufi, and P. Malfreyt. Coarse-Graining the liquid-liquid interface with the MARTINI force field: How is the interfacial tension reproduced? *Journal of Chemical Theory and Computation*, 11:3818–3828, 2015.

[243] F. Noé, I. Horenko, C. Schütte, and J. C. Smith. Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *Journal of Chemical Physics*, 126:155102, 2007.

[244] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *Journal of Chemical Physics*, 139:090901, 2013.

[245] W. G. Noid, G. S. Ayton, S. Izvekov, and G. A. Voth. The multiscale coarse-graining method. In G. A. Voth, editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, pages 21–40. CRC Press, New York, 2008.

[246] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *Journal of Chemical Physics*, 128:244114, 2008.

[247] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. a. Voth. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *Journal of Chemical Physics*, 128:244115, 2008.

[248] W. G. Noid, J.-W. Chu, G. S. Ayton, and G. A. Voth. Multiscale coarse-graining and structural correlations: connections to liquid-state theory. *Journal of Physical Chemistry B*, 111:4116–4127, 2007.

[249] S. Nose. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81:511–519, 1984.

[250] T. Oda, M. Iwasa, T. Aihara, Y. Maéda, and A. Narita. The nature of the globular- to fibrous-actin transition. *Nature*, 457:441–445, 2009.

[251] A. R. Oganov, R. Martonák, A. Laio, P. Raiteri, and M. Parrinello. Anisotropy of Earth's D" layer and stacking faults in the MgSiO3 post-perovskite phase. *Nature*, 438:1142–1144, 2005.

[252] Y. Oono. Application of dimensional regularization to single chain polymer static properties: Conformational space renormalization of polymers. III. *Journal of Chemical Physics*, 74:6458–6466, 1981.

[253] A. Ostermann, R. Waschipky, F. G. Parak, and G. U. Nienhaus. Ligand binding and conformational motions in myoglobin. *Nature*, 404:205–208, 2000.

[254] L. R. Otterbein, P. Graceffa, and R. Dominguez. The crystal structure of uncomplexed actin in the ADP state. *Science*, 293:708–711, 2001.

[255] H. Paliwal and M. R. Shirts. Using multistate reweighting to rapidly and efficiently explore molecular simulation parameters space for nonbonded interactions. *Journal of Chemical Theory and Computation*, 9:4700–4717, 2013.

[256] A. C. Pan and B. Roux. Building Markov state models along pathways to determine free energies and rates of transitions. *Journal of Chemical Physics*, 129:064107, 2008.

[257] V. S. Pande, K. Beauchamp, and G. R. Bowman. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, 52:99–105, 2010.

[258] G. A. Pavliotis and A. M. Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts Applied in Mathematics*. Springer New York, New York, NY, 2008.

[259] J. Pfaendtner, D. Branduardi, M. Parrinello, T. D. Pollard, and G. A. Voth. Nucleotide-dependent conformational states of actin. *Proceedings of the National Academy of Sciences of the United States of America*, 106:12723–12728, 2009.

[260] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781–1802, 2005.

[261] S. E. V. Phillips. Structure and refinement of oxymyoglobin at 1.6 A resolution. *Journal of Molecular Biology*, 142:531–554, 1980.

[262] S. Piana, K. Lindorff-Larsen, and D. E. Shaw. Protein folding kinetics and thermodynamics from atomistic simulation. *Proceedings of the National Academy of Sciences of the United States of America*, 109:17845–17850, 2012.

[263] S. Piana and A. Laio. A bias-exchange approach to protein folding. *Journal of Physical Chemistry B*, 111:4553–4559, 2007.

[264] K. S. Pitzer. Electrolyte theory – Improvements since Debye and Hückel. *Accounts of Chemical Research*, 10:371–377, 1977.

[265] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1–19, 1995.

[266] T. D. Pollard and G. G. Borisy. Cellular motility driven by assembly and disassembly of actin filaments. *Cell*, 112:453–465, 2003.

[267] T. D. Pollard and J. a. Cooper. Actin, a central player in cell shape and movement. *Science*, 326:1208–1212, 2009.

[268] B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30:838–855, 1992.

[269] M. Praprotnik, L. delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *Journal of Chemical Physics*, 123:224106, 2005.

[270] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé. Markov models of molecular kinetics: generation and validation. *Journal of Chemical Physics*, 134:174105, 2011.

[271] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–854, 2013.

[272] B. Qi, S. Muff, A. Caflisch, and A. R. Dinner. Extracting physically intuitive reaction coordinates from transition networks of a beta-sheet miniprotein. *Journal of Physical Chemistry B*, 114:6979–6989, 2010.

[273] H.-J. Qian, P. Carbone, X. Chen, H. A. Karimi-Varzaneh, C. C. Liew, and F. Müller-Plathe. Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures. *Macromolecules*, 41:9919–9929, 2008.

[274] E. K. Rains and H. C. Andersen. A Bayesian method for construction of Markov models to describe dynamics on various time-scales. *Journal of Chemical Physics*, 133:144113, 2010.

[275] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *Journal of Physical Chemistry B*, 110:3533–3539, 2006.

[276] D. Reith, M. Putz, and F. Muller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24:1624–1636, 2003.

[277] S. Reuveni, J. Klafter, and R. Granek. Dynamic structure factor of vibrating fractals: Proteins as a case study. *Physical Review E*, 85:011906, 2012.

[278] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[279] J. L. Robertson, L. Kolmakova-Partensky, and C. Miller. Design, function and structure of a monomeric ClC transporter. *Nature*, 468:844–847, 2010.

[280] G. J. Rocklin, D. L. Mobley, and K. A. Dill. Calculating the sensitivity and robustness of binding free energy calculations of force field parameters. *Journal of Chemical Physics*, 9:3072–3083, 2013.

[281] L. Rosso and M. E. Tuckerman. An adiabatic molecular dynamics method for the calculation of free energy profiles. *Molecular Simulation*, 28:91–112, 2002.

[282] E. Rosta and G. Hummer. Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *Journal of Chemical Theory and Computation*, 11:276–285, 2015.

[283] M. Rubinstein and R. H. Colby. *Polymer Physics*. Oxford UP, Oxford, 2003.

[284] J. F. Rudzinski and W. G. Noid. Investigation of coarse-grained mappings via an iterative generalized Yvon-Born-Green method. *Journal of Physical Chemistry B*, 118:8295–8312, 2014.

[285] J. F. Rudzinski and W. G. Noid. Coarse-graining entropy, forces, and structures. *Journal of Chemical Physics*, 135:214101, 2011.

[286] V. Rühle and C. Junghans. Hybrid approaches to coarse-graining using the VOTCA package: Liquid hexane. *Macromolecular Theory and Simulations*, 20:472–477, 2011.

[287] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *Journal of Chemical Theory and Computation*, 5:3211–3223, 2009.

[288] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977.

[289] M. G. Saunders and G. A. Voth. Comparison between actin filament models: Coarse-graining reveals essential differences. *Structure (Oxford, U. K.)*, 20:641–653, 2012.

[290] M. G. Saunders, J. Tempkin, J. Weare, A. R. Dinner, B. Roux, and G. A. Voth. Nucleotide regulation of the structure and dynamics of G-actin. *Biophysical Journal*, 106:1710–1720, 2014.

[291] J.-W. Schen, C. Li, N. F. A. van der Vegt, and C. Peter. Transferability of coarse grained potentials: Implicit solvent models for hydrated ions. *Journal of Chemical Theory and Computation*, 7:1916–1927, 2011.

[292] H. A. Scheraga, M. Khalili, and A. Liwo. Protein-folding dynamics: Overview of molecular simulation techniques. *Annual Review of Physical Chemistry*, 58:57–83, 2007.

[293] F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, and P. A. Anfinrud. Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science*, 300:1944–1947, 2003.

[294] P. Shaffer, O. Valsson, and M. Parrinello. Enhanced, targeted sampling of high-dimensional free-energy landscapes using variationally enhanced sampling, with an application to chignolin. *Proceedings of the National Academy of Sciences of the United States of America*, 113:1150–1155, 2016.

[295] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330:341–346, 2010.

[296] M. S. Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *Journal of Chemical Physics*, 129:144108, 2008.

[297] M. S. Shell. Systematic coarse-graining of potential energy landscapes and dynamics in liquids. *Journal of Chemical Physics*, 137:084503, 2012.

[298] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *Journal of Chemical Physics*, 129:124105, 2008.

[299] M. R. Shirts and V. S. Pande. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *Journal of Chemical Physics*, 122:144107, 2005.

[300] P. B. Sigler, Z. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich. Structure and function in GroEL-mediated protein folding. *Annual Review of Biochemistry*, 67:581–608, 1998.

[301] M. Singh, D. Dhabal, A. H. Nguyen, V. Molinero, and C. Chakavarty. Triplet correlations dominate the transition from simple to tetrahedral liquids. *Physical Review Letters*, 112:147801, 2014.

[302] S. Singh, C.-c. Chiu, and J. J. de Pablo. Flux tempered metadynamics. *Journal of Statistical Physics*, 145:932–945, 2011.

[303] A. V. Sinitskiy, M. G. Saunders, and G. A. Voth. Optimal number of coarse-grained sites in different components of large biomolecular complexes. *Journal of Physical Chemistry B*, 116:8363–8374, 2012.

[304] A. V. Sinitskiy and G. A. Voth. Coarse-graining of proteins based on elastic network models. *Chemical Physics*, 422:165–174, 2012.

[305] L. Skjaerven, B. Grant, A. Muga, K. Teigen, J. A. McCammon, N. Reuter, and A. Martinez. Conformational sampling and nucleotide-dependent transitions of the GroEL subunit probed by unbiased molecular dynamics simulations. *PLoS Computational Biology*, 7:e1002004, 2011.

[306] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *Journal of Chemical Physics*, 109:7737–7744, 1998.

[307] V. Srajer, Z. Ren, T. Y. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff, and K. Moffat. Protein conformational relaxation and ligand migration in myoglobin: a nanosecond to millisecond molecular movie from time-resolved Laue X-ray diffraction. *Biochemistry*, 40:13802–13815, 2001.

[308] T. Stecher, N. Bernstein, and G. Csányi. Free energy surface reconstruction from umbrella samples using Gaussian process regression. *Journal of Chemical Theory and Computation*, 10:4079–4097, 2014.

[309] F. H. Stillinger, H. Sakai, and S. Torquato. Statistical mechanical models with effective potentials: Definitions, applications, and thermodynamic consequences. *Journal of Chemical Physics*, 117:288–296, 2002.

[310] F. H. Stillinger and T. A. Weber. Computer simulation of local order in condensed phases of silicon. *Physical Review B*, 31:5262–5271, 1985.

[311] G. Stölting, M. Fischer, and C. Fahlke. CLC channel function and dysfunction in health and disease. *Frontiers in Physiology*, 5:378, 2014.

[312] J. Sun, D. D. Klug, R. Martonák, J. A. Montoya, M.-S. Lee, S. Scandolo, and E. Tosatti. High-pressure polymeric phases of carbon dioxide. *Proceedings of the National Academy of Sciences of the United States of America*, 106:6077–6081, 2009.

[313] L. Sutto, S. Marsili, and F. L. Gervasio. New advances in metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2:771–779, 2012.

[314] R. H. Swendsen and J. S. Wang. Replica Monte Carlo simulation of spin glasses. *Physical Review Letters*, 57:2607–2609, 1986.

[315] C. F. Tejero and E. Lomba. Density-dependent interactions and thermodynamic consistency in integral equation theories. *Molecular Physics*, 107:349–355, 2009.

[316] E. Thiede, B. Van Koten, and J. Weare. Sharp entrywise perturbation bounds for Markov chains. 2014.

[317] I. F. Thorpe, D. P. Goldenberg, and G. A. Voth. Exploration of transferability in multiscale coarse-grained peptide models. *Journal of Physical Chemistry B*, 115:11911–11926, 2011.

[318] I. F. Thorpe, J. Zhou, and G. A. Voth. Peptide folding using multiscale coarse-grained models. *Journal of Physical Chemistry B*, 112:13079–13090, 2008.

[319] P. Tiwary, J. F. Dama, and M. Parrinello. A perturbative solution to metadynamics ordinary differential equation. *Journal of Chemical Physics*, 143:234112, 2015.

[320] P. Tiwary and M. Parrinello. A time-independent free energy estimator for metadynamics. *Journal of Physical Chemistry B*, 119:736–742, 2014.

[321] G. Torrie and J. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

[322] V. Tozzini. Minimalist models for proteins: a comparative analysis. *Quarterly Reviews of Biophysics*, 43:333–371, 2010.

[323] S. Trebst, D. A. Huse, and M. Troyer. Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations. *Physical Review E*, 70:046701, 2004.

[324] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185:604–613, 2013.

[325] G. A. Tribello, M. Ceriotti, and M. Parrinello. A self-learning algorithm for biased molecular dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 107:17509–17514, 2010.

[326] G. A. Tribello, M. Ceriotti, and M. Parrinello. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 109:5196–201, 2012.

[327] E. Trizac and Y. Levin. Renormalized jellium model for charge-stabilized colloidal suspensions. *Physical Review E*, 69:031403, 2004.

[328] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49:61–74, 1998.

[329] O. Valsson and M. Parrinello. Variational approach to enhanced sampling and free energy calculations. *Physical Review Letters*, 113:090601, 2014.

[330] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26:1701–1718, 2005.

[331] R. van Roij, M. Dijkstra, and J.-P. Hansen. Phase diagram of charge-stabilized colloidal suspensions: Van der Waals instability without attractive forces. *Physical Review E*, 59:2010–2025, 1999.

[332] J. VandeVondele and U. Rothlisberger. Canonical adiabatic free energy sampling (CAFES): A novel method for the exploration of free energy surfaces. *Journal of Physical Chemistry B*, 106:203–208, 2002.

[333] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31:NA–NA, 2009.

[334] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159:98–103, 1967.

[335] M. C. Villet and G. H. Fredrickson. Numerical coarse-graining of fluid field theories. *Journal of Chemical Physics*, 132:034109, 2010.

[336] A. Vitalis and A. Caflisch. Efficient construction of mesostate networks from molecular dynamics trajectories. *Journal of Chemical Theory and Computation*, 8:1108–1120, 2012.

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

[338] J. W. Wagner, J. F. Dama, and G. A. Voth. Predicting the sensitivity of multiscale coarse-grained models to their underlying fine-grained model parameters. *Journal of Chemical Theory and Computation*, 11:3547–3560, 2015.

[339] D. J. Wales. Perspective: Insight into reaction coordinates and dynamics from the potential energy landscape. *Journal of Chemical Physics*, 142:130901, 2015.

[340] D. Wang and G. A. Voth. Proton transport pathway in the ClC Cl-/H+ antiporter. *Biophysical Journal*, 97:121–131, 2009.

[341] F. G. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86:2050–2053, 2001.

[342] F. Wang and G. A. Voth. A linear-scaling self-consistent generalization of the multi-state empirical valence bond method for multiple excess protons in aqueous systems. *Journal of Chemical Physics*, 122:144105, 2005.

[343] H. Wang, C. Schütte, and L. Delle Site. Adaptive resolution simulation (AdResS): A smooth thermodynamic and structural transition from atomistic to coarse grained resolution and vice versa in a grand canonical fashion. *Journal of Chemical Theory and Computation*, 8:2878–2887, 2012.

[344] L.-P. Wang, J. Chen, and T. Van Voorhis. Systematic parameterization of polarizable force fields from quantum chemistry data. *Journal of Chemical Theory and Computation*, 9:452–460, 2013.

[345] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande. Systematic improvement of a classical molecular model of water. *Journal of Physical Chemistry B*, 117:9956–9972, 2013.

[346] Y. Wang, S. Feng, and G. A. Voth. Transferable coarse-grained models for ionic liquids. *Journal of Chemical Theory and Computation*, 5:1091–1098, 2009.

[347] Y. T. Wang and G. A. Voth. Molecular dynamics simulations of polyglutamine aggregation using solvent-free multiscale coarse-grained models. *Journal of Physical Chemistry B*, 114:8735–8743, 2010.

[348] Y. Wang, S. Izvekov, T. Yan, and G. a. Voth. Multiscale coarse-graining of ionic liquids. *Journal of Physical Chemistry B*, 110:3564–3575, 2006.

[349] Z.-G. Wang. Concentration fluctuation in binary polymer blends: chi parameter, spinodal and Ginzburg criterion. *Journal of Chemical Physics*, 117:481–500, 2002.

[350] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining equilibrium structure of simple liquids. *Journal of Chemical Physics*, 54:5237–5247, 1971.

[351] A. Wegner. Head to tail polymerization of actin. *Journal of Molecular Biology*, 108:139–150, 1976.

[352] T. Wendt, D. Taylor, K. M. Trybus, and K. Taylor. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proceedings of the National Academy of Sciences of the United States of America*, 98:4361–4366, 2001.

[353] A. D. White, J. F. Dama, and G. A. Voth. Designing free energy surfaces that match experimental data with metadynamics. *Journal of Chemical Theory and Computation*, 11:2451–2460, 2015.

[354] J. K. Whitmer, A. M. Fluitt, L. Antony, J. Qin, M. McGovern, and J. J. de Pablo. Sculpting bespoke mountains: Determining free energies with basis expansions. *Journal of Chemical Physics*, 143:044101, 2015.

[355] C. Willcox. Understanding hierarchical neural network behaviour: a renormalization group approach. *Journal of Physics A*, 24:2655–2664, 1991.

[356] K. G. Wilson. The renormalization group and critical phenomena. *Reviews of Modern Physics*, 55:583–600, 1983.

[357] W. Wojtas-Niziurski, Y. Meng, B. Roux, and S. Bernèche. Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. *Journal of Chemical Theory and Computation*, 9:1885–1895, 2013.

[358] C. F. Wong. Systematic sensitivity analyses in free energy perturbation calculations. *Journal of the American Chemical Society*, 113:3208–3209, 1991.

[359] C. F. Wong and H. Rabitz. Sensitivity analyses and principal component analysis in free energy calculations. *Journal of Physical Chemistry*, 95:9230–9628, 1991.

[360] C. F. Wong, T. Thacher, and H. Rabitz. Sensitivity analysis in biomolecular simulations. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, volume 12, pages 281–326. Wiley & Sons, New York, 1998.

[361] K.-y. Wong and D. M. York. Exact Relation between Potential of Mean Force and Free-Energy Profile. *Journal of Chemical Theory and Computation*, 8:3998–4003, 2012.

[362] Z. Xu, A. L. Horwich, and P. B. Sigler. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*, 388:741–750, 1997.

[363] L. Yang, G. Song, and R. L. Jernigan. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophysical Journal*, 93:920–929, 2007.

[364] J. Zavadlav, M. N. Melo, S. J. Marrink, and M. Praprotnik. Adaptive resolution simulation of polarizable supramolecular coarse-grained water models. *Journal of Chemical Physics*, 142:244118, 2015.

[365] J. Zavadlav, M. N. Melo, A. V. Cunha, A. H. de Vries, S. J. Marrink, and M. Praprotnik. Adaptive resolution simulation of MARTINI solvents. *Journal of Chemical Theory and Computation*, 10:2591–2598, 2014.

[366] Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. a. Voth. A systematic methodology for defining coarse-grained sites in large biomolecules. *Biophysical Journal*, 95:5073–5083, 2008.

[367] Z. Zhang, J. Pfaendtner, A. Grafmüller, and G. a. Voth. Defining coarse-grained representations of large biomolecules and biomolecular complexes from elastic network models. *Biophysical Journal*, 97:2327–2337, 2009.

[368] Z. Zhang and G. A. Voth. Coarse-grained representations of large biomolecular complexes from low-resolution structural data. *Journal of Chemical Theory and Computation*, 6:2990–3002, 2010.

[369] L. Zheng, M. Chen, and W. Yang. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105:20227–20232, 2008.

[370] W. Zheng, B. Qi, M. A. Rohrdanz, A. Caflisch, A. R. Dinner, and C. Clementi. Delineation of folding pathways of a β-sheet miniprotein. *Journal of Physical Chemistry B*, 115:13065–13074, 2011.

[371] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophysical Journal*, 92:4289–4303, 2007.

[372] S.-B. Zhu and C. F. Wong. Sensitivity Analysis of Water Thermodynamics. *Journal of Chemical Physics*, 98:8892–8899, 1993.