UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

UNDERSTANDING AND DESIGN OF MOLECULAR DYNAMICS BASED SIMULATION METHODS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

CHRISTOPHER ADAM HIXSON
Norman, Oklahoma
2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

UNDERSTANDING AND DESIGN OF MOLECULAR DYNAMICS BASED SIMULATION METHODS

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

_____

Prof. Ralph A. Wheeler, Chair

_____

Prof. Roger Frech

_____

Prof. Wai Tak Yip

_____

Prof. George B. Richter-Addo

_____

Prof. Kieran Mullen

Acknowledgements

A variety of institutions and individuals have joined to make the work presented here possible. Firstly, I thank my alma matter, the University of Tennessee, Chattanooga for providing such an excellent introduction to the field of chemistry. I also, obviously, thank the University of Oklahoma for giving me the chance to hone my chemical education, and for providing me with an Alumni Fellowship during my time here. I also would like to acknowledge the various granting agencies. Without the support of the Oklahoma Center for the Advancement of Science and Technology, the National Science Foundation, and the Department of Energy much of this work could not have been done. I also owe NCSA a great debt for the use of their computers as well as a larger debt to the Oklahoma Supercomputing Center for Education and Research, and Dr. Henry Neeman particularly for advice, support, and resources. Individually, I thank the members of my committee, who have given a great deal of support to me. Particularly, I owe a great deal to Professor Wheeler, my thesis advisor, for showing great confidence in my abilities. Also, Professor Mullen has been a source of inspiration in my conversations with him. The various member of my group have been nearly a family to me, and thanks especially go to Scott Boesch, Dr. Haitao Dong, Dr. Zunnan Huang, Jermont Chen, and Kurt Brorsen. Finally I owe the deepest debt to my ever loving family, and my partner of nine years, Shawn Carter.

Table of Contents

List of Tables

List of Figures

Abstract

Several new strategies for employing modified molecular dynamics (MD) protocols for finding energy minima were developed. First, we derived a new method of searching for the lowest energy conformation by employing the results of our work to explain "mean-field" molecular dynamics simulation methods. Called ensembles extracted by atomic coordinate transformations (EXACT), the method allows simulations to be performed with a variable degree of approximation. Then, we examine the previously published locally enhanced sampling (LES) approximation. The method makes copies of a small part of interest in a larger system, and allows the dynamics to unfold. The method works by making the copies invisible to each other and allowing them to interact only with the remainder of the system, called the bath. The bath, on the other hand, interacts with an averaged representation of the copied part. This averaged interaction allows the copied particles to move into geometries they might not visit in a conventional MD simulation, which allows a greater variety of structures to be sampled. We derive the algorithm by copying the entire system, and then by employing holonomic constraints between the bath particles between the various systems. Using this new approach, we explore several issues previously noted in the literature. We also use the EXACT approximation method to illustrate the nature of the LES approximation. Finally, we present another optimization method, which adds pressure along with temperature in analogue with simulated annealing. The method is tested against simulated annealing for condensed phase systems of argon, monoglyme, and tetraglyme. The method noticeably improves the

results for the glyme systems, but does not seem to hurt the results for the argon system.

Previously Published Works

This dissertation is composed of chapters which contain previously published works. This includes:

- Chapter 2, published as Christopher Adam Hixson and Ralph A. Wheeler, Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles. *Physical Review E* 2001, 64, 026701.

- Chapter 3, published as Chirstopher Adam Hixson and Ralph A. Wheeler, Practical multiple-copy methods for sampling statistical mechanical ensembles. *Chemical Physics Letters* 2004, 386, 330-335.

- Chapter 4, published as Christopher Adam Hixson, Jermont Chen, Zunnan Huang, and Ralph A. Wheeler, New perspectives on multiple-copy, mean-field molecular dynamics methods. *Journal of Molecular Graphics and Modeling* 2004, 22, (5), 349-357.

Each of these chapters is reprinted here by permission of the copyright holder.

CHAPTER 1

An Introduction to Computational Methods for Enhanced Sampling Dynamics

PROLOGUE

> *The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.*
>
> *Paul A. M. Dirac (attributed)*

Theoretical chemistry owes its existence almost entirely to modern physics, which recursively owes its existence to chemistry. Before people observed that clouds of hydrogen atoms absorbed electromagnetic radiation in an odd, discrete pattern (among other interesting phenomena), physics was considered almost a dead science. The discovery that these lines could be explained by the newly emerging quantum mechanics propelled physics to new heights, and simultaneously gave chemistry the beginnings of its now universally recognized fundamental underpinnings. Given this fact, I like to think that Dirac's words were prophetic. We now do routinely solve these equations, and other ones Gibbs might be more familiar with, with a degree of accuracy that might impress even the likes ones of the great scientists of the 20th century, Paul A. M. Dirac.

1.1 BASIC OVERVIEW OF THEORETICAL CHEMISTRY

The practice of theoretical chemistry is an important part of modern chemical research because the field has demonstrated its ability to make predictions,[1-6] to

resolve disagreements,[7-14] and to determine or estimate results.[15-18] Advancements in theoretical chemistry are also important research efforts in themselves because of the potential for impacting important areas of human understanding, such as protein folding.[19] Like any other scientific discipline, theoretical chemistry can be subdivided using a variety of classification schemes. One important way to divide this field is between qualitative and quantitative theories. For example, qualitative models that serve as a guide to understanding physical phenomena are useful for a variety of reasons. They provide insight, which can guide research activities, but also serve to define the language scientists use to communicate their research. For example, even though bond hybridization has not been used for decades among quantitative practitioners of theoretical chemistry, it is still one of the most basic bonding models taught to young scientists – from high school to graduate level chemistry. The power of this theory is now purely qualitative; it is an important way of communicating results for a large part of the chemical literature. More commonly used by contemporary theoretical chemists, though, are theories designed to produce an accurate answer to a quantitative problem. Quantitative theories are now generally the preferred approach, firstly, because quantitative problems can be expressed most easily in mathematics (which is the natural language of a theoretical science), and ultimately as computer programs. But it is also true that quantitative theories are favored because oftentimes more qualitative theories introduce misleading uncertainty. Problems in chemistry are notoriously sensitive to approximate treatments. Small differences in the value of the energy of a system, for example, might correspond to dramatic differences in the observed structure. To be most

useful, theoretical methods must be able to produce answers with "chemical accuracy"[20] (a few kcal mol^{-1}) or even "spectroscopic accuracy"[21] (a few cm^{-1}). Striving to achieve such results, two separate[7, 22] (increasingly compatible) approaches to theoretically solving quantitative chemical problems now exist: the quantum and classical approaches.

1.1.1 Quantitative Theories

In the first of these approaches, the quantum approach,[22, 23] problems are solved by describing the interactions between nuclei and electrons, the very things which determine the behavior of chemical systems. Strictly speaking, this method approaches chemical problems using the most appropriate tool for the job – by calculating an approximate solution to a molecule's Schrödinger equation (this, of course, glosses over differences between wavefunction methods and density functional methods, semi-empirical methods and ab initio methods, variational and nonvariational methods, etc.) [24] Information about wavefunctions, electron densities, energy levels, vibrational and rotational structure of molecules, and accurate predictions for molecular geometry are provided by such treatment. These data can be of great help in interpreting spectra, for example, since quantum mechanics can tell a great deal about the state of individual (or only weakly interacting) molecules.

In the second approach, the classical approach,[7] molecular systems are treated as billiard balls connected by springs. The system is designed so as to hold each molecule into a desirable shape which is generally assumed to bear some relation to an experimentally observed geometry. Before computers, similar simulations were performed using real mechanical molecular models.[25] However, methods that use the

classical approach are fraught with all the dangers inherent in using an empirical model to solve a problem. The results are heavily dependant on the quality of the model, and the quality varies greatly from system to system. Because of such concerns, the quantum approach is generally used in each technically feasible opportunity.

Unfortunately, the quantum approach requires more computer resources than is generally practically available if the size of the molecule of interest is too large or, certainly, if the behavior of large collections of relatively strongly-interacting molecules is to be studied. In this situation, which is typically the domain of statistical mechanics and thermodynamics, the classical approximation is overwhelmingly favored – if for no other reason than the fact that it is the only approach practically possible when using a realistic system is required for the analysis. Of course, quantum molecular dynamics has recently begun to bridge this gap, and further improvements are topics of active research.[26-28] Also, sometimes quantum problems can be reformulated to be solved using molecular dynamics or *Monte Carlo* methods using a classical potential, as in path integral techniques.[29, 30] Still, results from the classical approach can be used to calculate diffusion times, thermodynamic information, paths, and generally any other measurement where complete knowledge of the trajectory of the complete system is sufficient information.[31, 32] The work presented in the following chapters generally advances the classical approach.

1.1.2 Molecular Dynamics and *Monte Carlo* – Classical dynamics applied to chemistry

The classical approach mentioned above is deemed classical because the particles can be assumed to have simultaneously known locations and momentum, with a potential energy generally defined to be:

$$V(\vec{X}) = \sum_{overatompairs}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{eq_iq_j}{4\pi\varepsilon_0 r_{ij}}\right) + \sum_{overbondpairs} k_n(r_{ij} - r_n)^2 + \sum_{overanglepairs} k_n(\theta_{ijk} - \theta_n)^2 + \sum_{overtorsionpairs}\sum_{overtorsionterms} K_n(1 + \cos(m(\varphi_{nm} - \varphi_{0m}) - \gamma_m)).$$

The first term of this equation describes the van der Waals and electrostatic potentials. The van der Waals interaction is represented by a (12-6) Lennard Jones potential,[33] with A and B acting as the required constants, provided by the designer of the potential energy function, and r is the distance between the atom pairs. In the electrostatic term, the q's are the partial-charge of the atom which is also provided by the designer of the potential energy function. The next two terms represent the springs in the ball and spring model. The first of these represents the spring holding the molecule's bond in shape, the second acts to hold the angles into shape, and the final term (which is not strictly harmonic, but does have a similar shape) holds the dihedral angles fixed in their proper shape.[7] Each of these terms use constants provided by the designer of the potential energy function, as well. Such parameterized potential energy functions are usually called "force fields"[34] in the literature of this field. This force field is used (at least when using this approach to

calculate statistical and thermodynamic quantities) to generate a collection of states (or particular examples of the system) that belong to a desired ensemble. Among the many ensembles available to be considered, the canonical ensemble (NVT) and the isothermal-isobaric ensemble (NPT) are most commonly encountered in computational chemistry. There are (basically) two competing methods available to generate such collections. In the first technique, *Monte Carlo*,[35] some version of Metropolis importance sampling[36] is generally employed. This method, simply explained, takes an initial state, and then performs a trial move. The trial move is accepted or rejected depending (at least in the canonical ensemble) solely on the energy difference between the two states. If the energy of the state after the trial move is lower than the previous state's energy, then the move is always accepted. If the energy is higher, the state is accepted with a probability corresponding to the Boltzmann distribution:

$$\rho(\Delta E) = \exp(-\beta \Delta E),$$

where ΔE is the difference in energy between the two states, β is $(kT)^{-1}$, the Boltzmann constant is k, and T is the temperature. This method has several advantages, not the least of which is computational. *Monte Carlo* methods are among the easiest methods to parallelize, and computer scientists describe it as an "embarrassingly parallel"[35] computational problem. However, despite its advantages, it is deficient in certain chemical problems, most notably those requiring knowledge of a trajectory, because the "path" a *Monte Carlo* calculation takes is typically not physical.

Molecular dynamics, on the other hand, takes a system from one state to the next following Newton's laws of motion. This can be accomplished using a variety of algorithms; one common example is the velocity Verlet algorithm.[7] Molecular dynamics trajectories are suitable replacements for most applications of *Monte Carlo*, plus they represent physical paths. However, since the path the simulation takes must be physical, molecular dynamics is generally considered to have a few disadvantages. First, it is much more difficult to make a MD calculation parallel. Efforts that improve accuracy and computational efficiency have been taken to improve electrostatic potential calculations.[37] More importantly, molecular dynamics simulations undersample some molecular motions and have been shown to require more computational effort to achieve results that converge as well as comparable *Monte Carlo* simulations.[38, 39] The methods described in the next section were designed, in part, to provide researchers methods to improve calculations based on molecular dynamics.

1.2 MD METHODS TO IMPROVE "SAMPLING"

It has been shown that molecular dynamics converges 2-3 times slower than *Monte Carlo* calculations do when estimating structural and thermodynamic properties of a typical condensed phase system.[39] Though the previous fact is perhaps somewhat controversial, it is true that a common complaint found in the literature is that the MD method provides poor "sampling." That is to say that, even though a system might be assumed to be ergodic (which is to say that all areas of the system can be sampled given enough time), typical molecular dynamics simulations only slowly visit all accessible areas. The systems will explore a single energy "well" for

a long period of time, at least at temperatures and pressures of usual chemical interest, with only rare transitions to another configuration. To address this particular complaint, a variety of methods have been proposed seeking to improve the performance of molecular dynamics. In addition, some tasks, such as finding minima, take considerable time, regardless of the method, and related efforts have been made to improve their performance. One of the simplest improvements that can be made to either molecular dynamics or *Monte Carlo* simulations is simulated annealing,[36, 40] which aids in finding minima. Other techniques designed to improve "sampling" include locally enhanced sampling (LES),[41, 42] collisional LES (cLES),[43] replica exchange,[44] and other less frequently used ideas.[45-48] These methods will be introduced briefly below.

1.2.1 Simulated annealing

Simulated annealing is an approach taken to optimize a quantity, often molecular geometry when applied to chemical problems, and can be coupled with either molecular dynamics,[49-51] *Monte Carlo*,[52-54] or other techniques capable of generating a suitable ensemble.[36] The method is quite general, and has been used to study a wide variety of situations where a "cost function" can be defined. The cost function, analogous to the energy in chemical problems, is a function for which extrema are sought. Examples of other problems simulated annealing is well suited for include the "traveling salesman" problem, optimally designing electrical circuits, reconstructing images and sounds, and many others.[36]

Simulated annealing has been proven to be an effective optimization technique in certain circumstances[55] and its validity can be rather easily understood using a few

concepts known to chemists. In its initial state, a system is located in some well inside the system's potential energy surface. From its initial position, subsequent heating increases the probability that the system will move to ever higher energies, until the rugged features of the potential energy surface no longer constrain it. Finally, cooling increases the probability that the system will lose energy to the surroundings, and it will once again become trapped inside a well in the potential energy surface. The Boltzmann distribution makes it seem reasonable that the system will have a better chance of becoming trapped in a deeper well, and repeated applications of the heating/cooling cycle make it seem even more reasonable. Of course, control over the convergence of simulated annealing methods requires careful study of the "cooling schedule."[56]

Simulated annealing has been used widely as an optimization technique for chemical systems. For example, it is still a widely popular method for crystallographic and NMR structure refinement[51, 53] and is used as a geometry optimization technique for both small molecules[50] and peptides.[54] It can also be used to optimize mathematical functions, for example, solving the Ginzberg-Landau equations for a magnetic system. [52]

1.2.2 "Mean Field" Methods - Locally Enhanced Sampling

Locally enhanced sampling (LES) is a simulation method developed in 1990[41] and initially designed to provide increased efficiency in observing escape times and trajectories of a gas molecule trapped inside a globular protein. It belongs to a wider family of techniques called "mean-field" methods[42] that includes the multiple-copy simultaneous search (MCSS) algorithm,[57] which is used in drug design. In addition, a

version of LES with improved handling of collisions was also developed, which improved the estimated escape times, as well as other properties of the simulation.[43] This method, called cLES, or collisional LES, detects collisions between atoms and handles them as a special case. LES was inspired by an earlier method, called trajectory bundles, designed to study the vibrational modes of small molecules.[58]

LES, simply described, is a simulation technique where the system under study is divided into two parts. The first part, a small part of the system of interest to the researcher, is copied several times. In the original LES application, the copied part was the diffusing gas particles. None of the copies interact with each other, but instead interact normally with the remainder of the system, called the bath. The bath in the original example was the protein. The particles in the bath, of course, interact normally with each other, but interact with the average structure of each of the copied particles. The average interaction is both the source of the advantages, and as was discovered, the source of many undesirable properties in "mean-field" simulations.

LES has been applied beyond its initial role, and is widely touted as a method that improves the performance of molecular dynamics in searching for low energy states of a system[59-61] and in finding free-energy differences,[62] besides the original gas diffusion example. It has also been shown to be useful in understanding the effects of a mutation in myoglobin.[63] Each of these tasks use the brute force approach offered by pure molecular dynamics and represent studies where LES has had mixed success.

Because of its relative popularity, LES has been implemented in the majority of commercial and high-quality academic molecular dynamics programs including CHARMM,[64, 65] AMBER,[66] and NAMD.[67] It is used, however, despite several issues

that have been know almost since the method was invented. Among the issues that the initial presentation did not address include the "temperature-disparity problem,"[68-70] which is the observation that the temperature of the copied portion of the simulation is always hotter than the bath; the related cooling problem,[69] which is observed when an excited atom in the bath fails to cool as it would be expected to; and the potential energy surface problem,[71] so named because minima on the LES potential energy surface do not always correspond to minima on the real potential energy surface.

1.2.3 Replica exchange

The replica-exchange molecular dynamics method, belongs to a superset of methods called "generalized ensemble" methods, of which simulated tempering and the multicanonical algorithm are members.[44] Replica exchange, in particular, has become a popular tool for finding free energy minima for peptides and for calculating free-energy differences and other thermodynamic quantities. Replica exchange has been of particular use to the study of protein folding. Though much younger than simulated annealing, the replica exchange method has been implemented in most of the popular molecular dynamics packages, including CHARMM,[64, 65] AMBER,[66] and NAMD.[67]

Replica exchange works by making several copies of the entire system, but simulating each at a different temperature, ranging from very cold to very hot. The copies do not directly interact. One system is kept at each of the temperatures. At a predefined interval, the simulation attempts to exchange the temperature of one of the systems with another (by exchanging atomic velocities), and the temperature change

is accepted or rejected according to a Metropolis-like probability. Changes between systems closer in temperature are more likely than those further apart, so generally only exchanges between neighboring temperatures are attempted. Simulating in this way has been shown to improve the rate of conformational sampling, and since the method was formulated to allow obtaining thermodynamic quantities, it has been so used.

In the first published work involving replica exchange molecular dynamics,[44] a pentapeptide called met-enkephalin was studied. This work demonstrated that the method evenly sampled a wide range of potential energy values over the course of the simulation, whereas conventional molecular dynamics sampled approximately the same range, but with gaps. Replica exchange also sampled a greater variety of torsion angle conformations. Because of its sampling ability is has been applied to measuring free energy differences using umbrella sampling,[72] and also to measure the free energy of folding for a model peptide.[73] But because of its great ability to explore conformational space, it has been used most often to search pathways important to protein folding.[74-76]

1.2.4 Other interesting methods

In addition to the widely used methods discussed above, there are many other ideas that deserve attention. Despite their relatively less frequent use, many methods are clever attempts to improve sampling, and can provide insight for future efforts. Among these are included the diffusion equation method,[47] the Gaussian phase packet method,[46] SWARM-MD,[45] and a variable transformation method.[48] In first method listed, the diffusion equation method, the function to be optimized (generally the

potential energy function of the system) is transformed in such a way as to introduce a smoothing parameter, which reduces the depth of the least "important" shallower wells. The minima are found using the transformed function, and then returned to their original form. In the Gaussian phase packet method, dynamics were derived for the case that atoms are not located at fixed points, but rather somewhere inside a Gaussian distribution. The averaged nature of the interactions between particles smoothed potential energy barriers. SWARM-MD, on the other hand, allows that important parts of the system can be copied. Then the simulation proceeds a flocking algorithm which causes the copied parts to mimic the swarm-like way birds fly which causes the algorithm to sample parts of the energy surface it would not ordinarily explore. Finally, Zhu, Tuckerman, Samuelson, and Martyna[48] designed a clever variable transformation that when applied to the canonical partition function created an effective potential energy function that was much smoother than the original. The method was shown to sample configurations of relatively complicated molecules effectively in a manner that preserved all of the properties desirable in a molecular dynamics calculation. The approach, though clever and deserving of further study, proved unwieldy.

1.3 OUR CONTRIBUTIONS TO THE FIELD

The three main methods described above: simulated annealing, "mean field" methods, and replica exchange are each common methods used to study systems of chemical interest using molecular dynamics, and each has been addressed by this group. This dissertation addresses simulated annealing and the "mean field" methods directly, in particular "mean field" methods are heavily addressed in the following

chapters. We derive mean field equations to illuminate the limitations of LES and correct them. First, approximations to our equations are shown to provide an alternate derivation of LES, the most used "mean-field" method, and thus clarify several of its limitations. We also implemented the equations by designing a computational method to further illuminate (and correct) limitations of LES. In the final chapter, we demonstrate a method that uses variable pressure, along with temperature, to conduct simulated annealing to find low energy structures of model polymer systems.

All of the work described in Chapters 2-4 was performed using "folly", a suite of programs developed by our group. Providing a complete molecular dynamics implementation, folly reads the initial state of the system, integrates the appropriate equations of motion, and provides output consisting of the details of the motion. Using the AMBER 94 and 99 force fields, folly implements a large number of the features in popular MD programs including the velocity Verlet integration algorithm, Ewald sums, the generalized born/surface area (GB/SA) implicit solvent model, the SHAKE algorithm for maintaining constant bond distances, and the Nosé-Hoover chain thermostat. Each of these features was not only implemented with conventional MD, but also in our implementation of LES and our EXACT approximation (the ensembles extracted by atomic coordinate transformation approximation). Extending some of these methods to be used in the LES and EXACT approximations is partially the subject of Chapter Three.

1.4 REFERENCES

1. Bianco, R.; Hynes, J. T., Heterogeneous Reactions Important in Atmospheric Ozone Depletion: A Theoretical Perspective. *Accounts of Chemical Research* **2006,** 39, (2), 159-165.

2. DesJarlais, R. L.; Dixon, J. S., A shape- and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. *Journal of Computer-Aided Molecular Design* **1994,** 8, (3), 231-42.

3. Enemark, J. H.; Feltham, R. D., Principles of structure, bonding, and reactivity for metal nitrosyl complexes. *Coordination Chemistry Reviews* **1974,** 13, (4), 339-406.

4. Kirby, K. P., Applications of molecular structure methods to problems in astrochemistry. *Advanced Series in Physical Chemistry* **1995,** 2, 1241-79.

5. Walden, S. E.; Wheeler, R. A., Protein conformational gate controlling binding site preference and migration for ubiquinone-B in the photosynthetic reaction center of Rhodobacter sphaeroides. *Journal of Physical Chemistry B* **2002,** 106, (11), 3001-3006.

6. Woodward, R. B.; Hoffmann, R., Stereochemistry of Electrocyclic Reactions. *Journal of the American Chemical Society* **1965,** 87, (2), 395-397.

7. Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*. Clarendon Press: Oxford, 1987.

8. Allen, T. L.; Fink, W. H.; Power, P. P., Theoretical studies of multiple bonds in gallium-gallium and germanium-germanium compounds. *Dalton* **2000**, (3), 407-412.

9. Cotton, F. A.; Cowley, A. H.; Feng, X., The Use of Density Functional Theory To Understand and Predict Structures and Bonding in Main Group Compounds with Multiple Bonds. *Journal of the American Chemical Society* **1998,** 120, (8), 1795-1799.

10. Cotton, F. A.; Feng, X., Remarks on the Gallium to Iron Bond in an ArGaFe(CO)4 Molecule. *Organometalics* **1998,** 17, (2), 128-130.

11. Su, J.; Li, X.-W.; Crittendon, R. C.; Campana, C. F.; Robinson, G. H., Experimental Confirmation of an Iron-Gallium Multiple Bond: Synthesis, Structure, and Bonding of a Ferrogallyne. *Organometalics* **1997,** 16, (21), 4511-4513.

12. Su, J.; Li, X.-W.; Crittendon, R. C.; Robinson, G. H., How Short is a Ga.tplbond.Ga Triple Bond? Synthesis and Molecular Structure of Na2[Mes*2C6H3Ga.tplbond.GaC6H3Mes*2] (Mes* = 2,4,6-i-Pr3C6H2): The First Gallyne. *Journal of the American Chemical Society* **1997,** 119, (23), 5471-5472.

13. Xie, Y.; Grev, R. S.; Gu, J.; Schaefer, H. F., III; Schleyer, P. v. R.; Su, J.; Li, X.-W.; Robinson, G. H., The Nature of the Gallium-Gallium Triple Bond. *Journal of the American Chemical Society* **1998,** 120, (15), 3773-3780.

14. Xie, Y.; Schaefer, H. F., III; Robinson, G. H., The gallium-gallium triple bond in a realistic model. A density functional theory study of Na2[(C6H5)2C6H3GaGaC6H3(C6H5)2]. *Chemical Physics Letters* **2000,** 317, (1,2), 174-180.

15. Hisatani, K.; Okajima, K.; Kamide, K., Quantum chemical conformational analysis on isotactic polyacrylonitrile using molecular model compounds. *Polymer Journal (Tokyo)* **1996,** 28, (2), 99-105.

16. Scott, K. A.; Randles, L. G.; Moran, S. J.; Daggett, V.; Clarke, J., The folding pathway of spectrin R17 from experiment and simulation: Using experimentally validated MD simulations to characterize states hinted at by experiment. *Journal of Molecular Biology* **2006,** 359, (1), 159-173.

17. Visser, H.; Dube, C. E.; Armstrong, W. H.; Sauer, K.; Tachandra, V. K., FTIR Spectra and Normal-Mode Analysis of a Tetranuclear Manganese Adamantane-like Complex in Two Electrochemically Prepared Oxidation States: Relevance to the Oxygen-Evolving Complex of Photosystem II. *Journal of the American Chemical Society* **2002,** 124, (37), 11008-11017.

18. Wang, B.; Merz, K. M., Jr., A Fast QM/MM (Quantum Mechanical/Molecular Mechanical) Approach to Calculate Nuclear Magnetic Resonance Chemical Shifts for Macromolecules. *Journal of Chemical Theory and Computation* **2006,** 2, (1), 209-215.

19. Suits, F.; Pitman, M. C.; Pitera, J. W.; Swope, W. C.; Germain, R. S., Overview of molecular dynamics techniques and early scientific results from the Blue gene project. *IBM Journal of Research and Development* **2005,** 49, (2/3), 475-487.

20. Helgaker, T.; Ruden, T. A.; Jorgensen, P.; Olsen, J.; Klopper, W., A priori calculation of molecular properties to chemical accuracy. *Journal of Physical Organic Chemistry* **2004,** 17, (11), 913-933.

21. Kutzelnigg, W.; Von Herigonte, P., Electron correlation at the dawn of the 21st century. *Advances in Quantum Chemistry* **1999,** 36, 185-229.

22. Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry*. Dover Publications, Inc.: Mineola, NY, 1996.

23. Parr, R. G.; Yang, W., *Density-Functional Theory of Atoms and Molecules*. Oxford Science Publications: New York, 1989.

24. Levine, I. N., *Quantum Chemistry*. 5th ed.; Prentice Hall: 1999.

25. Kettering, C. F.; Shutts, L. W.; Andrews, D. H., A Representation of the Dynamic Properties of Molecules by Mechanical Models. *Physical Review* **1930,** 36, 531-543.

26. Elstner, M.; Frauenheim, T.; Suhai, S., An approximate DFT method for QM/MM simulations of biological structures and processes. *THEOCHEM* **2003,** 632, 29-41.

27. Titmuss, S. J.; Cummins, P. L.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E., Comparison of linear-scaling semiempirical methods and combined quantum mechanical/molecular methods applied to enzyme reactions. *Chemical Physics Letters* **2000,** 320, (1,2), 169-176.

28. Titmuss, S. J.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E., Comparison of linear-scaling semiempirical methods and combined quantum mechanical/molecular mechanical methods for enymic reactions. II. An energy decomposition analysis. *Journal of Computational Chemistry* **2002,** 23, (14), 1314-1322.

29. Feynman, R. P., *Statistical Mechanics: A Set of Lectures*. W. A. Benjamin, Inc.: Reading, MA, 1972.

30. Makri, N., Time-dependent quantum methods for large systems. *Annual Review of Physical Chemistry* **1999,** 50, 167-191.

31. Brooks, C.; Case, D. A., Simulations of peptide conformational dynamics and thermodynamics. *Chemical Reviews* **1993,** 93, (7), 2487-2502.

32. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **1993,** 93, (7), 2395-2417.

33. Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B., *Molecular Theory of Gases and Liquids*. John Wiley and Sons: New York, 1964.

34. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.

35. Wilson, M. R.; Ilnytskyi, J. M., Parallel computer simulation techniques for the study of macromolecules. In *Computer simulations of liquid crystals and polymers*, Pasini, P.; Zannoni, C.; Zumer, S., Eds. Kluwer: 2004; pp 335-356.

36. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

37. Sagui, C.; Darden, T., Multigrid methods for classical molecular dynamics simulations of biomolecules. *Journal of Chemical Physics* **2001,** 114, (15), 6578-6591.

38. Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips, J., G. N., A sampling problem in molecular dynamics simulations of macromolecules. *Proceeding of the National Academy of Sciences* **1995,** 92, 3288-3292.

39. Jorgensen, W. L.; Tirado-Rives, J., Monte carlo vs molecular dynamics for conformational sampling. *Journal of Physical Chemistry* **1996,** 100, 14508-513.

40. Tsallis, C.; Stariolo, D. A., Generalized simulated annealing. *Physica A* **1996,** 233, (1-2), 395-406.

41. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

42. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

43. Ulitsky, A.; Elber, R., Application of the Locally Enhanced Sampling (Les) and a Mean-Field with a Binary Collision Correction (Cles) to the Simulation of Ar Diffusion and No Recombination in Myoglobin. *Journal of Physical Chemistry* **1994,** 98, (3), 1034-1043.

44. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** 314, 141-151.

45. Huber, T.; van Gunsteren, W. F., SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A* **1998,** 102, (29), 5937-5943.

46. Ma, J. P.; Hsu, D.; Straub, J. E., Approximate Solution of the Classical Liouville Equation Using Gaussian Phase Packet Dynamics - Application to Enhanced Equilibrium Averaging and Global Optimization. *Journal of Chemical Physics* **1993,** 99, (5), 4024-4035.

47. Piela, L.; Kostrowicki, J.; Scheraga, H. A., The Multiple-Minima Problem in the Conformational-Analysis of Molecules - Deformation of the Potential-Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry* **1989,** 93, (8), 3339-3346.

48. Zhu, Z. W.; Tuckerman, M. E.; Samuelson, S. O.; Martyna, G. J., Using novel variable transformations to enhance conformational sampling in molecular dynamics. *Physical Review Letters* **2002,** 88, (10).

49. Brunger, A. T.; Krukowski, A., Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallographica A* **1990,** 46, 585-593.

50. Hohl, D.; Jones, R. O.; Car, R.; Parrinello, M., Structure of sulfur clusters using simulated annealing: S2 to S13. *Journal of Chemical Physics* **1988,** 89, (11), 6823-6835.

51. Nilges, M.; Gronenborn, A. M.; Brunger, A. T.; Clore, G. M., Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Engineering* **1988,** 2, (1), 27-38.

52. Doria, M. M.; Gubernatis, J. E.; Rainer, D., Solving the Ginzberg-Landau equations by simulated annealing. *Physical Review B* **1990,** 41, (10), 6335-6340.

53. Goodsell, D. S.; Olson, A. J., Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Genetics* **1990,** 8, 195-202.

54. Wilson, S. R.; Cui, W. L., Applications of simulated annealing to peptides. *Biopolymers* **1990,** 29, (1), 225-35.

55. Catoni, O. In *Simulated annealing algorithms and Markov chains with rare transitions*, Seminaire de Probabilites, Strasbourg, 1999; Stringer-Verlang: Strasbourg, 1999.

56. Cohn, H.; Fielding, M., Simulated annealing: Searching for an optimal temperature schedule. *SIOPT* **1999,** 9, (3), 779-802.

57. Miranker, A.; Karplus, M., Functionality Maps of Binding-Sites - a Multiple Copy Simultaneous Search Method. *Proteins-Structure Function and Genetics* **1991,** 11, (1), 29-34.

58. Gerber, R. B.; Buch, V.; Ratner, M. A., Time-Dependent Self-Consistent Field Approximation for Intramolecular Energy-Transfer .1. Formulation and Application to Dissociation of Vanderwaals Molecules. *Journal of Chemical Physics* **1982,** 77, (6), 3022-3030.

59. Roitberg, A.; Elber, R., Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* **1991,** 95, (12), 9277-87.

60. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

61. Simmerling, C.; Miller, J. L.; Kollman, P. A., Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid. *Journal of the American Chemical Society* **1998,** 120, (29), 7149-7155.

62. Simmerling, C.; Fox, T.; Kollman, P. A., Use of locally enhanced sampling in free energy calculations: Testing and application to the alpha ->beta anomerization of glucose. *Journal of the American Chemical Society* **1998,** 120, (23), 5771-5782.

63. Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N., Jr.; Dou, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; et al., Structural and functional effects of apolar mutations of the distal valine in myoglobin. *Journal of Molecular Biology* **1995,** 245, (4), 416-36.

64. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Sqaminathan, S.; Karplus, M., CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** 4, 187-217.

65. MacKerell, J., A. D.; Brooks, B.; Brooks III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M., CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. In *The Encyclopedia of Computational Chemistry*, Schleyer, P. v. R.; al., e., Eds. John Wiley and Sons: Chichester, 1998; Vol. 1, pp 271-277.

66. Case, D. A.; Darden, T.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*, University of California, San Francisco: 2004.

67. Phillips, J. C.; al., e., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** 26, 1781-1802.

68. Straub, J. E.; Karplus, M., Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation. *Journal of Chemical Physics* **1991,** 94, (10), 6737-6739.

69. Ulitsky, A.; Elber, R., The Thermal-Equilibrium Aspects of the Time-Dependent Hartree and the Locally Enhanced Sampling Approximations - Formal Properties, a Correction, and Computational Examples for Rare-Gas Clusters. *Journal of Chemical Physics* **1993,** 98, (4), 3380-3388.

70. Zheng, W. M.; Zheng, Q., An analytical derivation of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1997,** 106, (3), 1191-1194.

71. Stultz, C. M.; Karplus, M., On the potential surface of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1998,** 109, (20), 8809-8815.

72. Murata, K.; Sugita, Y.; Okamoto, Y., Molecular dynamics simulations of DNA dimers based on replica-exchange umbrella sampling. II. Free energy analysis. *Journal of Theoretical and Computational Chemistry* **2005,** 4, (2), 433-448.

73. Baumketner, A.; Shea, J.-E., The Thermodynamics of Folding of a Beta Hairpin Peptide Probed Through Replica Exchange Molecular Dynamics Simulations. *Theoretical Chemistry Accounts* **2006,** 116, (1-3), 262-273.

74. Gnankaran, S.; Nussinov, R.; Garcia, A. E., Atomic-level description of amyloid beta-dimer formation. *Journal of the American Chemical Society* **2006,** 128, (7), 2158-2159.

75. Larios, E.; Pitera, J. W.; Swope, W. C.; Gruebele, M., Correlation of early orientational order of engineered lambda6-85 structure with kinetics and thermodynamics. *Chemical Physics* **2006,** 323, (1), 45-53.

76. Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L., The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure. *Journal of Molecular Biology* **2006,** 360, (5), 1094-1107.

CHAPTER 2

Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles

2.1 INTRODUCTION

More than ten years ago, Elber and Karplus presented a multiple-copy molecular dynamics method designed to accelerate a simulation's convergence to a global energy minimum.[1] The method, called locally enhanced sampling LES, is based on creating a set of non-interacting copies of a small subsystem of primary interest and allowing a larger subsystem, the "bath," to interact with each copy of the sub-system. The force each copied atom experiences from the bath is the total force the corresponding real atom would experience. The bath atoms, on the other hand, experience the average of the forces due to the copied atoms. As a result, energy barriers that copies of the subsystem must overcome to avoid being trapped in local energy minima are decreased compared to those in a conventional molecular dynamics simulation. LES and related mean field methods, have been used successfully in a variety of optimization problems,[1-16] but they suffer limitations common to other *ad hoc* geometry optimization methods:[17-25] since the underlying energy surface and/or its sampling is modified, the methods do not generate trajectories that correspond to any of the familiar statistical mechanical ensembles. So, phase-space averaging over these trajectories is not useful in a statistical

mechanical sense and any information obtained in this manner (free energy differences, radial distribution functions, or even temperatures) must be used with caution. It should come as no surprise, then, that data acquired from LES has been found to violate fundamental principles, including the equipartition of energy theorem.[26, 27] It has been claimed[26-28] that this manifests itself as the "temperature disparity problem," which is a failure of the subsystem and bath temperatures to reach the same equilibrium value. On a separate issue, Stultz and Karplus have provided a proof that minima located using LES are not necessarily minima on the original energy surface.[29]

Although optimization methods based on mean field theory show great practical utility,[1-16, 30-32] current multiple-copy implementations are clearly flawed. The current contribution was inspired by LES, but our approach is intended to improve phase-space sampling by using a rigorously derived method, so our approximations are well known and may be controlled. By controlling the approximations, we hope to generate trajectories that more closely approximate trajectories expected of systems belonging to one of the well-studied statistical mechanical ensembles. Our starting point is fundamentally different from that of Elber and Karplus, but it is nonetheless instructive to reiterate briefly their rationale for LES. Next we describe known drawbacks of LES and attempts to remedy them. Finally, we derive our generalization of multiple-copy methods, present a numerical test, and relate our algorithm to existing multiple-copy methods.

2.2 CONVENTIONAL MULTIPLE-COPY MOLECULAR DYNAMICS

As it was originally presented,[1] LES flows from the assumption that the classical phase-space density can be written as

$$\rho\left(\vec{\mathbf{X}},t\right)=\rho_s\left(\vec{\mathbf{X}}_s,t\right)\rho_b\left(\vec{\mathbf{X}}_b,t\right), \qquad (2.1)$$

where ρ_s is the density of the subsystem to be copied, ρ_b is the density of the bath, and $\vec{\mathbf{X}}$ is the vector that indicates the system's location in phase space. It is also assumed that the bath's density can be written as a single delta function:

$$\rho_b\left(\vec{\mathbf{X}},t\right)=\delta\left(\vec{\mathbf{X}}_b(t)\right), \qquad (2.2)$$

while the copied subsystem's density can be taken to be a "swarm" of delta functions, such as

$$\rho_s\left(\vec{\mathbf{X}}_s,t\right)=\sum_{k=1}^{C} w_{sk}\delta\left(\vec{\mathbf{X}}_s(t)\right). \qquad (2.3)$$

Equation 2.3 is a construct describing the positions in phase space of the various copies, weighted by w_{sk}. In LES, the weights are taken to be $1/C$, where C is the number of copies. Requiring that this form of the phase-space density satisfies Hamilton's equations of motion for ensemble averages of the individual particles and momenta derived from the Liouville equation[33, 34] gives analogous differential equations describing motions of the bath atoms and the copied system atoms:

$$\dot{q}_{i,k}=\frac{\partial H_k}{\partial p_{i,k}}, \qquad (2.4a)$$

$$\dot{p}_{i,k}=-\frac{\partial H_k}{\partial q_{i,k}}, \qquad (2.4b)$$

$$\dot{Q}_i = \sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial P_i}, \qquad (2.4c)$$

$$\dot{P}_i = -\sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial Q_i}. \qquad (2.4d)$$

Here, lower case variables refer to the copied subsystem and uppercase variables refer to the "bath." The *i* refers to the coordinate index, while *k* indicates the copy. The chosen form for the phase-space density is never rigorously justified, but rather is rationalized by analogy to the time-dependent Hartree (TDH) approximation for quantum mechanical wave functions. This analogy was previously used to obtain a similar, rigorous method intended for small molecules,[35] after taking the quantum expression to the classical limit. The original method is rigorous because the entire small system is replicated and "trajectory bundles" are simulated, whereas LES replicates only a small part of the system. Although the quantum mechanical analogy is very clever and insightful, it provides limited prospects for improving approximate multiple-copy simulation methods and has left many workers wondering exactly how the approximation alters the dynamics.

To understand and resolve the limitations and uncertainties inherent in LES, subsequent workers have tried to put multiple-copy methods on a more stable foundation. Most notably, Zheng and Zheng[28] claim a derivation of LES by starting with copies of the entire system. Then, they perform a unitary transformation of coordinates for particles of the bath and integrate out of the phase-space density all of the transformed variables except for those corresponding to the mean of the bath's coordinates. The transformation employed was of the form

$$Q_i = \frac{1}{n}\left(\sum_{k=1}^{C} Q_{i,k}\right), \qquad (2.5a)$$

$$P_i = \frac{1}{n}\left(\sum_{k=1}^{C} P_{i,k}\right), \qquad (2.5b)$$

$$Q'_{i,l} = \frac{1}{n}\left(Q_{i,1} + \sum_{k=2}^{C} c_{l,k} Q_{i,k}\right), \qquad (2.5c)$$

$$P'_{i,l} = \frac{1}{n}\left(P_{i,1} + \sum_{k=2}^{C} c_{l,k} P_{i,k}\right), \qquad (2.5d)$$

where we refer to $\{Q_i\}_{i=1}^{`N}$ and $\{P_i\}_{i=1}^{`N}$ as the ‘‘major ’’ variables, $\{Q'_{i,l}\}_{i=1,l=2}^{`N,C}$ and $\{P'_{i,l}\}_{i=1,l=2}^{`N,C}$ as the “minor” variables, and $\{c_{i,j}\}$ are the transform coefficients. The major variables correspond to the average position of the bath coordinates for all of the copies, and the minor variables can be viewed as an orthogonal set of vectors that describe the fine details of the dynamics. There are, of course, constraints that define the transformation. Not only must the coefficients generate an orthogonal transformation, it is also convenient to normalize them such that

$$1 + \sum_{k=2}^{C} c_{n,k} c_{m,k} = C\delta_{m,n}$$

with the additional constraint that

$$1 + \sum_{k=2}^{C} c_{n,k} = 0.$$

In their work, Zheng and Zheng used a probability density appropriate for the canonical ensemble and only considered the special case of a harmonic potential energy function. After transforming and integrating out the minor variables, their result seemed to imply that by choosing the value of *n* used in the transformation

above, the temperature disparity problem could be solved. The choice of $n = \sqrt{C}$ makes the transformation canonical, which when applied to the arguments in references 26 and 27, the authors felt, would correct the equipartition problem for the harmonic potential considered.

To investigate the general temperature disparity problem, an originator of LES showed that LES conserves the following Hamiltonian:[27]

$$H_{LES} = \sum_{i=1}^{N} \frac{P_i^2}{2m_i} + \sum_{i>j=1}^{N} V\left(Q_i, Q_j\right) + \frac{1}{C}\sum_{k=1}^{C}\left(\sum_{i=1}^{S} \frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S} V\left(q_{i,k}, q_{j,k}\right) + \sum_{i=1}^{S}\sum_{j=1}^{N} V\left(q_{i,k}, Q_j\right)\right), \tag{2.6}$$

where S is the number of copied particles, C is the number of times the particles are copied, and N is the number of uncopied "bath" particles. Lowercase variables refer to the copied particles, while uppercase variables indicate the "bath." Dynamical information is obtained by integrating the equations

$$\dot{q}_{i,k} = C\frac{\partial H_{LES}}{\partial p_{i,k}}, \tag{2.7a}$$

$$\dot{p}_{i,k} = -C\frac{\partial H_{LES}}{\partial q_{i,k}}, \tag{2.7b}$$

$$\dot{Q}_i = \frac{\partial H_{LES}}{\partial P_i}, \tag{2.7c}$$

$$\dot{P}_i = -\frac{\partial H_{LES}}{\partial Q_i}. \tag{2.7d}$$

This is an approximation to the dynamics of C systems.

Straub and Karplus, who originally introduced the Lagrangian generator of this Hamiltonian, propose energy scaling as a solution for the temperature disparity

problem.[26] Ulitsky and Elber derived the Hamiltonian given above[27] and noted that the TDH approximation does not follow the classical virial theorem and therefore violates energy equipartitioning, and claimed that this fault is the source of some of the more apparent limitations of LES. To remedy this, Ulitsky and Elber appealed to the Boltzmann equation of transport theory and created an algorithm that monitored "collisions" between atoms of copied particles and the bath, in analogy to the collision integral. Then, they modified their dynamics to treat the collision exactly, instead of using the LES approximation. With this "collisional" LES (cLES) scheme, they achieved improved behavior in the dynamics of systems studied.[27, 31]

2.3 RIGOROUS ALGORITHM FOR MULTIPLE-COPY DYNAMICS

Inspired by the insightful approach of Zheng and Zheng,[28] we decided to take an exact expression for an ensemble of C independent copies of a system, and transform it to obtain an exact multiple-copy dynamics for systems with a more general potential energy function than the harmonic form considered by Zheng and Zheng. Using a transform similar to the one given above makes this possible. The key is to realize that the reverse transform has the form

$$Q_{i,1} = \frac{C}{n}\left(Q_i + \sum_{k=2}^{C} Q'_{i,k}\right), \qquad (2.8a)$$

$$P_{i,1} = \frac{C}{n}\left(P_i + \sum_{k=2}^{C} P'_{i,k}\right), \qquad (2.8b)$$

$$Q_{i,k} = \frac{C}{n}\left(Q_i + \sum_{l=2}^{C} c_{l,k} Q'_{i,l}\right), \qquad (2.8c)$$

$$P_{i,k} = \frac{C}{n}\left(P_i + \sum_{l=2}^{C} c_{l,k} P'_{i,l}\right), \qquad (2.8d)$$

which can then be applied to the exact Hamiltonian of C noninteracting copies of a system. This can be done by first writing the Hamiltonian of a collection of C equivalent copies of the same system assuming a two-body form for the potential energy:

$$H = \sum_{k=1}^{C}\left(\sum_{i=1}^{S}\frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S}V\left(q_{i,k},q_{j,k}\right) + \sum_{i=1}^{S}\sum_{j=1}^{N}V\left(q_{i,k},Q_{j,k}\right) + \sum_{i=1}^{N}\frac{P_{i,k}^2}{2M_i} + \sum_{i>j=1}^{N}V\left(Q_{i,k},Q_{j,k}\right)\right) \qquad (2.9)$$

Here the first two terms are the energy due to the copied subsystem, the middle term is the interaction energy between the copies and the bath, and the final two terms are the energy due to the bath. Applying the transform yields

$$\tilde{H} = \sum_{i=1}^{S}\sum_{k=1}^{C}\frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S}\sum_{k=1}^{C}\left(q_{i,k},q_{j,k}\right) + \sum_{j=1}^{N}\frac{n^2P_j^2}{2M_jC} + \sum_{i>j=1}^{N}\sum_{k=1}^{C}V\left(\frac{C}{n}\left(Q_i+\sum_{k=2}^{C}Q'_{i,k}\right),\frac{C}{n}\left(Q_j+\sum_{k=2}^{C}Q'_{j,k}\right)\right) + \sum_{j=1}^{C}\sum_{k=2}^{C}\frac{n^2P'^2_{j,k}}{2M_jC} + \sum_{i=1}^{N}\sum_{j=1}^{S}\sum_{k=1}^{C}V\left(q_{j,k},\frac{C}{n}\left(Q_i+\sum_{k=2}^{C}Q'_{i,k}\right)\right). \qquad (2.10)$$

Here, again, the first two terms are the energy of the copied particles. The transform has split the bath energy into three terms in Eq. 2.10: the kinetic energy of the major variables, the potential energy due to the bath, and the kinetic energy due to the minor variables. The last term in Eq. 2.10 is the interaction energy between the copies and the bath.

Obviously, the original Hamiltonian can be used to generate equations of motion, which can be integrated to generate the desired trajectories. The transformed Hamiltonian's equations of motions, though, are more interesting. The untransformed, "copied" part is just the same as in the untransformed representation, but the "bath" part is more complex. For example, the time derivative,

$$\dot{Q}_i = \frac{1}{n}\left(\sum_{k=1}^{C} \dot{Q}_{i,k}\right),$$

can be rewritten as

$$\dot{Q}_i = \frac{1}{n}\sum_{k=1}^{C}\left(\frac{\partial \tilde{H}}{\partial P_i}\frac{\partial P_i}{\partial P_{i,k}}\right),$$

which can finally be reduced to

$$\dot{Q}_i = \frac{C}{n^2}\frac{\partial \tilde{H}}{\partial P_i}. \qquad (2.11)$$

Given all of this, the entire set of equations of motion can be rewritten as

$$\dot{q}_{i,k} = \frac{\partial \tilde{H}}{\partial p_{i,k}}, \qquad (2.12a)$$

$$\dot{p}_{i,k} = -\frac{\partial \tilde{H}}{\partial q_{i,k}}, \qquad (2.12b)$$

$$\dot{Q}_i = \frac{C}{n^2}\frac{\partial \tilde{H}}{\partial P_i}, \qquad (2.12c)$$

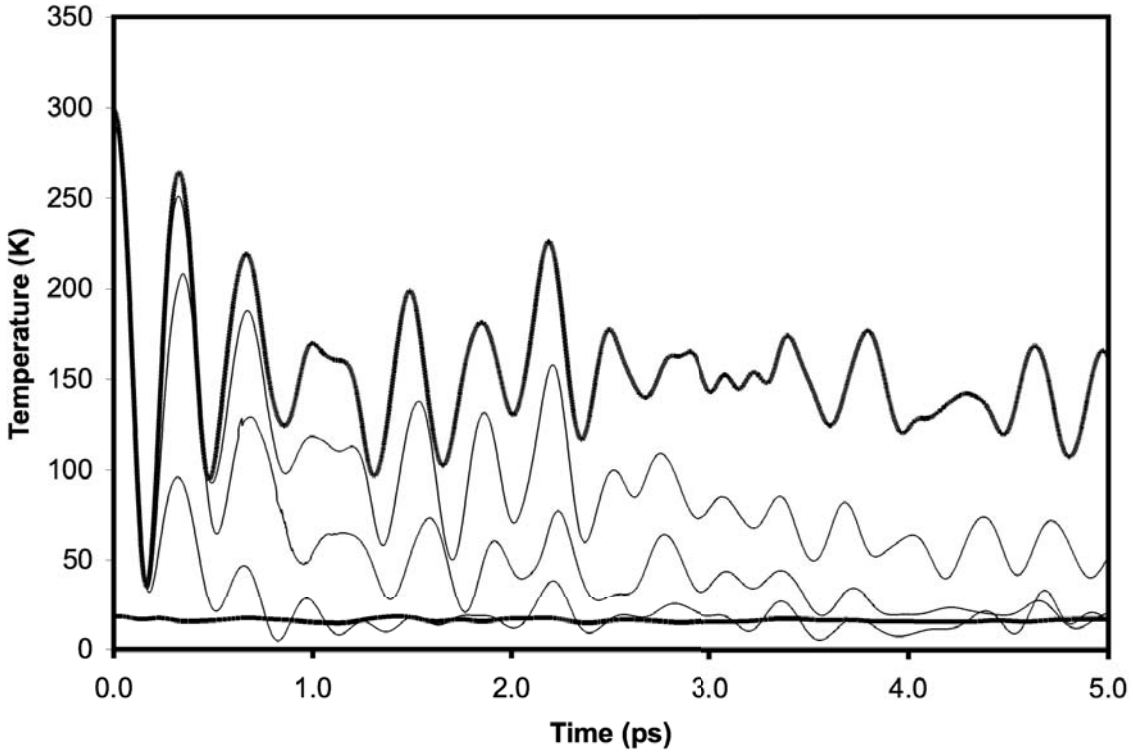
$$\dot{P}_i = -\frac{C}{n^2}\frac{\partial \tilde{H}}{\partial Q_i}, \qquad (2.12d)$$

$$\dot{Q}'_{i,k} = \frac{C}{n^2}\frac{\partial \tilde{H}}{\partial P'_{i,k}}, \qquad (2.12e)$$

$$\dot{P}'_{i,k} = -\frac{C}{n^2}\frac{\partial \tilde{H}}{\partial Q'_{i,k}}. \qquad (2.12f)$$

These equations are the exact equations of motion in the transformed representation. It should be noted that if $n = \sqrt{C}$, we have a canonical transformation, as Zheng and Zheng accomplished for harmonic potentials. But if $n = C$, the point transform generated yields equations of motion that are very similar to the LES equations. Now, if all of the bath particles start from the same initial positions with the same velocities in all of the copies and all of the minor variables are ignored (assumed to vanish), the LES equations of motion are recovered. We claim that all of the error present in this sort of multiple-copy method comes from this holonomic constraint and that previously noted faults flow from this idea. Numerical simulations presented in Fig. 2.1 support this view for the temperature disparity problem. The flattest curve in Fig. 2.1 shows the temperature of an argon bath as a function of time and the curve that oscillates around it is the temperature of a single argon atom calculated using exact dynamics. The top, bold curve is the temperature of a single argon atom, represented by four copies, in the LES approximation. Clearly, the LES particle's temperature does not relax to equilibrium as it should, and its fluctuations are much larger than in exact dynamics. Our algorithm can be used to generate each curve in Fig. 2.1, including those intermediate between LES and exact dynamics. These curves were generated using a scheme practically similar to cLES, but with a fundamental difference in implementation and basis. In cLES, "collisions" were detected by physical proximity between a copied particle and a bath particle, whereas the correction presented here calculates the "minor" variables' forces (the $\dot{P}'$'s from above), which are neglected in both LES and cLES. Our method still ignores these forces when they are sufficiently small, but when the force exerted on the minor

Figure 2.1. The flattest curve is the temperature as a function of time for a bath of 64 Ar atoms calculated using molecular dynamics. The curve that oscillates around it is the temperature of a single Ar atom relaxing to equilibrium calculated using exact molecular dynamics. The bold curve is the temperature of one Ar calculated using the LES approximation. The curves between the exact and LES curves were calculated using the algorithm presented.



variable is greater than some threshold, the particle is then removed from the bath and treated exactly, as if a copied particle. By adjusting the threshold force, the dynamics can be scaled from purely LES to exact dynamics, in principle visiting all points in between. This allows the computational advantage of LES to be largely maintained, while providing a rigorous procedure that leads to improved dynamics.

2.4 COMPARISONS WITH CONVENTIONAL MULTIPLE-COPY DYNAMICS

As most of the work produced to explain conventional multiple-copy dynamics depends on a phase-space density approach[1, 6, 26-28] often involving the Liouville equation, further comparisons of our work with previously published methods would be helped by casting our work in such a formalism. First, we consider a formal representation of the density that satisfies the Liouville equation,[34]

$$\rho\left(\vec{\mathbf{X}},t\right) = e^{-\hat{\mathbf{L}}t}\rho\left(\vec{\mathbf{X}},0\right) \tag{2.13}$$

where $\hat{\mathbf{L}}$ is a Liouville operator,

$$\begin{aligned}\hat{\mathbf{L}} = &\sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}} - \frac{\partial H}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right)\\ &+\sum_{i=1}^{N}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial P_{i,k}}\frac{\partial}{\partial Q_{i,k}} - \frac{\partial H}{\partial Q_{i,k}}\frac{\partial}{\partial P_{i,k}}\right)\end{aligned} \tag{2.14}$$

and $\rho\left(\vec{\mathbf{X}},0\right)$ is the density at time zero:

$$\begin{aligned}\rho\left(\vec{\mathbf{X}},0\right) = &\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(q_{i,k}-q_{i,k,0}\right)\delta\left(p_{i,k}-p_{i,k,0}\right)\right)\\ &\times\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(Q_{i,k}-Q_{i,k,0}\right)\delta\left(P_{i,k}-P_{i,k,0}\right)\right).\end{aligned} \tag{2.15}$$

Next, we apply the same point coordinate transformation as before to both the Liouville operator and the initial probability density. This generates the transformed Liouvillian

$$\hat{\mathbf{L}} = \sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial \tilde{H}}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}} - \frac{\partial \tilde{H}}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right)$$

$$+\frac{C}{n^2}\left[\sum_{i=1}^{N}\left(\frac{\partial\tilde{H}}{\partial P_i}\frac{\partial}{\partial Q_i}-\frac{\partial\tilde{H}}{\partial Q_i}\frac{\partial}{\partial P_i}\right)+\sum_{i=1}^{N}\sum_{k=2}^{C}\left(\frac{\partial\tilde{H}}{\partial P'_{i,k}}\frac{\partial}{\partial Q'_{i,k}}-\frac{\partial\tilde{H}}{\partial Q'_{i,k}}\frac{\partial}{\partial P'_{i,k}}\right)\right]. \qquad (2.16)$$

After applying the Jacobian of the transformation, $\mathrm{J}\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)$, and assuming that the bath's initial positions and velocities are the same in all of the copies, the transform also generates the density*,

$$\rho\left(\vec{\mathbf{X}},0\right)=\frac{1}{\mathrm{J}\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)}\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(q_{i,k}-q_{i,k,0}\right)\delta\left(p_{i,k}-p_{i,k,0}\right)\right)$$
$$\times\left(\prod_{i=1}^{N}\delta\left(Q_i-Q_{i,0}\right)\delta\left(P_i-P_{i,0}\right)\prod_{k=2}^{C}\delta\left(Q'_{i,k}\right)\delta\left(P'_{i,k}\right)\right). \qquad (2.17)$$

Thus the exact time-dependent phase-space density in the transformed representation, $\tilde{\rho}\left(\tilde{\vec{\mathbf{X}}},t\right)$, is formed by using the transformed Liouvillian and transformed initial densities in the above formal expression. One can then follow the example provided in Zheng and Zheng's work to arrive at a reduced density, $\tilde{\rho}'\left(\tilde{\vec{\mathbf{X}}},t\right)$, that can be used to generate LES- type equations of motion. This is simply done by integrating over the minor variables

$$\tilde{\rho}'\left(\tilde{\vec{\mathbf{X}}},t\right)=\int d\gamma'\left|\mathrm{J}\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)\right|e^{\hat{\mathbf{L}}t}\tilde{\rho}\left(\tilde{\vec{\mathbf{X}}},0\right), \qquad (2.18)$$

where $d\gamma'$ is the volume element in the space that the minor variables occupy. Not only does this integration directly generate a density that can be used to obtain the LES equations of motion, but it also serves to enforce the holonomic constraints that were found above to result in LES.

---

* This was obtained via the identity $\delta(x-x')\delta(y-y')=\frac{1}{|J|}\delta(\alpha-\alpha')\delta(\beta-\beta')$, c.f. in CRC Standard Mathematical Tables and Formulae, 30th ed. (CRC Press, Boca Raton, FL, 1996).

The approach presented to generate a multiple-copy method also sheds some light onto the violation of energy equipartitioning[26] or equivalently, the incorrect virial given by LES.[27] One of the well known results of classical mechanics is that the time or assuming ergodicity, the ensemble average of the kinetic energy can be expressed in the following way:[36]

$$\overline{T} = -\tfrac{1}{2}\sum_i \langle q_i F_i \rangle, \tilde{\rho}'\left(\tilde{\vec{\mathbf{X}}}, t\right) = \int d\gamma' \left|\mathrm{J}\left(\vec{\mathbf{X}}, \tilde{\vec{\mathbf{X}}}\right)\right| e^{\hat{\mathbf{L}}t} \tilde{\rho}\left(\tilde{\vec{\mathbf{X}}}, 0\right), \qquad (2.19)$$

where $\langle A \rangle$ is the ensemble average of quantity *A*, and the summation is taken over all the degrees of freedom in the system. In the canonical ensemble, this produces the familiar result that $\overline{T} = {}^{N}\!/_{2\beta}$, where *N* is the number of degrees of freedom in the system, β is ${}^{1}\!/_{kT}$, and it is assumed that the potential diverges on the surface of the volume of integration. This assumption corresponds physically to considering a bound system. In this special case of a bound system, Ulitsky and Elber showed that the virial theorem generates a strange result, $\overline{T} = {}^{(N+CS)}\!/_{2\beta}$, when applied.[27] From this expression, it appears that the temperature of the copied subsystem is *C* times hotter than the bath's. Our approach confirms this result, but demonstrates that this too can be traced to a neglect of the minor variables.

In the canonical ensemble, evaluating $\langle q_i F_i \rangle$ for a bound system entails solving the integral

$$\langle q_i F_i \rangle = -\frac{1}{Z}\int d\Gamma dq_i q_i \frac{\partial H}{\partial q_i} e^{-\beta \mathrm{H}(q_i, \Gamma)}. \qquad (2.20)$$

Here, *Z* is the partition function and q_i can be any of the coordinates described in this paper, be it major, minor, or untransformed. The integral is evaluated most easily by

integrating by parts and its exact value is $-1/\beta$. Therefore, the virial theorem generates an average kinetic energy for an ensemble of C systems as simply

$$\overline{T} = -\tfrac{1}{2}\sum_{i=1}^{S}\sum_{k=1}^{C}\left\langle q_{i,k}\frac{\partial H}{\partial q_{i,k}}\right\rangle - \tfrac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{C}\left\langle Q_{i,k}\frac{\partial H}{\partial Q_{i,k}}\right\rangle$$
$$= \frac{(SC+NC)}{2\beta}$$

After making the transformation,

$$\left\langle Q_{i,k}\frac{\partial H}{\partial Q_{i,k}}\right\rangle = \left\langle \frac{C^2}{n^3}\left(Q_i + \sum_{i=2}^{C} c_{l,k}Q'_{i,l}\right)\left(\frac{\partial H}{\partial Q_i} + \sum_{m=2}^{C} c_{m,k}\frac{\partial H}{\partial Q'_{i,m}}\right)\right\rangle,$$

and using the result in the virial theorem, the exact result, $\overline{T} = (NC+CS)/2\beta$, is generated. However, if the minor variables are ignored, one obtains the result $\overline{T} = (N+CS)/2\beta$.

So, the problem with the LES virial is simply a counting problem and the average kinetic energy takes the Ulitsky-Elber form because of the reduced number of degrees of freedom found in LES.

2.5 CONCLUSIONS

Starting from an ensemble of identical systems and applying a point transformation to the coordinates of a large number of "bath" particles generates an algorithm for efficiently replicating the dynamics of the ensemble. The transformation gives a description of the bath in terms of "major" variables located at the average phase-space position of equivalent atoms and a set of "minor" variables describing the finer details of the bath dynamics. Numerical tests show that the algorithm can recover exact dynamics or give dynamics identical to conventional multiple-copy dynamics,[1] if the minor variables are neglected. If the minor variables are included in the dynamics only when the corresponding forces exceed a chosen

threshold, results of intermediate accuracy are obtained. Varying the threshold force controls the accuracy of the calculation and the computer time required.

Applying the point transformation to an ensemble of identical systems also opens a new perspective on conventional multiple-copy dynamics. First, neglecting Hamilton's equations of motion for the minor variables of the bath gives the equations of motion that define LES.[1] Second, applying the same point transformation to the Liouville equation and probability density, followed by enforcing the holonomic constraint that the minor variables vanish, gives a probability density similar to that assumed for LES. Third, neglecting the minor variables gives the same incorrect virial as LES, but including the minor variables yields the correct virial for the ensemble. Finally, the success of collisional LES is understandable, as cLES includes the minor variable dynamics in an empirical way.

Since mean field theories such as LES have proven extremely useful for locating global minima on complex free energy surfaces,[1-16] work is currently underway to develop the algorithm described here into a similar tool. In addition, we are investigating the limitations of the algorithm for calculating ensemble average properties accurately and quickly.

2.6 REFERENCES

1. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

2. Finkelstein, A. V.; Reva, B. A., A search for the most stable folds of protein chains. *Nature (London, United Kingdom)* **1991,** 351, (6326), 497-9.

3. Hofacker, I.; Schulten, K., Oxygen and proton pathways in cytochrome c oxidase. *Proteins: Structure, Function, and Genetics* **1998,** 30, (1), 100-107.

4. Koehl, P.; Delarue, M., Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *Journal of Molecular Biology* **1994,** 239, (2), 249-75.

5. Koehl, P.; Delarue, M., A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modelling. *Nature Structural Biology* **1995,** 2, (2), 163-70.

6. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

7. Kono, H.; Doi, J., Energy Minimization Method Using Automata Network for Sequence and Side-Chain Conformation Prediction from Given Backbone Geometry. *Proteins-Structure Function and Genetics* **1994,** 19, (3), 244-255.

8. Lee, C., Predicting Protein Mutant Energetics by Self-consistent Ensemble Optimization. *Journal of Molecular Biology* **1994,** 236, (3), 918-939.

9. Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N., Jr.; Dou, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; et al., Structural and functional effects of apolar mutations of the distal valine in myoglobin. *Journal of Molecular Biology* **1995,** 245, (4), 416-36.

10. Reva, B. A.; Finkelstein, A. V., A New Approach to the Design of a Sequence with the Highest Affinity for a Molecular-Surface. *Protein Engineering* **1992,** 5, (7), 625-628.

11. Roitberg, A.; Elber, R., Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* **1991,** 95, (12), 9277-87.

12. Simmerling, C.; Fox, T.; Kollman, P. A., Use of locally enhanced sampling in free energy calculations: Testing and application to the alpha ->beta anomerization of glucose. *Journal of the American Chemical Society* **1998,** 120, (23), 5771-5782.

13. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

14. Vasquez, M., An evaluation of discrete and continuum search techniques for conformational analysis of side chains in proteins. *Biopolymers* **1995,** 36, 53-70.

15. Zheng, Q.; Kyle, D. J., Multiple copy sampling: rigid versus flexible protein. *Proteins* **1994,** 19, (4), 324-9.

16. Zheng, Q.; Rosenfeld, R.; DeLisi, C.; Kyle, D. J., Multiple copy sampling in protein loop modeling: computational efficiency and sensitivity to dihedral angle perturbations. *Protein Science* **1994,** 3, (3), 493-506.

17. Huber, G. A.; McCammon, J. A., Weighted-ensemble simulated annealing: Faster optimization on hierarchical energy surfaces. *Physical Review E* **1997,** 55, (4), 4822-4825.

18. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Optimization methods for conformational sampling using a Boltzmann-weighted mean field approach. *Biopolymers* **1996,** 39, (1), 103-114.

19. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics. *Journal of Physical Chemistry A* **1997,** 101, (33), 5926-5930.

20. Huber, T.; van Gunsteren, W. F., SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A* **1998,** 102, (29), 5937-5943.

21. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

22. Ma, J. P.; Hsu, D.; Straub, J. E., Approximate Solution of the Classical Liouville Equation Using Gaussian Phase Packet Dynamics - Application to Enhanced Equilibrium Averaging and Global Optimization. *Journal of Chemical Physics* **1993,** 99, (5), 4024-4035.

23. Maranas, C. D.; Floudas, C. A., A Deterministic Global Optimization Approach for Molecular-Structure Determination. *Journal of Chemical Physics* **1994,** 100, (2), 1247-1261.

24. Piela, L.; Kostrowicki, J.; Scheraga, H. A., The Multiple-Minima Problem in the Conformational-Analysis of Molecules - Deformation of the Potential-Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry* **1989,** 93, (8), 3339-3346.

25. Tsallis, C.; Stariolo, D. A., Generalized simulated annealing. *Physica A* **1996,** 233, (1-2), 395-406.

26. Straub, J. E.; Karplus, M., Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation. *Journal of Chemical Physics* **1991,** 94, (10), 6737-6739.

27. Ulitsky, A.; Elber, R., The Thermal-Equilibrium Aspects of the Time-Dependent Hartree and the Locally Enhanced Sampling Approximations - Formal Properties, a Correction, and Computational Examples for Rare-Gas Clusters. *Journal of Chemical Physics* **1993,** 98, (4), 3380-3388.

28. Zheng, W. M.; Zheng, Q., An analytical derivation of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1997,** 106, (3), 1191-1194.

29. Stultz, C. M.; Karplus, M., On the potential surface of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1998,** 109, (20), 8809-8815.

30. Simmerling, C.; Miller, J. L.; Kollman, P. A., Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid. *Journal of the American Chemical Society* **1998,** 120, (29), 7149-7155.

31. Ulitsky, A.; Elber, R., Application of the Locally Enhanced Sampling (Les) and a Mean-Field with a Binary Collision Correction (Cles) to the Simulation of Ar Diffusion and No Recombination in Myoglobin. *Journal of Physical Chemistry* **1994,** 98, (3), 1034-1043.

32. Verkhivker, G.; Elber, R.; Nowak, W., Locally Enhanced Sampling in Free-Energy Calculations - Application of Mean Field Approximation to Accurate Calculation of Free-Energy Differences. *Journal of Chemical Physics* **1992,** 97, (10), 7838-7841.

33. Liboff, R. L., *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*. J. Wiley: New York, 1998.

34. Prigogine, I., *Non-equilibrium statistical mechanics*. Interscience: New York, 1962.

35. Gerber, R. B.; Buch, V.; Ratner, M. A., Time-Dependent Self-Consistent Field Approximation for Intramolecular Energy-Transfer .1. Formulation and Application to Dissociation of Vanderwaals Molecules. *Journal of Chemical Physics* **1982,** 77, (6), 3022-3030.

36. Goldstein, H., *Classical Mechanics*. Addison-Wesley: Reading, MA, 1950.

CHAPTER 3

Practical multiple-copy methods for sampling classical statistical mechanical ensembles

3.1 INTRODUCTION

Molecular dynamics (MD) is a technique that has found widespread application to problems ranging from the purely classical to the purely quantum mechanical,[1, 2] from finding minima on potential energy surfaces to approximating integrals required to find statistical mechanical phase-space averages.[2-4] Despite its widespread use, MD easily allows a system to become trapped in shallow, local wells in its potential energy surface. Diverse methods have been realized to correct this.[5-13] Perhaps the most accepted, and certainly most widely used of these techniques is simulated annealing.[8, 13, 14]

Another class of techniques used to more efficiently explore the potential energy surface are the mean field methods.[9] Since their early use,[15, 16] mean field methods have been used to generate approximate classical trajectories. These trajectories have been used for a wide variety of purposes, but have generally been employed in four areas that require large amounts of computational effort: finding global energy minima of complex systems,[17-19] studying non-equilibrium behavior such as ligand diffusion,[20, 21] increasing sampling during the calculation of a free energy difference,[18] and searching for molecules that bind to an active site.[22-24] The benefits of using mean field methods, and the locally-enhanced sampling (LES) method in

particular, are generally thought to overshadow any uncertainties[25-29] caused by such an approximation.

LES, as it was originally described,[15] allows a small part of the system of interest to be copied several times. Each copy feels the same force that the corresponding real particle would feel, while the rest of the system (the uncopied part, or "bath") feels the average of the forces contributed by the copied atoms (the mean field). Thus, the force on the bath particles is much the same as it would be in a conventional simulation, differing only due to the average nature of the interactions with the copies. This interaction force is calculated with a relatively small effort. Of course, this rather severe approximation is not particularly accurate in non-equilibrium situations or in finding global minima. Various workers have noted that copied particles' non-equilibrium behavior in the LES approximation wanders far from expectations.[26-29]

In this contribution we develop our previously derived algorithm[25] into a new computational method, called Ensembles EXtracted from Atomic Coordinate Transformations (EXACT). The EXACT approximation is rigorously justified by our previous work[25] and generates trajectories for the evaluation of approximate ensemble averages, but should not be confused with exact (conventional) molecular dynamics (MD). The level of accuracy in the EXACT approximation can be scaled by a user-defined parameter from a conventional mean field technique, LES, to the conventional molecular dynamics of an ensemble of particles. We therefore briefly review a key result of our purely classical mechanical approach in the next section. Then we derive the equations of motion that generalize our algorithm from

approximating the microcanonical ensemble to approximating the canonical ensemble, using the Nosé-Hoover Chain method.[30] Then, we describe the computational details of the EXACT approximation. Finally, using tests involving Lennard-Jones clusters, we illustrate some advantages our method provides over conventional multiple-copy methods and conventional MD.

3.2 THEORETICAL BACKGROUND

It is amazing that the same result, i.e. the LES equations of motion, can be obtained from three widely different starting points. In the original presentation, the result was obtained by taking the classical limit of a quantum-mechanical, time-dependent Hartree self-consistent-field procedure.[16] Elber and Karplus also obtained the LES equations of motion by assuming a particular form for the classical phase-space density distribution and then integrating this distribution using the Liouville equation.[15] Finally, we obtained[25] them by replicating the entire system and then applying a point transformation to the "bath" coordinates. Our approach generates a transformed Hamiltonian in the untransformed "copied" coordinates, a set of "major" bath coordinates, and a set of "minor" bath coordinates. The "major" coordinates correspond to the coordinates of the bath particles in the LES approximation, and the "minor" set of coordinates are simply ignored in LES.

The point transformation required to obtain our result is defined previously[25, 29] and is used to write the system's Hamiltonian in a new representation:

$$\tilde{H} = \sum_{i=1}^{S}\sum_{k=1}^{C}\frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S}\sum_{k=1}^{C}\left(q_{i,k},q_{j,k}\right) + \sum_{j=1}^{N}\frac{CP_j^2}{2M_j}$$
$$+ \sum_{i>j=1}^{N}\sum_{k=1}^{C}V\left(\left(Q_i + \sum_{k=2}^{C}Q'_{i,k}\right),\left(Q_j + \sum_{k=2}^{C}Q'_{j,k}\right)\right)$$
$$+ \sum_{j=1}^{C}\sum_{k=2}^{C}\frac{CP'^2_{j,k}}{2M_j} + \sum_{i=1}^{N}\sum_{j=1}^{S}\sum_{k=1}^{C}V\left(q_{j,k},\left(Q_i + \sum_{k=2}^{C}Q'_{i,k}\right)\right). \qquad (3.1)$$

The coefficients of this transformation are needed in the implementation of the EXACT approximation, and details for their calculation are provided in supplementary material. The first two terms refer to the copied particles (lowercase variables). The next three terms in Eq. (3.1) are the kinetic energy of the major variables (capitalized unprimed variables), the potential energy due to the bath, and the kinetic energy due to the minor variables (capitalized unprimed variables). The last term in Eq. (3.1) represents the potential energy of interaction between the copies and the bath. This result demonstrates that if all of the bath particles start from the same initial positions and velocities in all of the copies and all of the minor variables' coordinates are holonomically constrained to vanish, the LES Hamiltonian is recovered.

3.3 EXTENSION TO THE CONSTANT TEMPERATURE ENSEMBLE

To extend this algorithm to constant temperature ensembles, we use the Nosé-Hoover Chain method.[30] The Nose-Hoover Chain (NHC) method for controlling temperature in molecular dynamics works by appending to the original Hamiltonian an extended system designed to act as a heat bath. This is done, qualitatively, in analogy to the procedure used in the derivation of the canonical ensemble's probability density expression. The heat bath is composed of "chains" of particles that

are linearly coupled to the original system and each other. One end of the chain interacts with the system and the second member of the chain, the second member of the chain interacts with the first and the third members, until the other end of the chain is reached. Thus, the following Hamiltonian is conserved for NHC chain length of 2 and C copies of a system of interest:

$$H_{NHC} = H + \sum_{k=1}^{C}\left(\sum_{i=1}^{2}\frac{p_{\eta_i,k}^2}{2\mu_i} + k_B T\left[D\eta_{1,k} + \eta_{2,k}\right]\right). \qquad (3.2)$$

In this equation, H is just the Hamiltonian of the system, $p_{\eta_i,k}$ and $\eta_{i,k}$ are the momenta and "position" of the heat bath's chain particles, and μ_i is the "mass" of the chain particle. The subscripts on $p_{\eta_i,k}$ indicate chain particle *i* in copy *k*. The temperature of the system is denoted T, k_B is Boltzmann's constant, and D is the number of particles to be copied. The exact "non-Hamiltonian"[2, 30, 31] equations of motion then can be written as

$$\dot{q}_{i,k} = \frac{\partial H_{NHC}}{\partial p_{i,k}},\ \dot{p}_{i,k} = -\frac{\partial H_{NHC}}{\partial q_{i,k}} - p_{i,k}\frac{\partial H_{NHC}}{\partial p_{\eta_i,k}},$$
$$\dot{\eta}_{i,k} = \frac{\partial H_{NHC}}{\partial p_{\eta_i,k}}, \qquad \dot{p}_{\eta_1,k} = -\frac{\partial H_{NHC}}{\partial \eta_{1,k}} - 2K_k - 2p_{\eta_1,k}\frac{\partial H_{NHC}}{\partial p_{\eta_2,k}},$$
$$\dot{p}_{\eta_2,k} = -\frac{\partial H_{NHC}}{\partial \eta_{2,k}} + \frac{p_{\eta_1,k}}{\mu_1}, \qquad \dot{Q}_{i,k} = \frac{\partial H_{NHC}}{\partial P_{i,k}},$$
$$\dot{P}_{i,k} = -\frac{\partial H_{NHC}}{\partial Q_{i,k}} - P_{i,k}\frac{\partial H_{NHC}}{\partial p_{\eta_i,k}}.$$

Here, K_k is the kinetic energy contained in the *k*th copy. The transform described in Section (3.2) can then be applied to Eq. (3.2) to generate the transformed Hamiltonian and the equations of motion that control the temperature in the EXACT approximation. The transformation leaves the equations of motion for the copied particles and the bath momenta ($\dot{Q}_i$) unchanged from Eq. (3.3). The forces on the

major variables, on the other hand, can be expressed as

$$\dot{P}_i = \frac{1}{C}\sum_{k=1}^{C}\dot{P}_{i,k}\,, \qquad (3.4)$$

which can be rewritten exactly (using the properties of the transformation coefficients[25]) as:

$$\dot{P}_i = -\frac{1}{C}\left(\frac{\partial H_{NHC}}{\partial Q_i} + P_i\sum_{k=1}^{C}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}} + \sum_{l=2}^{C}P'_{i,l}\sum_{k=1}^{C}c_{l,k}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}}\right). \qquad (3.5)$$

Setting the minor variables to zero allows the final term to be ignored and leaves the first two terms as the force required to maintain a constant temperature in the EXACT approximation. Doing a transformation similar to the one used to derive Eq. (3.5) gives the temperature controlled force on the minor variables. Applying the holonomic constraint to eliminate the "minor" variables of the bath yields the final expression:

$$\dot{P}_{i,l} = -\frac{1}{C}\left[\frac{\partial H_{NHC}}{\partial Q'_{i,l}} + P_i\left(\frac{\partial H_{NHC}}{\partial p_{\eta_1,1}} + \sum_{k=2}^{C}c_{l,k}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}}\right)\right]. \qquad (3.6)$$

Equation (3.6) is used in constant temperature implementations of the EXACT approximation.

3.4 COMPUTATIONAL METHODS

The computational goal of this work is to demonstrate the ability of the EXACT approximation to provide a more accurate alternative to LES. All calculations reported in this work were obtained by using our own implementations of conventional MD,[2] LES,[15] and the EXACT approximation. We integrated the equations of motion using the velocity-Verlet[2, 3, 32] algorithm with a time step of 1 fs.

Instantaneous quantities were calculated using information obtained at each time step, while phase space averages (such as the pair distribution function) were obtained by using data collected at every tenth time step.

The system studied was a 64 atom Ar[33] cluster at 20 ± 5 K, in vacuum. This system was selected for its relative simplicity and the fact that it has been studied before[28] in other evaluations of mean field methods. This system was used to test the behavior of EXACT as it approached the conventional MD limit. The Ar cluster was equilibrated for at least 100 ps. Before data collection, in the constant energy formalism the system maintained temperature fluctuations on the order of 5-10 K. Constant temperature simulations using the Nosé-Hoover chain[30] method as described in Section 3 were performed to assess the relaxation properties of a hot Ar atom. The Lennard-Jones parameters ($\varepsilon = 0.238$ kcal mol^{-1} and $\sigma = 3.405$ Å) were obtained from previous work on these Ar clusters.[28]

3.4.1 The "EXACT" approximation Careful inspection of Eq. (3.1) reveals that the LES approximation is good only in regions of phase-space where the minor variables are small and the corresponding pseudoparticles may be ignored. The EXACT approximation accounts for this by using Eq. (3.1) along with an initial holonomic constraint to force the minor pseudoparticles to vanish. Along the way the forces on the minor pseudoparticles are calculated at each step (while still assuming the holonomic constraint) and compared to the force on the corresponding major pseudoparticle by calculating the ratio, $\varphi_{i,l} = \left| \frac{P'_{i,l}}{P_i} \right|$. For any coordinate in which the ratio, $\varphi_{i,l}$, is greater than a user-defined tolerance, Φ, the associated particle is removed from the bath and treated exactly (as a copied particle) for a user-defined

number of time steps, and then returned to the bath in a manner that ensures that the energy is conserved. In general, smaller values of Φ produce more exact results. In the limit, $\Phi \to 0$, conventional dynamics for an ensemble of system copies is produced. Conversely, $\Phi \to \infty$ produces LES. Practically, the value of Φ should be chosen to provide balance between improved accuracy (smaller values) and computational efficiency (larger values).

Actually calculating the force on the minor variables requires little more effort than an LES simulation would. For momenta independent of spatial coordinates we can express the forces as derivatives of the potential energy:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \sum_{k=1}^{C} \left(\frac{\partial V_{B,k}}{\partial Q'_{i,m}} + \frac{\partial V_{X,k}}{\partial Q'_{i,m}} \right) \qquad (3.7)$$

In this representation of the force on the minor pseudoparticle, $V_{B,k}$ is the bath-bath interaction's potential energy in the *k*th copy and $V_{X,k}$ is the bath-copy interaction potential energy due to the *k*th copy. Since the bath's potential energy is the same for each copy due to the holonomic constraints, $V_{B,k}$ is the same in each copy and is replaced by the common value, V_B . Further, Eq. (3.7) can be rewritten:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_B}{\partial Q_i} \left(1 + \sum_{k=2}^{C} c_{m,k} \right) + \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} \qquad , \qquad (3.8)$$

and further simplified (using properties of the transform coefficients) to:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} c_{m,k} \frac{\partial V_{X,k}}{\partial Q_i} . \qquad (3.9)$$

Eq. (3.9) indicates that the force on the minor variables is entirely derived from the force on the major variables, a quantity calculated even in the LES limit.

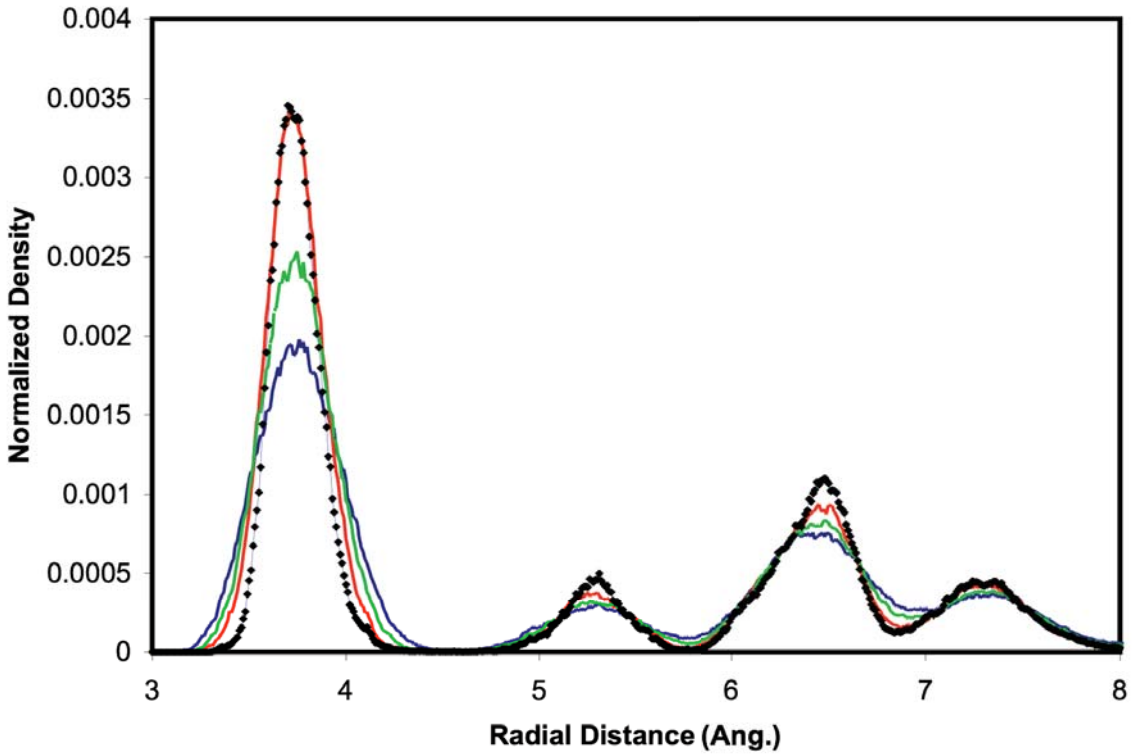
3.5 COMPUTATIONAL TESTS

Tests were performed to compare the results generated by conventional MD, LES, and the EXACT approximation. The tests performed are similar to those used by Ulitsky and Elber[28] to test their cLES method for calculating phase space averages and efficiently returning systems to equilibrium. These tests include finding the time required for a hot (300 K) Ar atom to cool inside a cluster of Ar atoms at 17 K using both constant temperature and constant energy methods, and calculating the pair distribution function of an Ar atom in an Ar cluster at 20 K. The Ulitsky-Elber tests were well chosen, as they comprise a good sample of the techniques required in molecular dynamics including a phase-space average (the distribution function) and equilibration properties (Ar cooling).

3.5.1 Pair distribution function as indicators of enhanced sampling Inspection of the Fig. (3.1) illustrates both a major advantage and disadvantage of LES. The peaks in the pair distribution function generated by LES are located in the correct places, but are clearly too broad and short compared to the MD result. So LES indeed samples regions of space forbidden to particles in a conventional MD simulation. The EXACT approximation, however, allows some control over the extent of sampling. The result with the threshold $\Phi = 0.8$ generates a clearly improved distribution with a taller and narrower initial peak. The initial peak due to the calculation with $\Phi = 0.3$ is nearly indistinguishable from the MD result, differing only by a slight broadening of the tails of the peaks.

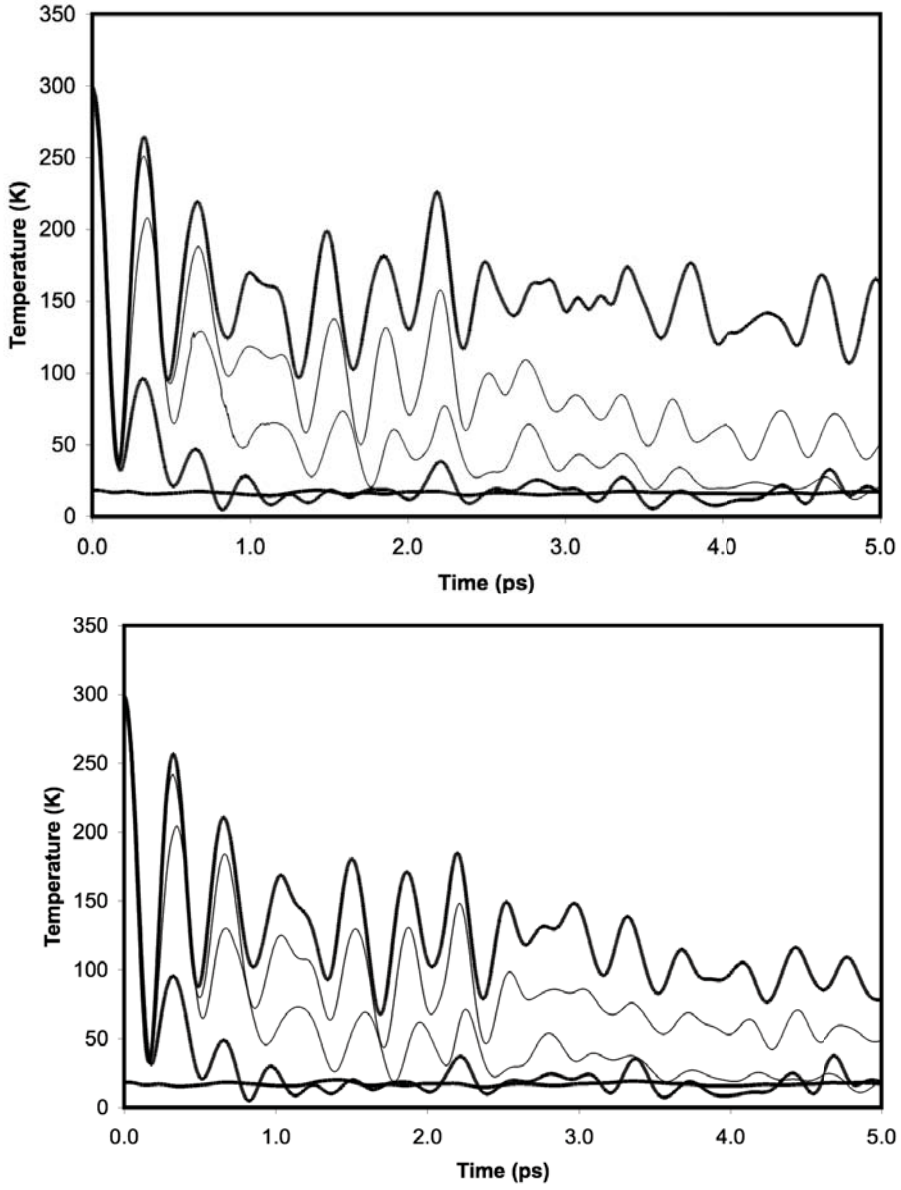
3.5.2 Temperature equilibration times The Ar cooling simulation described above was performed using conventional MD, LES, and the constant energy

Figure 3.1. Normalized density distribution for a single, copied Ar atom in a bath of Ar. Simulations correspond to exact MD (highest, narrowest peak, marked by diamonds), LES(shortest, broadest peak, blue), and intermediate cases, generated using the EXACT approximation (green and red).



implementation of the EXACT algorithm, and the results are shown in Fig. (3.2a). Temperatures were estimated by calculating the instantaneous kinetic energy.[3] This figure illustrates a well-documented flaw in LES, termed the "temperature-disparity problem." This problem manifests itself as the failure of a hot particle to cool properly, and has been explained in several ways.[25, 26, 28, 29] Comparing LES with the conventional MD result demonstrates the spectacular nature of the difficulty. Whereas Ar cools quickly to the bath temperature using conventional MD, in the LES approximation the single particle temperature does not seem to cool at all. The two

Figure 3.2. Comparison of temperature vs. time for the bath of 64 Ar atoms (flattest curve near the bottom) with the temperature of one Ar atom in exact MD(bold curve oscillating about the bath temperature), LES (bold curve at the top showing large oscillations), and intermediate cases, generated using the EXACT approximation for a) (top) constant energy simulations and b) (bottom) constant temperature simulations.



additional curves in the figure illustrate results generated by the EXACT

approximation using thresholds Φ=0.8 and 0.3. The threshold of 0.8 is clearly intermediate between LES and conventional MD, as it cools to about 50K in the 5 ps of simulation time. The result with Φ = 0.3, though, cools to the temperature of the bath within 5 ps and represents a more obvious improvement over LES.

Figure (3.2b) depicts Ar cooling in a constant temperature scheme. The particle treated with LES seems to be slowly approaching equilibrium, however the temperature is still about 100K after 5 ps. It was noted[26] that a possible correction to the "temperature disparity problem" might be to perform a constant temperature simulation. Figure (3.2b) shows that using the Nosé-Hoover chain method does not resolve the issue. The EXACT approximation is a better solution.

3.6 CONCLUSIONS

We have used our rigorous derivation of a multiple-copy mean-field algorithm to generate a new computational method called Ensembles Extracted from Atomic Coordinate Transformations (EXACT), and have extended the method to approximate the canonical ensemble using a transformed Nosé-Hoover formalism. The EXACT approximation works by incorporating the effects of minor variables that are ignored in conventional multiple-copy methods, but which complete the description of conventional molecular dynamics. The minor variables' effects are treated in the EXACT approximation by calculating the force on a corresponding pseudoparticle, and then comparing it to the force the bath particle would feel by calculating their ratio, $\varphi_{i,1}$. This ratio is compared to a user-defined tolerance, Φ. If $\varphi_{i,1} > \Phi$, the particle is removed from the bath and copied for a user-defined number of timesteps.

After this, it is returned to the bath in a manner that conserves the energy. The user can thus control the accuracy of the dynamics by judiciously selecting Φ. As $\Phi \rightarrow 0$, the dynamics become exact; as $\Phi \rightarrow \infty$, conventional multiple-copy dynamics results. EXACT performed better than LES in each test case, and differs from cLES because of the scaling property found in Φ, which scales both the level of accuracy and the amount of computer time spent to obtain results of the desired accuracy.

3.7 REFERENCES

1. Feynman, R. P., *Statistical Mechanics: A Set of Lectures*. W. A. Benjamin, Inc.: Reading, MA, 1972.

2. Tuckerman, M. E.; Martyna, G. J., Understanding modern molecular dynamics: Techniques and applications. *Journal of Physical Chemistry B* **2000,** 104, (2), 159-178.

3. Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*. Clarendon Press: Oxford, 1987.

4. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **1993,** 93, (7), 2395-2417.

5. Berne, B. J.; Straub, J. E., Novel methods of sampling phase space in the simulation of biological systems. *Current Opinion in Structural Biology* **1997,** 7, (2), 181-189.

6. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Optimization methods for conformational sampling using a Boltzmann-weighted mean field approach. *Biopolymers* **1996,** 39, (1), 103-114.

7. Huber, T.; van Gunsteren, W. F., SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A* **1998,** 102, (29), 5937-5943.

8. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

9. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

10. Ma, J. P.; Hsu, D.; Straub, J. E., Approximate Solution of the Classical Liouville Equation Using Gaussian Phase Packet Dynamics - Application to Enhanced Equilibrium Averaging and Global Optimization. *Journal of Chemical Physics* **1993,** 99, (5), 4024-4035.

11. Maranas, C. D.; Floudas, C. A., A Deterministic Global Optimization Approach for Molecular-Structure Determination. *Journal of Chemical Physics* **1994,** 100, (2), 1247-1261.

12. Piela, L.; Kostrowicki, J.; Scheraga, H. A., The Multiple-Minima Problem in the Conformational-Analysis of Molecules - Deformation of the Potential-Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry* **1989,** 93, (8), 3339-3346.

13. Tsallis, C.; Stariolo, D. A., Generalized simulated annealing. *Physica A* **1996,** 233, (1-2), 395-406.

14. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics. *Journal of Physical Chemistry A* **1997,** 101, (33), 5926-5930.

15. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

16. Gerber, R. B.; Buch, V.; Ratner, M. A., Time-Dependent Self-Consistent Field Approximation for Intramolecular Energy-Transfer .1. Formulation and Application to Dissociation of Vanderwaals Molecules. *Journal of Chemical Physics* **1982,** 77, (6), 3022-3030.

17. Roitberg, A.; Elber, R., Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* **1991,** 95, (12), 9277-87.

18. Simmerling, C.; Fox, T.; Kollman, P. A., Use of locally enhanced sampling in free energy calculations: Testing and application to the alpha ->beta anomerization of glucose. *Journal of the American Chemical Society* **1998,** 120, (23), 5771-5782.

19. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

20. Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N., Jr.; Dou, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; et al., Structural and functional effects of apolar mutations of the distal valine in myoglobin. *Journal of Molecular Biology* **1995,** 245, (4), 416-36.

21. Ulitsky, A.; Elber, R., Application of the Locally Enhanced Sampling (Les) and a Mean-Field with a Binary Collision Correction (Cles) to the Simulation of Ar Diffusion and No Recombination in Myoglobin. *Journal of Physical Chemistry* **1994,** 98, (3), 1034-1043.

22. Caflisch, A.; Miranker, A.; Karplus, M., Multiple Copy Simultaneous Search and Construction of Ligands in Binding-Sites - Application to Inhibitors of Hiv-1 Aspartic Proteinase. *Journal of Medicinal Chemistry* **1993,** 36, (15), 2142-2167.

23. Carlson, H. A.; Masukawa, K. M.; McCammon, J. A., Method for including the dynamic fluctuations of a protein in computer-aided drug design. *Journal of Physical Chemistry A* **1999,** 103, (49), 10213-10219.

24. Miranker, A.; Karplus, M., Functionality Maps of Binding-Sites - a Multiple Copy Simultaneous Search Method. *Proteins-Structure Function and Genetics* **1991,** 11, (1), 29-34.

25. Hixson, C. A.; Wheeler, R. A., Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles. *Physical Review E* **2001,** 64, 026701.

26. Straub, J. E.; Karplus, M., Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation. *Journal of Chemical Physics* **1991,** 94, (10), 6737-6739.

27. Stultz, C. M.; Karplus, M., On the potential surface of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1998,** 109, (20), 8809-8815.

28. Ulitsky, A.; Elber, R., The Thermal-Equilibrium Aspects of the Time-Dependent Hartree and the Locally Enhanced Sampling Approximations - Formal Properties, a Correction, and Computational Examples for Rare-Gas Clusters. *Journal of Chemical Physics* **1993,** 98, (4), 3380-3388.

29. Zheng, W. M.; Zheng, Q., An analytical derivation of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1997,** 106, (3), 1191-1194.

30. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *Journal of Chemical Physics* **1992,** 97, (4), 2635-2643.

31. Tuckerman, M. E.; Mundy, C. J.; Martyna, G. J., On the classical statistical mechanics of non-Hamiltonian systems. *Europhysics Letters* **1999,** 45, (2), 149-155.

32. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R., A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *Journal of Chemical Physics* **1982,** 76, (1), 637-649.

33. Hirschfelder, J. O.; Curtiss, C. F.; Bird, R. B., *Molecular Theory of Gases and Liquids*. John Wiley and Sons: New York, 1964.

CHAPTER 4

New perspectives on multiple-copy, mean-field molecular dynamics methods

4.1 INTRODUCTION

Molecular dynamics (MD) has been a part of computational chemistry for some time[1-3] and has proven its value for solving problems in classical mechanics and quantum mechanics.[3-5] Despite this fact, MD has several well known limitations. One well-known limitation is the difficulty an MD simulation experiences in moving from one energy well to another. For example, if one is interested in using MD to find global minima on a complicated potential energy surface, a long simulation must be performed before there is much chance that all available areas of phase space are explored. Practically, even this may not be enough, but there are methods available to address this difficulty.[6-14]

A popular method of addressing the above problem uses the same laws of physics in the solution that are the source of the problem. This method, called simulated annealing,[9, 14, 15] raises and lowers the temperature to facilitate the movement of the system between energy minima. Raising the temperature increases the probability that a barrier crossing can occur; subsequent temperature lowering traps the system, presumably in a lower energy well.

Other possible solutions for the sampling problem have been proposed.[11, 13, 16] One group of methods, though, uses an approximation of the laws of physics in order

to hasten energy minima transitions. These methods, called mean-field methods,[10, 17, 18] divide the system into a bath and a smaller, more interesting section. The smaller section is copied several times, and the simulation is run using a mean-field approximation. Stated another way, a group of non-interacting copies of the smaller system are allowed to interact with the larger bath. The force each particle in the copied sections experiences is the force it would normally feel, while the force each particle in the bath experiences is the averaged force of the bath's interaction with the copied sections. Despite the fact that the trajectories generated with such methods do necessarily not correspond to physically possible trajectories,[19-24] they have been used to find global minima,[25-28] to study non-equilibrium behavior,[29, 30] to enhance free-energy calculations,[31] and to search for molecules that bind to an active site.[32-34] Though many mean-field methods exist,[10, 17, 18, 20, 23, 32] this work will focus on the most commonly used method, called locally enhanced sampling (LES),[17] and our contribution, called Ensembles eXtracted from Atomic Coordinate Transformations (the EXACT approximation).[19, 20]

The mean-field approximation, though resulting in desirable features can also cause defects in mean-field trajectories and quantities derived from them.[19-24] The consequences of such approximations include the fact that mean-field trajectories violate the equipartition of energy theorem[21, 23, 24] and that local minima of a mean-field trajectory do not necessarily correspond to minima on a physically accurate trajectory.[22] The equipartition of energy violation is especially important and has been studied by many workers.[19-21, 23, 24]

This group's contributions[19, 20, 25] to this field include both theoretical and

practical contributions important for understanding and improving mean-field methods. For example, we have provided two new approaches that result in mean-field equations of motion. These methods each provide unique insights. Further, we developed a method that can be used to control the degree of approximation included in a calculations, allowing a researcher to decide the degree of sampling or accuracy a calculation should posses.

This paper is divided into three sections. The first reviews studies that inspired our work, the second reviews our theoretical contributions, and the last section shares some of our practical experience and characterizes some results.

4.2 THEORETICAL BACKGROUND

4.2.1 Original TDH approximation Gerber, Buch, and Ratner[18] originally devised a mean-field method by taking the classical limit of the time-dependent self-consistent field method (TDSCF). This method is described below by first describing the TDSCF method, and then explaining the path to the classical limit that defines the self-consistent trajectory bundles method. This method was used to characterize the classical vibrations of small molecules, and its copies are designed to represent different normal modes of the molecule's vibrations.

The TDSCF method starts with the assumption that a system can be described by a Hartree-product wavefunction,

$$\Psi(\mathbf{x},t) = \prod_{i=1}^{N} \psi_i(\mathbf{x}_i,t), \qquad (4.1)$$

where Ψ is the wavefunction, $\mathbf{x}$ is a vector describing all N degrees of freedom in the system, ψ_i is a wavefunction which describes the *i*th normal mode, and $\mathbf{x}_i$ is a vector

that describes that normal mode. Time is represented by t.

Applying Equation 4.1 to the time-dependent Schrödinger equation and then operating with $\langle \psi_j |$ gives

$$i\hbar \frac{\partial \varphi_i}{\partial t} = h_i^{SCF}(\mathbf{x}_i, t)\varphi_i(\mathbf{x}_i, t), \qquad (4.2)$$

where $\varphi_i(\mathbf{x}_i, t)$ is $e^{i\omega t}\psi_i(\mathbf{x}_i, t)$. Most importantly, $h_i^{SCF}(\mathbf{x}_i, t)$ is the kinetic and potential energies of the *i*th normal-mode plus an extra average term. So,

$$h_i^{SCF}(\mathbf{x}_i, t) = T_i + V(\mathbf{x}_i) + \overline{V}_i(\mathbf{x}_i, t). \qquad (4.3)$$

The final term of this expression is an averaged potential of the form

$$\overline{V}_i(\mathbf{x}_i, t) = \sum_{j \neq i} \langle \varphi_j | V_{ij}(\mathbf{x}_i, \mathbf{x}_j) | \varphi_i \rangle, \qquad (4.4)$$

where *i* and *j* are normal-mode labels. Thus, $\overline{V}_i(\mathbf{x}_i, \mathbf{x}_j)$ is the interaction energy between modes *i* and *j* . This term was the original basis for the development of mean-field methods, which include LES, and the self-consistent trajectory bundle method in particular. It is a "mean-field" term because it averages the interaction energy of the *i*th normal mode with all the other normal modes of the system.

The self-consistent trajectory bundle method is the classical correspondence of the above method, and redefines the average classical potential as

$$\overline{V}_i(\mathbf{x}_i, t) = \frac{1}{N} \sum_{i \neq j} \sum_{\lambda=1}^{C} V_{ij}(\mathbf{x}_i, \mathbf{x}_j^{(\lambda)}), \qquad (4.5)$$

where *i* and *j* are normal mode labels, and $\mathbf{x}_j^{(\lambda)}$ describes the λth copy of the *j*th normal mode. So in this method, all the normal modes are simultaneously simulated, but one is specially prepared in different states. These states interact with the other normal modes according to the potential described in Equation 4.5, but not with other

states.

4.2.2 Phase space TDH Elber and Karplus[17] begin their derivation of LES with the assumption that the classical phase-space density function can be expressed as a product of the "copied" sub-system's density and the "bath" sub-system's density:

$$\rho\left(\vec{\mathbf{X}},t\right) = \rho_s\left(\vec{\mathbf{X}}_s,t\right)\rho_b\left(\vec{\mathbf{X}}_b,t\right). \qquad (4.6)$$

Here $\vec{\mathbf{X}}$ is a vector representing all degrees of freedom in the system and *t* is the time. They further assumed that the bath's density can be written as a single delta function:

$$\rho_b\left(\vec{\mathbf{X}},t\right) = \delta\left(\vec{\mathbf{X}}_b(t)\right), \qquad (4.7)$$

while the copied sub-system's density can be written as a "swarm" of delta functions. Thus,

$$\rho_s\left(\vec{\mathbf{X}}_s,t\right) = \sum_{k=1}^{C} w_{sk}\delta\left(\vec{\mathbf{X}}_s(t)\right), \qquad (4.8)$$

where each delta function represents the positions in phase space of the various copies, and w_{sk} is a weighting function. In LES the weighting functions are generally taken to be $1/C$, where C is the number of copies. Elber and Karplus refer to this as the time-dependent Hartree (TDH) approximation, in analogy with Gerber, Buch, and Ratner's[18] previous work, described above.

Using this assumed density, they derive the approximate equations of motion by proving that

$$\frac{\partial\langle Q_j\rangle}{\partial t} = \left\langle \frac{\partial H}{\partial P_j} \right\rangle; \frac{\partial\langle P_j\rangle}{\partial t} = -\left\langle \frac{\partial H}{\partial Q_j} \right\rangle, \qquad (4.9)$$

where H is the Hamiltonian for one copy interacting with the bath, Q_j is a generalized coordinate of the *j*th particle, and P_j is the momentum of the *j*th particle.

They then use their assumed density to determine the LES equations of motion:

$$\dot{q}_{i,k} = \frac{\partial H_k}{\partial p_{i,k}}, \qquad (4.10a)$$

$$\dot{p}_{i,k} = -\frac{\partial H_k}{\partial q_{i,k}}, \qquad (4.10b)$$

$$\dot{Q}_i = \sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial P_i}, \qquad (4.10c)$$

$$\dot{P}_i = -\sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial Q_i}. \qquad (4.10d)$$

The index *i* refers to the particle and the index *k* refers to the copy. The lower case variables refer to the copied sub-system while the uppercase variables refer to the "bath."

4.2.3 Limitations of LES Because LES and other mean-field techniques employ an approximate dynamics, there are limits to the ability of these trajectories to model a Newtonian trajectory. One of the limitations include the fact that LES violates the equipartition of energy theorem.[21, 23] Some have speculated[21, 23, 24] that this violation results in the "temperature disparity problem," which is a failure of the sub-system and bath temperatures to reach the same equilibrium value. Another problem is that geometry-optimization problems solved in LES are definitive only if the global energy minimum is desired. Stultz and Karplus have proven that local-minima found using LES cannot be assumed to be minima on the original energy surface.[22]

4.3 THEORETICAL CONTRIBUTIONS

4.3.1 Classical mechanical approach We derived[19] mean-field equations of motion in an attempt to better understand the approximation involved in methods

such as LES and to improve on them. The derivation begins by writing the Hamiltonian of C non-interacting copies of the original system,

$$H = \sum_{k=1}^{C}\left(\sum_{i=1}^{S} \frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S} V\left(q_{i,k}, q_{j,k}\right) + \sum_{i=1}^{S}\sum_{j=1}^{N} V\left(q_{i,k}, Q_{j,k}\right) + \sum_{i=1}^{N} \frac{P_{i,k}^2}{2M_i} + \sum_{i>j=1}^{N} V\left(Q_{i,k}, Q_{j,k}\right)\right). \quad (4.11)$$

Lowercase variables refer to the sub-system of interest while uppercase variables refer to the bath. The constants C , N , and S represent the number of copies, the number of particles in the bath, and the number of particles in the sub-system of interest respectively. Upper and lowercase q's indicate positions, p's indicate momenta, and m's indicate masses. The first two terms are the energy due to the copied subsystem, the middle term is the interaction energy between the copies and the bath, and the final two terms are the energy due to the bath. This Hamiltonian, for the sake of simplicity, assumes a pair form of the potential. We then make a point transformation defined by

$$Q_i = \frac{1}{n}\left(\sum_{k=1}^{C} Q_{i,k} \right), \quad (4.12a)$$

$$P_i = \frac{1}{n}\left(\sum_{k=1}^{C} P_{i,k} \right), \quad (4.12b)$$

$$Q'_{i,l} = \frac{1}{n}\left(Q_{i,1} + \sum_{k=2}^{C} c_{l,k} Q_{i,k} \right), \quad (4.12c)$$

$$P'_{i,l} = \frac{1}{n}\left(P_{i,1} + \sum_{k=2}^{C} c_{l,k} P_{i,k} \right), \quad (4.12d)$$

where $\{Q_i\}_{i=1}^{N}$ and $\{P_i\}_{i=1}^{N}$ are the ‘‘major ’’ variables, $\{Q'_{i,l}\}_{i=1,l=2}^{N,C}$ and $\{P'_{i,l}\}_{i=1,l=2}^{N,C}$ are the “minor” variables, and $\{c_{i,j}\}$ are the transform coefficients. The constant C is equal to the number of copies. The “major” variables denote the average coordinates of all copies of the bath particles, and the “minor” variables complete the description of the exact dynamics. For future use, the transform coefficients possess three useful properties - orthonormality, Hermiticity, and zero-average:

$$1+\sum_{k=2}^{C} c_{n,k}c_{m,k} = C\delta_{m,n}\,, \qquad (4.13a)$$

$$c_{ij} = c_{ij}^{*}, \qquad (4.13b)$$

and

$$1+\sum_{k=2}^{C} c_{n,k} = 0\,. \qquad (4.13c)$$

Equation 4.13b is actually a more stringent requirement than necessary, as we expect all the coefficients to be real. This transformation was used in a previous work[24] attempting to explain LES, but its solution was limited to the harmonic approximation.

For a general potential, the transformed Hamiltonian can be written as[19]

$$\tilde{H} = \sum_{i=1}^{S}\sum_{k=1}^{C}\frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S}\sum_{k=1}^{C}\left(q_{i,k},q_{j,k}\right) + \sum_{j=1}^{N}\frac{n^2P_j^2}{2M_jC}$$
$$+\sum_{i>j=1}^{N}\sum_{k=1}^{C}V\left(\frac{C}{n}\left(Q_i+\sum_{k=2}^{C}Q'_{i,k}\right),\frac{C}{n}\left(Q_j+\sum_{k=2}^{C}Q'_{j,k}\right)\right)$$

$$+\sum_{j=1}^{C}\sum_{k=2}^{C}\frac{n^2 P'^2_{j,k}}{2M_j C}+\sum_{i=1}^{N}\sum_{j=1}^{S}\sum_{k=1}^{C}V\left(q_{j,k},\frac{C}{n}\left(Q_i+\sum_{k=2}^{C}Q'_{i,k}\right)\right). \quad (4.14)$$

The first two terms represent the Hamiltonian of the copied particles. The next three terms in Equation 4.14 are the kinetic energy of the major variables of the bath, the potential energy due to the bath, and the kinetic energy due to the minor variables of the bath. The final term contains the interaction energy between the two subsystems. The equations of motion for the major and minor variables can then be obtained:

$$\dot{q}_{i,k}=\frac{\partial\tilde{H}}{\partial p_{i,k}}, \quad (4.15a)$$

$$\dot{p}_{i,k}=-\frac{\partial\tilde{H}}{\partial q_{i,k}}, \quad (4.15b)$$

$$\dot{Q}_i=\frac{C}{n^2}\frac{\partial\tilde{H}}{\partial P_i}, \quad (4.15c)$$

$$\dot{P}_i=-\frac{C}{n^2}\frac{\partial\tilde{H}}{\partial Q_i}, \quad (4.15d)$$

$$\dot{Q}'_{i,k}=\frac{C}{n^2}\frac{\partial\tilde{H}}{\partial P'_{i,k}}, \quad (4.15e)$$

$$\dot{P}'_{i,k}=-\frac{C}{n^2}\frac{\partial\tilde{H}}{\partial Q'_{i,k}}. \quad (4.15f)$$

If the bath particles are initially in the same phase-space coordinates and all of the coordinates of the minor variables are neglected, the LES equations of motion are recovered.

4.3.2 Liouville operator approach A complimentary view of the results from the previous section can also be gained by using a phase-space density approach.[19] In fact, most theoretical work done regarding mean-field methods depends on a phase-

space density approach,[10, 17, 21, 23, 24] and often employs the Liouville equation.

The start of the analysis requires a density that satisfies the Liouville equation,[35, 36]

$$\rho\left(\vec{\mathbf{X}},t\right)=e^{-\hat{\mathbf{L}}t}\rho\left(\vec{\mathbf{X}},0\right) \tag{4.16}$$

where $\hat{\mathbf{L}}$ is a Liouville operator,

$$\begin{aligned}\hat{\mathbf{L}}=&\sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}}-\frac{\partial H}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right)\\&+\sum_{i=1}^{N}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial P_{i,k}}\frac{\partial}{\partial Q_{i,k}}-\frac{\partial H}{\partial Q_{i,k}}\frac{\partial}{\partial P_{i,k}}\right)\end{aligned} \tag{4.17}$$

$\rho\left(\vec{\mathbf{X}},0\right)$ is the initial phase-space density:

$$\begin{aligned}\rho\left(\vec{\mathbf{X}},0\right)=&\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(q_{i,k}-q_{i,k,0}\right)\delta\left(p_{i,k}-p_{i,k,0}\right)\right)\\&\times\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(Q_{i,k}-Q_{i,k,0}\right)\delta\left(P_{i,k}-P_{i,k,0}\right)\right),\end{aligned} \tag{4.18}$$

and $\vec{\mathbf{X}}$ is a point in phase-space. After applying the transformation defined by Equation 4.12 to the Liouville operator and the initial probability density, we obtain

$$\begin{aligned}\hat{\mathbf{L}}=&\sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial\tilde{H}}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}}-\frac{\partial\tilde{H}}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right)\\&+\frac{C}{n^2}\left[\sum_{i=1}^{N}\left(\frac{\partial\tilde{H}}{\partial P_i}\frac{\partial}{\partial Q_i}-\frac{\partial\tilde{H}}{\partial Q_i}\frac{\partial}{\partial P_i}\right)+\sum_{i=1}^{N}\sum_{k=2}^{C}\left(\frac{\partial\tilde{H}}{\partial P'_{i,k}}\frac{\partial}{\partial Q'_{i,k}}-\frac{\partial\tilde{H}}{\partial Q'_{i,k}}\frac{\partial}{\partial P'_{i,k}}\right)\right].\end{aligned} \tag{4.19}$$

With the assumption that the bath particles are in the same locations in each of the copies, the transformed density can be written as:

$$\begin{aligned}\rho\left(\vec{\mathbf{X}},0\right)=&\frac{1}{\mathrm{J}\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)}\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta\left(q_{i,k}-q_{i,k,0}\right)\delta\left(p_{i,k}-p_{i,k,0}\right)\right)\\&\times\left(\prod_{i=1}^{N}\delta\left(Q_i-Q_{i,0}\right)\delta\left(P_i-P_{i,0}\right)\prod_{k=2}^{C}\delta\left(Q'_{i,k}\right)\delta\left(P'_{i,k}\right)\right).\end{aligned} \tag{4.20}$$

Here, $J\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)$ is the Jacobian of the transformation and $\tilde{\vec{\mathbf{X}}}$ is a point in the transformed phase-space. The transformed time dependent phase-space density, $\tilde{\rho}\left(\tilde{\vec{\mathbf{X}}},t\right)$, is written using Equations 4.19 and 4.20 in the Liouville equation. Following Zheng and Zheng's work[24] the reduced density, $\tilde{\rho}\left(\tilde{\vec{\mathbf{X}}},t\right)$,that can be used to generate LES-type equations of motion is found when the minor variables are integrated out, as in

$$\tilde{\rho}'\left(\tilde{\vec{\mathbf{X}}},t\right)=\int d\gamma'\left|J\left(\vec{\mathbf{X}},\tilde{\vec{\mathbf{X}}}\right)\right|e^{\hat{\mathbf{L}}t}\tilde{\rho}\left(\tilde{\vec{\mathbf{X}}},0\right), \qquad (4.21)$$

where dγ is the minor variables' volume element. Solving Equation 4.21 serves the purpose of generating a density that can be used to obtain the LES equations of motion shown in Equation 4.15. It does so by enforcing the holonomic constraints shown to result in LES.

4.3.3 Explanation of the "temperature disparity" problem In the canonical ensemble, the average, $\langle q_i F_i\rangle$,for a bound system can be written as

$$\langle q_i F_i\rangle=-\frac{1}{Z}\int d\Gamma dq_i q_i\frac{\partial H}{\partial q_i}e^{-\beta\,\mathrm{H}(q_i,\Gamma)} \qquad (4.22)$$

Here, Q is the partition function, q_i is a general coordinate of the *i*th atom, and F_i is the force on the *i*th atom. The integral's exact value is $-1/\beta$, where β is $1/k_BT$ (k_B is the Boltzmann constant). Therefore, after applying the virial theorem the average kinetic energy for C copies of a structure is simply

$$\begin{aligned}\overline{T}&=-\tfrac{1}{2}\sum_{i=1}^{S}\sum_{k=1}^{C}\left\langle q_{i,k}\frac{\partial H}{\partial q_{i,k}}\right\rangle-\tfrac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{C}\left\langle Q_{i,k}\frac{\partial H}{\partial Q_{i,k}}\right\rangle\\&=\frac{(SC+NC)}{2\beta}\end{aligned} \qquad (4.23)$$

After making the transformation,

$$\left\langle Q_{i,k}\frac{\partial H}{\partial Q_{i,k}}\right\rangle = \left\langle \frac{C^2}{n^3}\left(Q_i + \sum_{i=2}^{C} c_{1,k}Q'_{i,l}\right)\left(\frac{\partial H}{\partial Q_i} + \sum_{m=2}^{C} c_{m,k}\frac{\partial H}{\partial Q'_{i,m}}\right)\right\rangle \quad (4.24)$$

and using the result in the virial theorem, the exact result $\overline{T} = {}^{(NC+CS)}\!/_{2\beta}$ is generated.

However, if the minor variables are constrained to vanish, the average kinetic energy is $\overline{T} = {}^{(N+CS)}\!/_{2\beta}$. Consequently, the LES virial is incorrect because mean-field systems have fewer degrees of freedom, and this manifests itself as the temperature disparity problem.

4.3.4 Analysis of validity of the LES approximation It should be noted that the work shown above presents a purely classical mechanical description of the approximations inherent in mean-field methods. It also allows some comments to be made regarding the accuracy of mean-field trajectories. For example, mean-field methods can be viewed as a collection of separate systems coupled together through the requirement that the particles labeled as a "bath" are required to occupy the same positions in each system.

Knowing this allows one to consider the approximation's range of validity. From this work it should be expected that mean-field methods can accurately reproduce the correct classical trajectories when the force on the "minor" variables are small. Two conditions might give rise to this. The first occurs when the interactions between the copied particles and the bath particles are small, as might be expected in a low density simulation. The second condition is that the simulation is of short enough duration that the dispersal of the copies of the bath particles is not important. If either condition is true, the trajectory generated with a mean-field method would be

expected to be accurate. Unfortunately, useful simulations are frequently impossible to perform under these conditions, because long trajectories need to be analyzed or because the system of interest is in a condensed phase. Our contribution described in the next section[20] is designed to alleviate this problem, and allow mean-field methods to be used to simulate condensed-phase systems more accurately.

4.4 PRACTICAL CONTRIBUTIONS

4.4.1 The EXACT approximation

4.4.1.1 Description of the algorithm The main result from Equations 4.14 and 4.15 is that the LES approximation can only be accurately applied where the minor variables are small enough to be ignored. We developed a method that accounts for the minor variable's influence called the EXACT approximation.[19, 20] Equation 4.14 along with an initial holonomic constraint set to force the minor variables to vanish is the basis, but corrections are applied when the forces on the minor variable at each step are large when compared to the force on the corresponding major variable. The ratio $\varphi_{i,l} = \left| \frac{P'_{i,l}}{P_i} \right|$, the force on the minor variable divided by the force on the major variable, is calculated and then compared with a user-defined tolerance, Φ. For any coordinate associated with a particle in which the ratio, $\varphi_{i,l}$, is greater than the tolerance ($\Phi < \varphi_{i,l}$), the particle is removed from the bath and treated exactly (as a copied particle) for a user-defined number of time steps, and then returned to the bath in a manner that ensures that the energy is conserved. As $\Phi \rightarrow 0$, conventional dynamics for an ensemble of system copies is produced; when $\Phi \rightarrow \infty$, LES results.

4.4.1.2 Calculation of the "minor" forces It is necessary to calculate the force

on the minor coordinates to use the EXACT approximation. This is accomplished economically by first writing the forces as derivatives of the potential energy:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \sum_{k=1}^{C} \left(\frac{\partial V_{B,k}}{\partial Q'_{i,m}} + \frac{\partial V_{X,k}}{\partial Q'_{i,m}} \right). \qquad (4.25)$$

We have earlier shown that this can be rewritten as:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_B}{\partial Q_i} \left(1 + \sum_{k=2}^{C} c_{m,k} \right) + \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} \sum_{k=2}^{C} c_{m,k} \frac{\partial V_{X,k}}{\partial Q_i} \quad (4.26)$$

Finally, applying Equation 4.13 to Equation 4.26 yields:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} c_{m,k} \frac{\partial V_{X,k}}{\partial Q_i}. \qquad (4.27)$$

Equation 4.27 is important because it shows that the forces on the minor variables can be calculated with trivial effort since $\frac{\partial V_{X,k}}{\partial Q_i}$, the forces between bath atoms and copies, is already calculated - it is the force on the corresponding major variable. This equation also requires that the transform coefficients be explicitly known, unlike in other mean-field methods. We have previously published a method to calculate them.[20]

4.4.2 Extension of LES and EXACT to constant temperature ensembles

Because the Nosé-Hoover Chain (NHC) method[37] has been shown to be a useful method for constant temperature simulations, we showed[20] how to implement this algorithm for the mean-field methods LES and EXACT. An extended system designed to act as a heat bath is appended to the original Hamiltonian of a system. Thus, the Hamiltonian for an NHC system of C non-interacting systems can be written as:

$$H_{NHC} = H + \sum_{k=1}^{C}\left(\sum_{i=1}^{2}\frac{p_{\eta_i,k}^2}{2\mu_i} + k_B T\left[D\eta_{1,k} + \eta_{2,k}\right]\right). \qquad (4.28)$$

In this equation, H is given by Equation 4.11, $p_{\eta_i,k}$ and $\eta_{i,k}$ are the momenta and "position" of the heat bath's chain particles, and μ_i is the "mass" of the chain particle. The subscripts on $p_{\eta_i,k}$ indicate that this is the momentum of chain particle *i* and D is the number of particles to be copied. From the above Hamiltonian the forces on the major variables can be expressed as

$$\dot{P}_i = \frac{1}{C}\sum_{k=1}^{C}\dot{P}_{i,k} \qquad (4.29)$$

which can be rewritten using Equations (4.12) and (4.13) as

$$\dot{P}_i = -\frac{1}{C}\left(\frac{\partial H_{NHC}}{\partial Q_i} + P_i\sum_{k=1}^{C}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}} + \sum_{l=2}^{C}P'_{i,l}\sum_{k=1}^{C}c_{l,k}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}}\right). \quad (4.30)$$

An assumption of both LES (and EXACT when the minor variables of the bath are ignored) is that the minor variables expressing the constant temperature constraint can be assumed to vanish, so the final term can be ignored. Thus, the first two terms comprise the force needed to ensure a constant temperature simulation. Doing a transformation similar to the one used to derive Equation 4.30 gives the temperature controlled force on the minor variables, after applying the holonomic constraint shown to yield the LES equations of motion:

$$\dot{P}_{i,l} = -\frac{1}{C}\left[\frac{\partial H_{NHC}}{\partial Q'_{i,l}} + P_i\left(\frac{\partial H_{NHC}}{\partial p_{\eta_1,1}} + \sum_{k=2}^{C}c_{l,k}\frac{\partial H_{NHC}}{\partial p_{\eta_1,k}}\right)\right]. \quad (4.31)$$

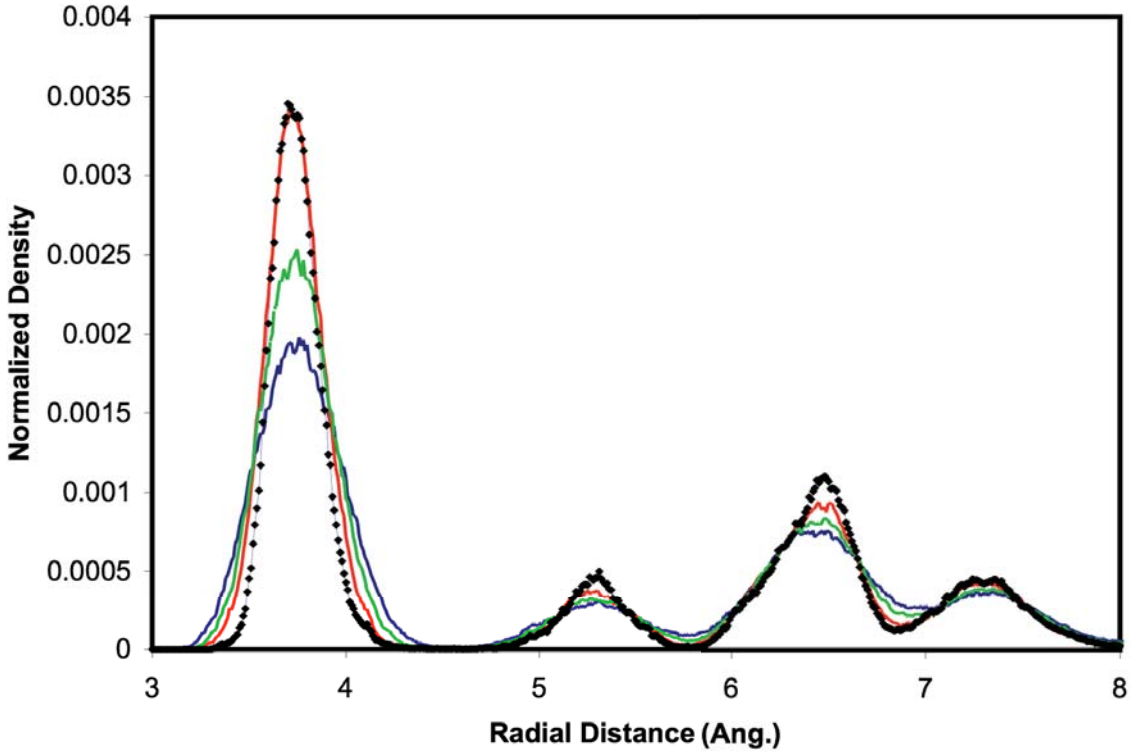
With these two equations, both LES and EXACT algorithms can be designed to simulate constant temperature trajectories. Similar logic can be used for barostats.[3, 38]

4.4.3 Tests of the EXACT approximation

4.4.3.1 Effect on simple pair distribution functions Figure 4.1 is a collection of pair distribution functions of a 64 atom cluster of argon.[20] The curve generated from purely LES data contains peaks located in the correct places, but they are too broad and short compared to the MD result. This is evidence that LES samples regions of space forbidden to particles in a conventional MD simulation. The EXACT approximation's results, however, scale between these two extremes. When the threshold is set with $\Phi = 0.8$, the simulation produces a taller and narrower initial peak. When the threshold is set with $\Phi = 0.3$, the simulation produces a result that is only slightly different from the conventional MD result. By carefully selecting the threshold, Φ, the simulation's sampling properties can range from conventional MD for C non-interacting copies to the mean-field (LES) limit.

4.4.3.2 Effect on cooling behavior of copied particles The same Ar cluster described above was used to perform a cooling simulation using conventional MD, LES, and the constant energy implementation of the EXACT algorithm.[20] The simulation consisted of choosing a particle near the center of the cluster, and raising its kinetic energy to 300K, and observing its coolings. Results are shown in Figure 4.2, and demonstrate the effect of the "temperature-disparity problem." The problem has been described as the failure of a hot particle to cool properly during a mean-field simulation, and has been the focus of several theoretical investigations. [19, 21, 23, 24] The crux of the problem is that although the Ar atom cools quickly to the bath temperature using conventional MD, LES simulations do not seem to demonstrate any cooling at all. Also shown in the figure are results generated by the EXACT approximation

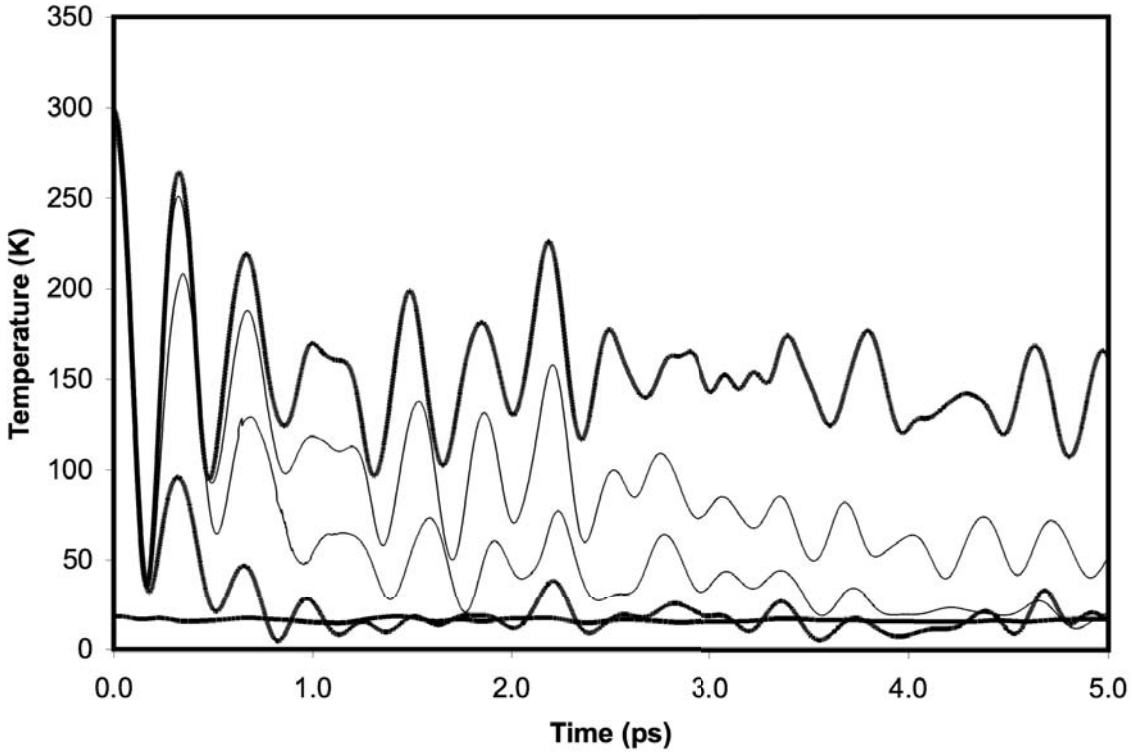
Figure 4.1. Normalized density distribution for a single, copied Ar atom in a bath of Ar. Simulations correspond to conventional MD (highest, narrowest peak, marked by diamonds), LES (shortest, broadest peak, blue), and intermediate cases, generated using the EXACT approximation (red and green). The EXACT approximation allows interpolation between conventional MD and the enhanced sampling of LES.



using thresholds Φ = 0.8 and 0.3. The 0.8 simulation improves the simulation, but not markedly. The Φ = 0.3 simulation, though, relaxes to the correct temperature within 5 ps. This is a clear improvement over LES.

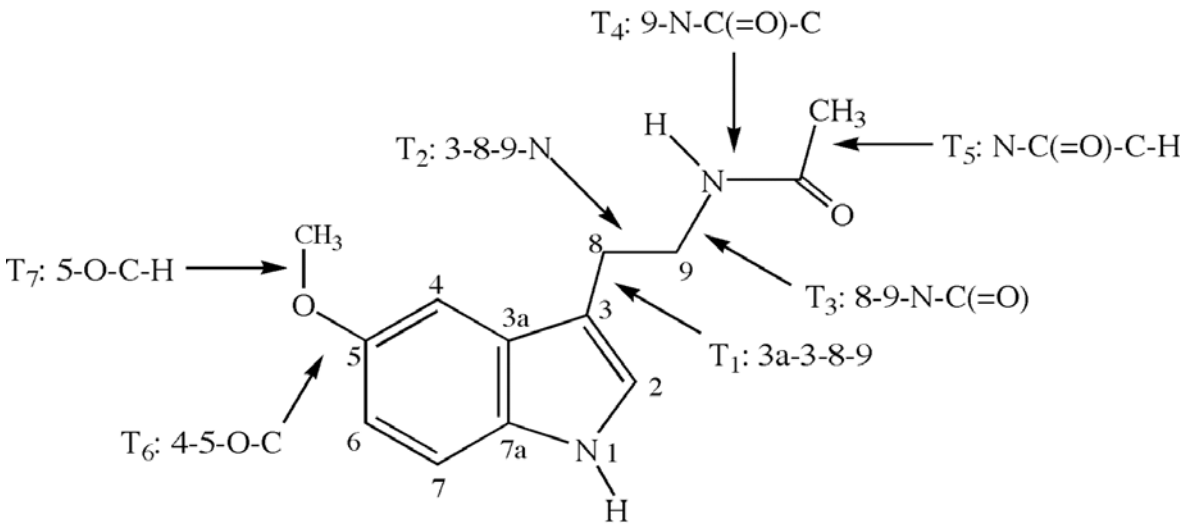
4.4.3.3 Effect on sampling a torsion angle of melatonin Melatonin,[39] a tryptophan derivative, is a molecule with seven flexible torsion angles, as shown in Figure 4.3. There are three main sections of melatonin which include the 5-methoxy group, the indole ring, and the peptide-like side chain located on the C3 position of

Figure 4.2. Comparison of temperature vs. time for the bath of 64 Ar atoms (flattest curve near the bottom) with the temperature of one Ar atom in conventional MD (bold curve oscillating about the bath temperature), LES (bold curve at the top showing large oscillations), and intermediate cases, generated using the EXACT approximation. The EXACT approximation allows the hot Ar to cool to the temperature of the bath but LES does not.



the ring. The dihedral angles of the methoxy group and side chain determine the molecule's spatial conformation, while the indole ring remains essentially planar. The methoxy group has two torsion angles - one represents the nearly free rotation of the hydrogens around the methyl group (T7) and the other determines the position of the methyl carbon group (T6). The peptide-like side chain contains the balance of the flexible torsion angles, of which only T1 , T2 , and T3 are both interesting and free to rotate, as T4 is a peptide bond, and T5 determines the conformation of a nearly free

Figure 4.3. Structure of melatonin, along with its standard numbering scheme. Interesting torsion angles are labeled as T_i .



methyl rotor.

Melatonin proves an excellent test system to illustrate the ability of a method to enhance conformational sampling, as its T1 torsion angle must overcome a relatively high energy barrier to rotate the side chain from one face of the indole ring to the other. As a consequence, T1 is expected to be relatively fixed with an angle near ±90°. It is around this torsion angle that the boundary is made between the copied region and the bath. Everything in the indole ring is designated bath, every atom in the sidechain is copied four times. Thus, the T1 torsion angle is expected to enjoy the bulk of the enhanced sampling in EXACT simulations.

For this test, a gas phase conventional MD simulation of melatonin was performed as described in Sec. (4.5) to observe the degree of sampling normally found in molecular dynamics simulations of this molecule. This was followed by a

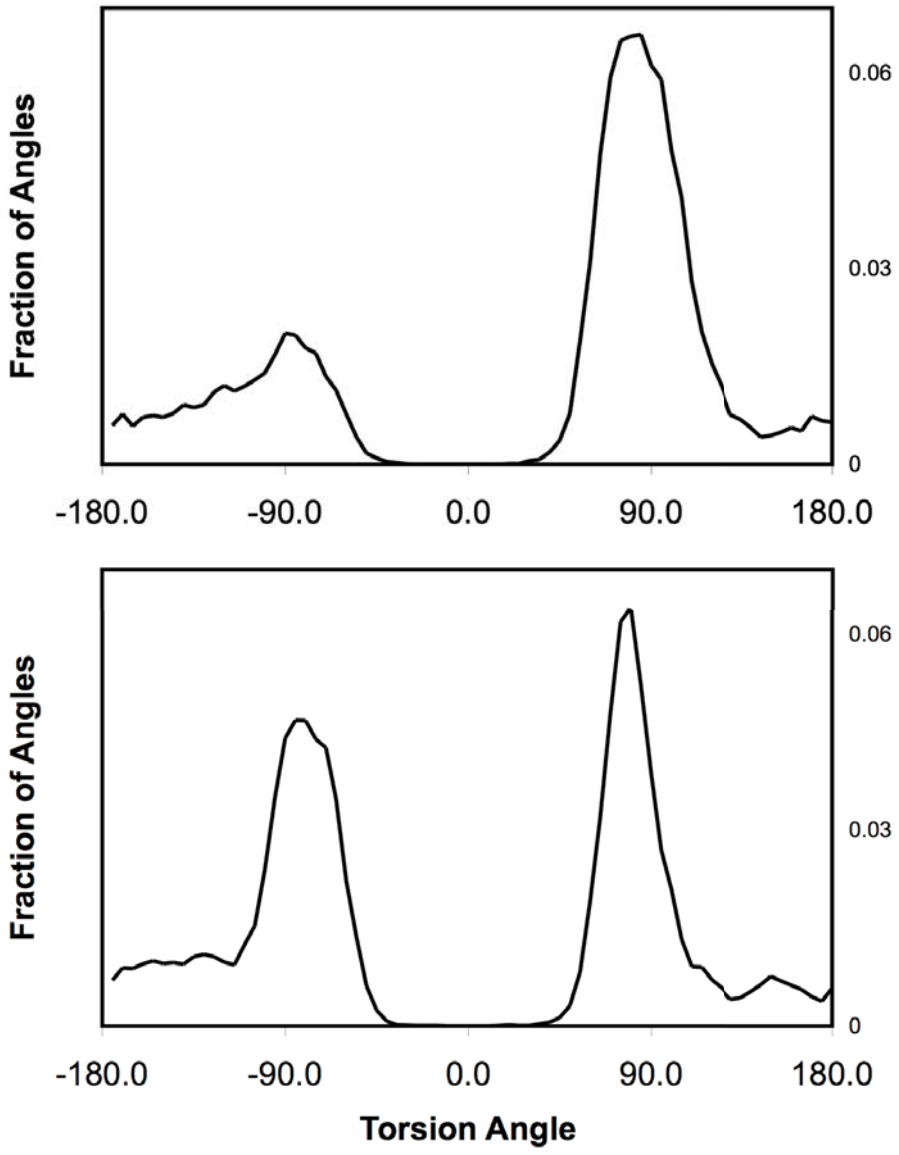
gas phase EXACT simulation expected to benefit from enhanced sampling, using a threshold expected to be near the LES limit ($\Phi \rightarrow \infty$). Each of these tests were allowed 300 ps of simulation time.

Histograms of the T1 dihedral are presented for both the gas phase conventional MD and EXACT simulation of 300 ps in Figure 4.4a and Figure 4.4b, respectively. The conventional simulation does not provide balanced sampling on both faces of melatonin's indole ring (defined by T1 = $\pm 90^{\circ}$). The positive dihedral region is greatly favored in the conventional simulation, because there was not adequate time for equal sampling to be achieved. For the EXACT simulation, both faces of melatonin's indole ring are sampled nearly equally. The centers of the peaks are at $\pm 90^{\circ}$, and are almost equal in area.

4.5 DETAILS OF MELATONIN SIMULATION

Simulations on melatonin were performed to ensure that EXACT could properly allow for the enhanced sampling that mean-field methods allow.[25] The melatonin simulations used the AMBER94 all-atom force field[40] with charges obtained using electrostatic fitting from electron densities obtained using Gaussian 94,[41] at the B3LYP/6-31G* level of theory. Gas-phase simulations of melatonin were performed using our own implementation, using the molecular modeling toolkit (MMTK)[42] to prepare the initial conditions. Temperature was controlled with the Nosé-Hoover chain algorithm.[37] Covalent bond distances involving hydrogen were constrained using the SHAKE algorithm[43] for conventional MD, and with a slightly modified SHAKE (to account for the common bath) in the EXACT approximation simulations. Atoms involving the side-chain of melatonin (which is the functional group attached

Figure 4.4. (a) Normalized histogram recording the number of times melatonin's T1 torsional angle visited a given value during a 300 ps gas-phase conventional MD simulation. The asymmetry indicates preferential sampling on one side of the indole ring. (b) Normalized histogram recording the number of times melatonin's T1 torsional angle visited a given value during a 300 ps gas-phase EXACT simulation. More balanced sampling on both sides of the indole ring is achieved, as indicated by the more symmetrical distribution of torsion angles.



to carbon 3 in the indole ring, as depicted in Figure 4.3) were copied four times in the EXACT simulations.

4.6 CONCLUSIONS

LES and other mean-field methods have been successfully used in a variety of applications[26, 27, 29-34] and have been the subject of a great deal of theoretical scrutiny.[19-24] Two new routes to obtain the mean-field equations of motion shown here add interesting insights to the understanding of mean-field methods. We show that mean-field methods can be viewed as a type of constrained dynamics.[19] These constraints reduce the number of degrees of freedom in the system, which affects the temperature equilibration properties of the system. Further, the classical-mechanical view demonstrates the degree of approximation in mean-field methods and indicates that these methods can be truly accurate under certain conditions. Finally, we use this knowledge to create a new method, called the EXACT approximation,[20] that is designed to allow a researcher to decide the degree of accuracy or enhanced sampling a simulation requires.

4.7 REFERENCES

1. Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*. Clarendon Press: Oxford, 1987.

2. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **1993,** 93, (7), 2395-2417.

3. Tuckerman, M. E.; Martyna, G. J., Understanding modern molecular dynamics: Techniques and applications. *Journal of Physical Chemistry B* **2000,** 104, (2), 159-178.

4. Feynman, R. P., *Statistical Mechanics: A Set of Lectures*. W. A. Benjamin, Inc.: Reading, MA, 1972.

5. Makri, N., Time-dependent quantum methods for large systems. *Annual Review of Physical Chemistry* **1999,** 50, 167-191.

6. Berne, B. J.; Straub, J. E., Novel methods of sampling phase space in the simulation of biological systems. *Current Opinion in Structural Biology* **1997,** 7, (2), 181-189.

7. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Optimization methods for conformational sampling using a Boltzmann-weighted mean field approach. *Biopolymers* **1996,** 39, (1), 103-114.

8. Huber, T.; van Gunsteren, W. F., SWARM-MD: Searching conformational space by cooperative molecular dynamics. *Journal of Physical Chemistry A* **1998,** 102, (29), 5937-5943.

9. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

10. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

11. Ma, J. P.; Hsu, D.; Straub, J. E., Approximate Solution of the Classical Liouville Equation Using Gaussian Phase Packet Dynamics - Application to Enhanced Equilibrium Averaging and Global Optimization. *Journal of Chemical Physics* **1993,** 99, (5), 4024-4035.

12. Maranas, C. D.; Floudas, C. A., A Deterministic Global Optimization Approach for Molecular-Structure Determination. *Journal of Chemical Physics* **1994,** 100, (2), 1247-1261.

13. Piela, L.; Kostrowicki, J.; Scheraga, H. A., The Multiple-Minima Problem in the Conformational-Analysis of Molecules - Deformation of the Potential-Energy Hypersurface by the Diffusion Equation Method. *Journal of Physical Chemistry* **1989,** 93, (8), 3339-3346.

14. Tsallis, C.; Stariolo, D. A., Generalized simulated annealing. *Physica A* **1996,** 233, (1-2), 395-406.

15. Huber, T.; Torda, A. E.; vanGunsteren, W. F., Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics. *Journal of Physical Chemistry A* **1997,** 101, (33), 5926-5930.

16. Zhu, Z. W.; Tuckerman, M. E.; Samuelson, S. O.; Martyna, G. J., Using novel variable transformations to enhance conformational sampling in molecular dynamics. *Physical Review Letters* **2002,** 88, (10).

17. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

18. Gerber, R. B.; Buch, V.; Ratner, M. A., Time-Dependent Self-Consistent Field Approximation for Intramolecular Energy-Transfer .1. Formulation and Application to Dissociation of Vanderwaals Molecules. *Journal of Chemical Physics* **1982,** 77, (6), 3022-3030.

19. Hixson, C. A.; Wheeler, R. A., Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles. *Physical Review E* **2001,** 64, 026701.

20. Hixson, C. A.; Wheeler, R. A., Practical multiple-copy methods for sampling classical statistical mechanical ensembles. *Chemical Physics Letters* **2004,** 386, 330-335.

21. Straub, J. E.; Karplus, M., Energy Equipartitioning in the Classical Time-Dependent Hartree Approximation. *Journal of Chemical Physics* **1991,** 94, (10), 6737-6739.

22. Stultz, C. M.; Karplus, M., On the potential surface of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1998,** 109, (20), 8809-8815.

23. Ulitsky, A.; Elber, R., The Thermal-Equilibrium Aspects of the Time-Dependent Hartree and the Locally Enhanced Sampling Approximations - Formal Properties, a Correction, and Computational Examples for Rare-Gas Clusters. *Journal of Chemical Physics* **1993,** 98, (4), 3380-3388.

24. Zheng, W. M.; Zheng, Q., An analytical derivation of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1997,** 106, (3), 1191-1194.

25. Chen, J. Tests of a new molecular dynamics method for enhanced conformational sampling. University of Oklahoma, Norman, OK, 2003.

26. Roitberg, A.; Elber, R., Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* **1991,** 95, (12), 9277-87.

27. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

28. Simmerling, C.; Miller, J. L.; Kollman, P. A., Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid. *Journal of the American Chemical Society* **1998,** 120, (29), 7149-7155.

29. Quillin, M. L.; Li, T.; Olson, J. S.; Phillips, G. N., Jr.; Dou, Y.; Ikeda-Saito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; et al., Structural and functional effects of apolar mutations of the distal valine in myoglobin. *Journal of Molecular Biology* **1995,** 245, (4), 416-36.

30. Ulitsky, A.; Elber, R., Application of the Locally Enhanced Sampling (Les) and a Mean-Field with a Binary Collision Correction (Cles) to the Simulation of Ar Diffusion and No Recombination in Myoglobin. *Journal of Physical Chemistry* **1994,** 98, (3), 1034-1043.

31. Simmerling, C.; Fox, T.; Kollman, P. A., Use of locally enhanced sampling in free energy calculations: Testing and application to the alpha ->beta anomerization of glucose. *Journal of the American Chemical Society* **1998,** 120, (23), 5771-5782.

32. Caflisch, A.; Miranker, A.; Karplus, M., Multiple Copy Simultaneous Search and Construction of Ligands in Binding-Sites - Application to Inhibitors of Hiv-1 Aspartic Proteinase. *Journal of Medicinal Chemistry* **1993,** 36, (15), 2142-2167.

33. Carlson, H. A.; Masukawa, K. M.; McCammon, J. A., Method for including the dynamic fluctuations of a protein in computer-aided drug design. *Journal of Physical Chemistry A* **1999,** 103, (49), 10213-10219.

34. Miranker, A.; Karplus, M., Functionality Maps of Binding-Sites - a Multiple Copy Simultaneous Search Method. *Proteins-Structure Function and Genetics* **1991,** 11, (1), 29-34.

35. Liboff, R. L., *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*. J. Wiley: New York, 1998.

36. Prigogine, I., *Non-equilibrium statistical mechanics*. Interscience: New York, 1962.

37. Martyna, G. J.; Klein, M. L.; Tuckerman, M., Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *Journal of Chemical Physics* **1992,** 97, (4), 2635-2643.

38. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant-Pressure Molecular-Dynamics Algorithms. *Journal of Chemical Physics* **1994,** 101, (5), 4177-4189.

39. Marco, M.; Vincenzo, P. P.; Gilberto, S.; Giorgio, T., Melatonin. *Current Medicinal Chemistry* **1999,** 6, (6), 501-518.

40. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.

41. Frisch, M. J.; al., e. *Gaussian 94*, Gaussian, Inc.: Pittsburgh, PA, 1995.

42. Hinsen, K., The molecular modeling toolkit: A new approach to molecular simulations. *Journal of Computational Chemistry* **2000,** 3, 79-85.

43. Palmer, B. J., Direct Application of Shake to the Velocity Verlet Algorithm. *Journal of Computational Physics* **1993,** 104, (2), 470-472.

CHAPTER 5

Mean-field molecular dynamics from a classical mechanical perspective

5.1 INTRODUCTION

Classical mechanical principles are important to learn, but are often underappreciated since many research problems in modern physics are more appropriately addressed by quantum mechanics. What follows is a description of a class of computational techniques used by chemists and physicists in modern research. In this contribution, we demonstrate how the concepts of the point transformation and holonomic constraint can be used to derive the equations of motion for a class of molecular dynamics based techniques, called multiple-copy, mean-field methods. First we introduce the field of molecular dynamics, and multiple-copy, mean-field simulation techniques in particular. Then we discuss point transformations and holonomic constraints and the particular techniques required to understand multiple-copy, mean-field molecular dynamics. Finally we discuss two published multiple-copy, mean-field methods and provide a simple example illustrating the two methods.

5.1.1 Background

Molecular dynamics[1] is a popular computational method designed to simulate the behavior of molecular systems by integrating the equations of motions of a model system obeying a classical, empirically parameterized potential energy function.[2]

This technique has formed the basis for a large number of applications which use the data generated by molecular dynamics simulations to extract secondary quantities for further analytical use. For example, molecular dynamics simulations have been used as the basis for estimating phase space averages including free energy differences[3] and radial distribution functions,[1] transport properties such as viscosity or diffusion coefficients,[4] and many other quantities. Molecular dynamics, particularly when coupled with simulated annealing,[5] is also the basis for geometry optimizations,[6, 7] where the search is performed along the physical path over which the system evolves. However, there are a variety of reasons that conventional molecular dynamics proves to be a poor choice for providing the underlying trajectory over which the calculation is performed, due to the properties of classical mechanics. For example, once a system reaches equilibrium, it only slowly explores phase space, which means that the great variety of configurations that the system can adopt is generally woefully undersampled.[8] Many techniques have been developed to address this issue. One class of such methods is called multiple-copy, mean-field methods.[9]

5.1.2 Multiple-copy, mean-field methods

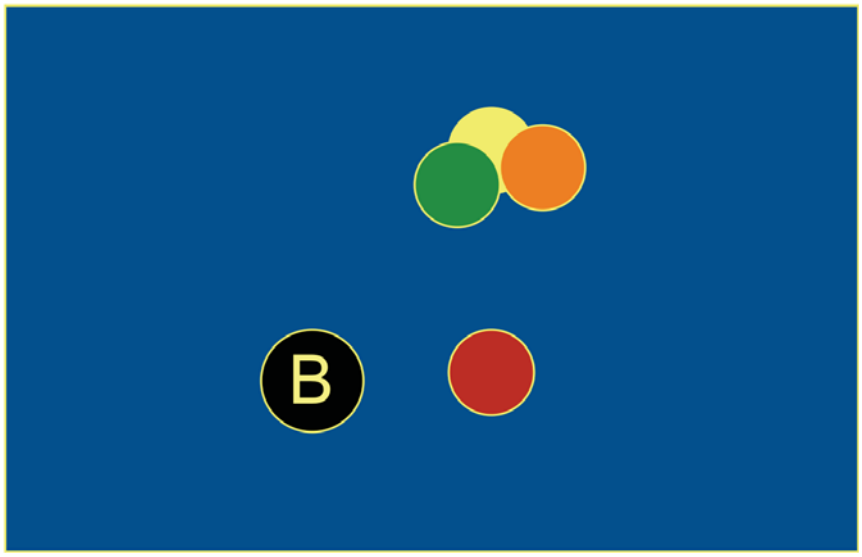
The multiple-copy, mean-field methods are related by a common way of constructing the Hamiltonian of the system to be simulated. This is accomplished by separating the system into two parts. First a bath is defined. This is the largest part of the system and contains the parts of the system that are required to exist, but are otherwise unimportant. The remainder of the system (which is composed of parts that are of most interest to the researcher) is copied some number of times. If C copies are made, then we can write the Hamiltonian of the system,

$$H_{total} = H_{bath} + \frac{1}{C}\left(\sum_{i=1}^{C} H_{copy}^{(i)} + H_{cross}^{(i)}\right), \qquad (5.1)$$

where H_{total} is the Hamiltonian of the entire system (which comprises all copies), H_{bath} is the Hamiltonian of the bath region of the system, $H^{(i)}_{copy}$ is the Hamiltonian of the ith copied part, and $H^{(i)}_{cross}$ covers the interaction between the bath and the ith copied part. The consequence of such a choice is that while each copied particle experiences the full force of each bath particle, the bath particles experience only an averaged force due to the copied particles.

Figure 5.1 illustrates how mean-field dynamics differs from conventional MD. The cluster of particles in the top right of the figure each exert and experience similar forces of interaction with the bath particle, because they are at similar distances. However, the red particle is much closer to the bath particle 'B,' so it exerts and experiences a much stronger force. Since the bath particle only experiences the average force exerted on it, it follows an altered trajectory compared to its trajectory in a classical MD simulation. This is so, because it experiences a much stronger force than the distant cluster would exert, but a much weaker force than the red particle ought to exert. This gives the red particle the ability to approach the bath particle much more closely than it would in a conventional MD simulation. Of course, as a consequence of such an approach, the dynamics is no longer "real," but this type of dynamics allows the system to explore a greater fraction of phase space than it would otherwise. This enhanced sampling has a significant disadvantage because it can significantly alter the ability of the simulation to provide accurate secondary quantities. For example, the radial distribution function is directly affected,[10] since

Figure 5.1 Depiction of an illustrative situation encountered in multiple-copy, mean-field simulations. The colored balls represent copied atoms, and the black ball containing the letter 'B' corresponds to a bath particle. Because the red particle is far from the others, its dynamics will be altered due to the way the forces are calculated in mean-field simulations.



the distribution of distances allowed over the course of the simulation is much broader. Since the radial distribution function is directly related to the partition function, all thermodynamics properties of the system are affected.

The method was originally explained in a variety of ways,[10] but can be simply explained by appealing to classical mechanics. A classical approach could construct a system where the entire structure is copied. Then, the coordinates of the bath particles are subjected to a point transformation.[11] Finally a holonomic constraint[11] is applied to selected variables in the new coordinate system. We have shown that selectively relaxing the constraints can dramatically improve the properties of the simulation, and have developed this idea into a new simulation technique.[12]

5.2 BRIEF REVIEW OF CLASSICAL MECHANICS CONCEPTS APPLIED TO MOLECULAR DYNAMICS

Before delving deeply into the task of describing this application of textbook concepts from classical mechanics, it is helpful to ensure that a common framework for understanding molecular dynamics is developed. In the following, ideas important to understand molecular dynamics are introduced, including the form of the force field and integration techniques. Then, a short discussion of point transformations and a description of the concept of the holonomic constraint follows.

5.2.1 Potential Energy Functions

Molecular potential energy functions generally have the following generic form[2]:

$$\begin{aligned} V(\vec{X}) = & \sum_{pairs}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{eq_iq_j}{4\pi\varepsilon_0 r_{ij}}\right) + \sum_{bonds} k_n(r_{ij} - r_n)^2 \\ & + \sum_{angles} k_n(\theta_{ijk} - \theta_n)^2 \\ & + \sum_{torsions}\sum_{terms} K_n(1 + \cos(m(\varphi_{nm} - \varphi_{0m}) - \gamma_m)). \end{aligned} \tag{5.2}$$

This function divides the potential energy into four contributing types of energy, each defined by a different summation term. The first sum, over pairs, is the non-bonded interaction and describes interactions between atoms separated by more than three chemical bonds. The first two terms of this part of the potential represent the (12,6)-Lennard Jones potential model for the van der Waals potential. The last term represents the Coulombic potential between atoms carrying fractional charges. Here, *r* represents the distance between the atom pair, the constants *A* and *B* (the Lennard Jones terms) are provided by the creator of the force field, while *q,* the partial charge

on each atom, must be determined (generally via a quantum mechanical calculation) for each new molecule to be simulated. The second sum, over bonded pairs of atoms, represents chemical bonds modeled using Hooke's law, where r_n is the equilibrium bond length and k is the spring constant. The third sum, over angles defined by atoms i, j, and k, uses Hooke's law, with θ_n representing the equilibrium angle. The last sum, over all torsion angles, is more complicated than the others. Each torsion angle can be represented by a sum over a Fourier-like expansion, with the number of terms and each constant provided by the creator of the force field.[2]

5.2.2 Integration techniques

Because the kinetic and potential energies are conjugate, we can write the system's Hamiltonian function:

$$H = \sum_{n=1}^{N} \frac{p_n^2}{2m_n} + V(\{x_n\}_{n=1}^{N}).$$

In the above expression, p_n is the momentum for the nth particle, m_n is the mass, and x_n is the position of the nth particle. The first term represents the kinetic energy of the system, which depends only on the momentum of each particle. The second is the function defined in Section 5.2.1, which depends only on the positions. Following Hamilton's rules, we can obtain the instantaneous force and momentum of each particle by differentiation. For example,

$$\dot{x}_n = \frac{\partial H}{\partial p_n}$$

is the velocity ($\dot{q}$ is the time derivative of the quantity q) of the nth particle, while

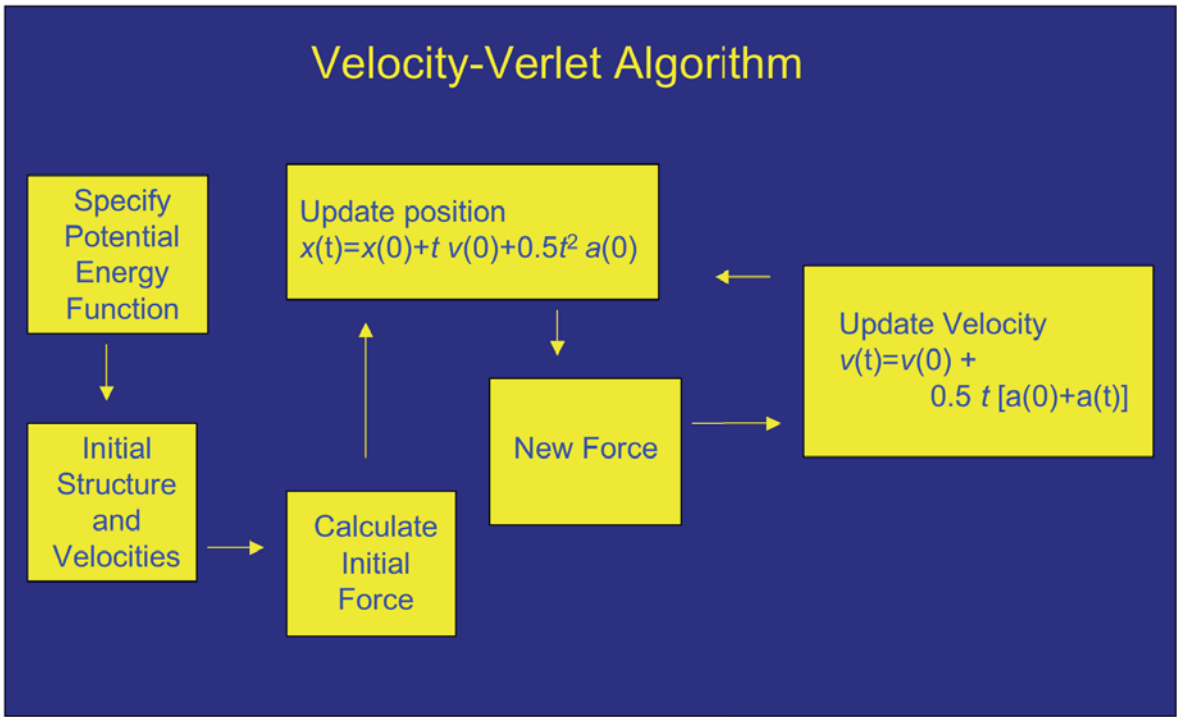
$$\dot{p}_n = -\frac{\partial H}{\partial x_n}$$

is the force on the *n*th particle.

Because exact solutions for the system of differential equations required to obtain the trajectory are not possible in such a many-body system, numerically integrating the equations of motion is generally performed. One of the most popular methods is called the velocity-Verlet algorithm, and it will be described here.[1]

Figure 5.2 contains a flow chart that illustrates the flow of a computer program implementing the velocity-Verlet algorithm. First, the initial positions and velocities must be somehow provided. The positions are often taken from an experimentally observed structure, or are contrived to test a hypothesis. Velocities are frequently assigned randomly according to the expected behavior of an ideal gas with a Maxwell-Boltzmann distribution of velocities corresponding to a selected average temperature and deviation in temperature. Once the initial state of the system is determined and the form of the potential energy function is selected, the initial force can be calculated. Now, the initial position, velocity, and force can be used to find the next set of positions for the system. The new positions are used to calculate a new force, which is then used to update the velocities. Now that the velocities and positions in the next time step are known, the cycle can repeat for as many steps as a user requires.

Figure 5.2 Flow chart illustrating the velocity-Verlet algorithm. This algorithm is used to integrate the equations of motion during a molecular dynamics simulation.



5.2.3 Point Transformations

Up to this point, we have treated the system using Cartesian coordinates implicitly. Of course, one of the advantages of using the Hamiltonian formalism is that it is consistent in any coordinate system. As long as a coordinate transformation is consistently applied, in the end, the dynamics will be independent of the coordinate system. One type of coordinate transformation that is of particular interest in explaining the results of this work is the point transformation. This type of coordinate transformation is quite general, and can be defined as any transformation that defines the new coordinates as functions solely of the old coordinates and the time. For example, for old coordinates $\{q_i\}_{i=1}^{N}$ and $\{p_i\}_{i=1}^{N}$, if the new coordinates can be written as functions of the old coordinates:

$$Q_i \equiv Q_i(q_1,...,p_N,t)$$
$$P_i \equiv P_i(q_1,...,p_N,t),$$

then a point transformation is defined. One common type of point transformation is the transformation from Cartesian coordinates to spherical coordinates.

One especially simple point transformation can be described by making the new coordinates linear combinations of the old coordinates. Such a transformation,

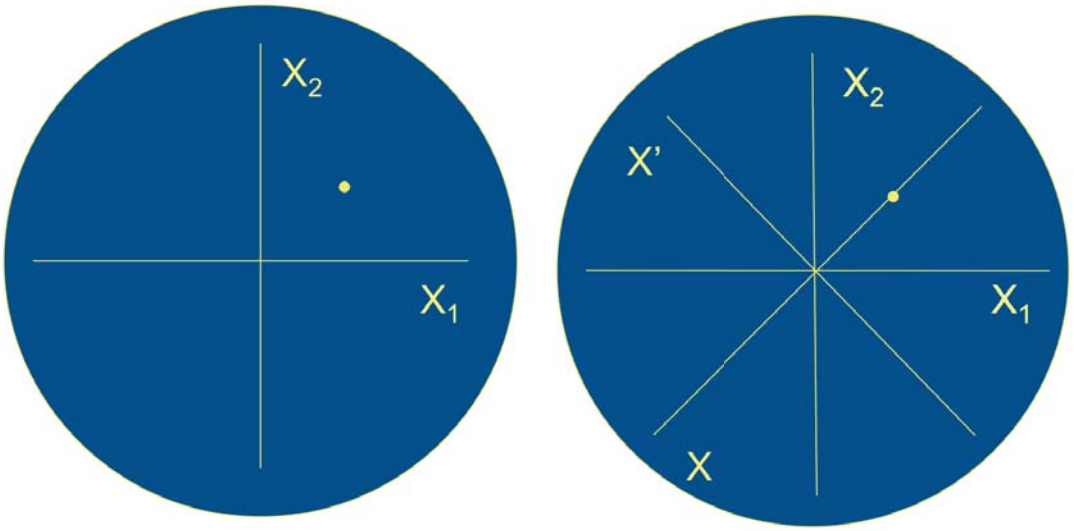
$$Q_i = \sum_{j=1}^{N} c_{i,j} q_i$$
$$P_i = \sum_{j=1}^{N} c_{i,j} p_i$$

can be completely defined by determining the transformation coefficients, $c_{i,j}$. This type of transformation is fundamentally important to understanding the classical mechanical derivation of multiple-copy, mean-field methods, and will be further discussed in the next section.

5.2.4 A Special Transformation

In the previous section, we introduced a type of point transformation defined by using linear combinations of the old coordinates to yield new coordinates. Such a transformation is entirely defined by its transformation coordinates and an infinite number of transformation coordinates are possible. Of the infinitely many combinations possible, which transformations are useful? One interesting problem to consider is illustrated in Figure 5.3.

Figure 5.3 a.) (left) Coordinate axis depicting the value of a coordinate for both copies of the system. If both copies' coordinate have the same value, then the point lies on the line of unit slope passing through the origin. b.) (right) Introducing two new coordinates, ***X*** and ***X'*** important to understanding multiple-copy, mean-field methods. If the point lies on the line described before, ***X'*** is zero.



First imagine some physical system. Inside this system is a coordinate which we designate *X*. Now, pretend that we make two copies of this entire system, and call the *X* coordinate of the first copy X_1 and the *X* coordinate of the second copy X_2. Now on the left in Figure 2.3 is a plot whose abscissa corresponds to the value of X_1 and whose ordinate corresponds to X_2. Since we made identical copies of the system, notice that the point is depicted as lying on the line of unit slope passing through the origin. Let us now define our goal in terms of identifying the coordinate system

where one coordinate, $\boldsymbol{X}$, represents the average of the coordinates in the original system, and $\boldsymbol{X}'$ is zero when the coordinates have the same value. This situation is depicted on the right in Figure 5.3. This simply corresponds to rotating the coordinate axes.

If we generalize, and rather than limit ourselves to two copies of the system, but instead allow some arbitrary C copies of the system, our transformation coordinates can be represented as elements of a CxC matrix and, in particular, we can define the transformation to have the following form:

$$Q_i = \tfrac{1}{C}\sum_{j=1}^{N} c_{i,j} q_i$$

$$P_i = \tfrac{1}{C}\sum_{j=1}^{N} c_{i,j} p_i .$$

We can set some limits on which possible values the elements of the matrix can take, but still be consistent with our goals described in the previous paragraph. For example, we could require that the matrix be symmetric. We definitely need the first row and column of the matrix to all be 1, because we want the coordinate corresponding to the first column to be the average value of the coordinate in all of the copies. The other terms are more loosely defined. Since we want to have an orthogonal basis, we want each column to be orthogonal and we want the length of each vector to be C. This corresponds to the following[13] mathematical constraints:

$$1 + \sum_{k=2}^{C} c_{n,k} c_{m,k} = C\delta_{n,m}$$

$$1 + \sum_{k=2}^{C} c_{n,k} = 0.$$

We can enforce these constraints while at the same time making a more tractable problem by making another simplification. If we choose to define the transformation coefficients by the following expression,

$$c_{1,k} = c_{k,1} = 1$$
$$c_{n,k} = y + \delta_{n,k}(x - y), \qquad \forall n,k > 1$$

i.e. by requiring that the first row and column are filled with 1's, the diagonal is filled with x's, and the rest is filled with y's then a solution can be determined algebraically.

For two copies, there is only one undetermined coefficient and its value is -1. For more copies than two, the following formula will provide the coefficients, which was determined by putting the above simplification into the expressions that define our constraint:

$$x = \frac{1 - c - \sqrt{c}}{1 + \sqrt{c}}$$
$$y = \frac{1}{1 + \sqrt{c}}.$$

Defining our coordinate transformation in this way ensures that one of the transformed variables corresponds to the average position across each of the copies while the rest remain zero as long as the positions of each copy are the same. We call the coordinate corresponding to the average a "major" variable. The other variables are denoted "minor".

5.2.5 Coordinate Transformations and Holonomic Constraints

Coordinate transformations in classical mechanics are useful in a variety of ways. Not only can a well planned coordinate scheme shed light on a problem, but, by expressing the dynamics in a natural coordinate scheme, intelligent use of constraints

can be made. One sort of constraint that might be imposed is a holonomic constraint. Holonomic constraints are functions of the coordinates of the system, but independent of time. One common example is a bond-length constraint. By fixing the length of a bond between two atoms, for example, the dynamics are altered to account for this.

Putting the concepts of the point transformation and holonomic constraint together, let us consider the consequences of applying a holonomic constraint to the "minor" variables generated by the coordinate transformation described in the previous section. If we require that the constraint be set to hold these "minor" variables at zero, it is apparent that the position of the constrained atoms would be required to remain the same across each of the copied systems. Selective application of such constraints is key to the implementation of multi-copy, mean-field methods.

5.3 SIMULTANEOUS APPLICATION OF COORDINATE TRANSFORMATION AND HOLONOMIC CONSTRAINT LEAD TO MULTIPLE-COPY, MEAN-FIELD EQUATIONS OF MOTION

We can derive the equations of motion for any classical system by first constructing its Hamiltonian function,

$$H(\{p_n\},\{P_n\},\{x_n\},\{X_n\}) = \sum_{n=1}^{N-B}\frac{p_n^2}{2m_n} + \sum_{n=1}^{B}\frac{P_n^2}{2m_n} + V(\{x_n\}_{n=1}^{N-B}) + V_{bath}(\{X_n\}_{n=1}^{B}) + V_{cross}(\{x_n\}_{n=1}^{N-B},\{X_n\}_{n=1}^{B}).$$

In this example, a basic system which includes bath coordinates (introduced in Section 5.1.2) in capital letters and the part of interest in lower case is presented. The first potential energy term contains coordinates for only the part of interest, the second term just the bath, and the final term is the interaction energy between the two

regions. N is the total number of degrees of freedom in the system, while B is the number of bath particles. We can now copy the system a number of times by first defining

$$H_i = H(\{p_{n,i}\},\{P_{n,i}\},\{x_{n,i}\},\{X_{n,i}\})$$

where the i subscript in the coordinates indicates that the coordinate belongs to the ith copy. Now we can construct the Hamiltonian of the system containing C copies:

$$\breve{H} = \sum_{i=1}^{C} H_i.$$

This is a perfectly correct Hamiltonian. We can make this a mean-field Hamiltonian, if two steps are taken. First, we apply the coordinate transformation described in Section 5.2.4 to all of the coordinates of the bath particles. Second, we holonomically constrain the "minor" variables to remain zero. Then the equations of motion for the "major" variables and those of the system of interest can be determined as described previously in this section. This can been seen in greater detail in Chapter 2. These equations of motion correspond to those of the previously published Locally Enhanced Sampling (LES) method.[4]

5.4 SELECTIVE USE OF CONSTRAINTS

As previously noted throughout this text, multiple-copy, mean-field methods in general are not well suited to collecting trajectories for calculating secondary properties. This is true because the trajectories produce a radial distribution function that contains broadened peaks relative to a conventional molecular dynamics simulation. So, at times, it might be desirable to maintain some of the enhanced sampling inherent in multiple-copy, mean-field simulations, but produce a trajectory

more capable of estimating secondary properties. This is the role that our algorithm "ensemble extracted from atomic coordinate transformations", the EXACT approximation, seeks to fill. The method works by recognizing that the previous section represents an accurate classical mechanical description of multiple-copy, mean-field methods. This understanding makes it plain that the holonomic constraints imposed on the "minor" variables constitutes the only difference between multiple-copy, mean-field and conventional molecular dynamics, and so selectively enforcing and relaxing such constraints is a logical method for interpolating between conventionally molecular dynamics methods and the LES approximation.

5.4.1 Relaxing the Constraints

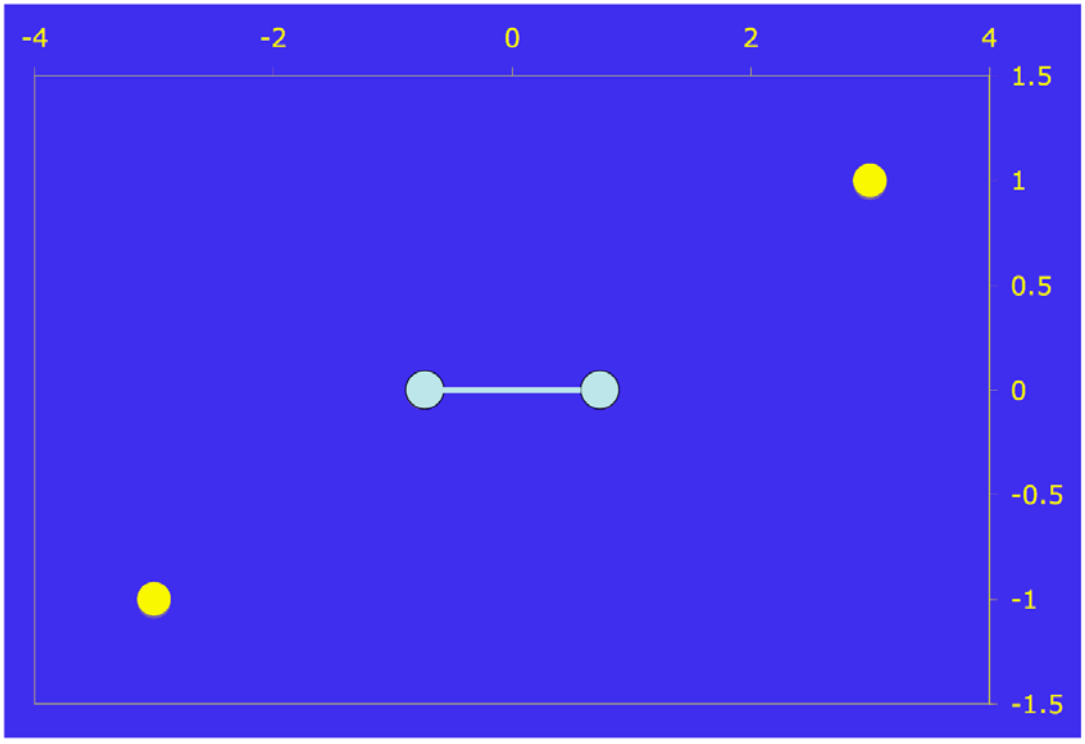
Some mechanism is necessary for determining when the constraints characteristic of multiple-copy, mean-field methods should be relaxed in order to improve the phase space averages produced by the trajectories. The method used in the EXACT approximation is as follows. At each step, the force exerted on the "minor" variables is calculated at each step. This is accomplished by calculating

$$\frac{\partial \breve{H}}{\partial Q'_{i,m}} = \sum_{k=1}^{C} c_{m,k} \frac{\partial V_{cross}}{\partial Q_i},$$

where $Q'_{i,m}$ is the "minor" variable of the *i*th degree of freedom in the *m*th copy and Q_i is the "major" variable of the *i*th degree of freedom. This was developed in detail previously.[12, 14] The calculated force is then compared with the force on the "major" variable as a ratio

$$\Phi_{i,l} = \left| \frac{P'_{i,l}}{P_i} \right|,$$

Figure 5.4 This is the arrangement of the initial positions (in unitless dimension) of the atoms in the two simulations described in Section 5.5 The yellow dots are free atoms and the blue dots (the bath) are atoms in a diatomic. Two trajectories were simulated. In the first, the yellow atoms were "fired" in such a way as to collide with the nearest blue atom. In the second trajectory, the yellow atoms were pushed to move along the y-axis toward the diatomic.



which is compared to a user-defined tolerance, Φ. If $\Phi_{I,l} > \Phi$ then the constraint is relaxed and the particle is removed from the bath. Of course, after the particle is removed from the bath, at some point, it makes sense to return it. This is accomplished rather simply: after a user-defined number of steps, the particle is returned to its average location, after checking that this is a safe place for it to be

returned. Of course, such a change breaks the flow of the dynamics. But its effect can be easily hidden in larger systems under the weight of the size of the averages. In smaller systems, sometimes it is preferable to skip the merging process altogether.

5.5 EXAMPLE TRAJECTORY

In the interests of illustrating the relationship between conventional MD, LES, and the EXACT approximation, we now present the results of two simple simulation trajectories performed on a system composed of four bodies. The simulation was performed using unitless variables. All four particles interact via the (12-6) Lennard-Jones potential (see the first term of Equation 5.2, $A=8x10^{-4}$, $B=4x10^{-5}$). Two of the particles had a mass of 20, while the other two had a mass of 16. The particles with mass 16 were joined together via Hooke's law, with an equilibrium distance of 1.5 and a spring constant of 0.1. The atoms were arranged as in Figure 5.4, and the trajectories were integrated with a time step of 0.1. The bonded atoms were started with no momentum, while the other atoms were given two separate trajectories (corresponding to the two simulations). In the first trajectory, the free atoms were each directed to collide with the nearest of the bonded atoms. In the second trajectory, the free atoms were directed along the y axis. Analogous simulations were performed using the LES and EXACT (threshold = 150) methods. In these simulations, the bonded atoms were defined as the bath, and the two copies corresponded to the two trajectories described above. Figure 5.5 depicts the results of

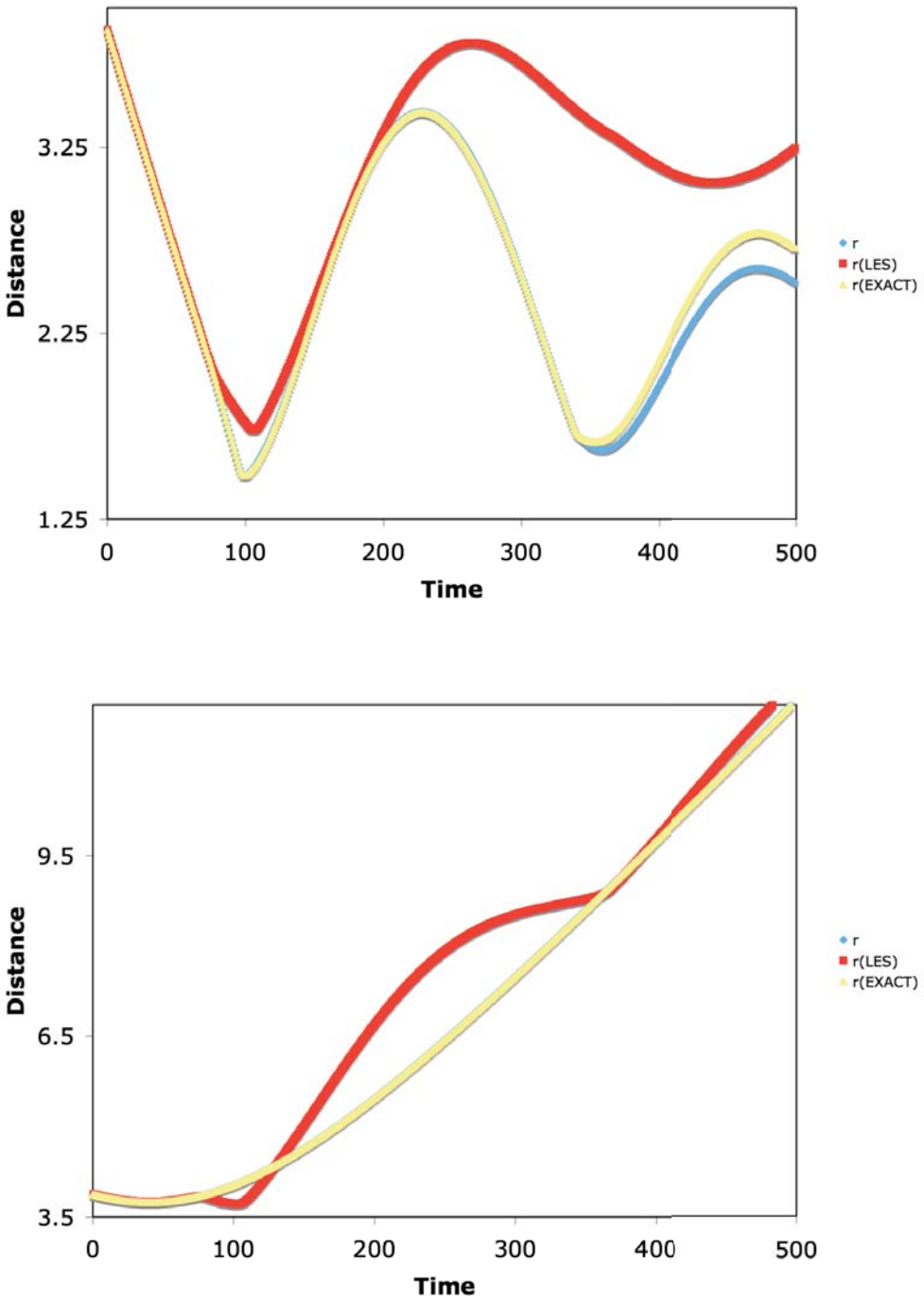


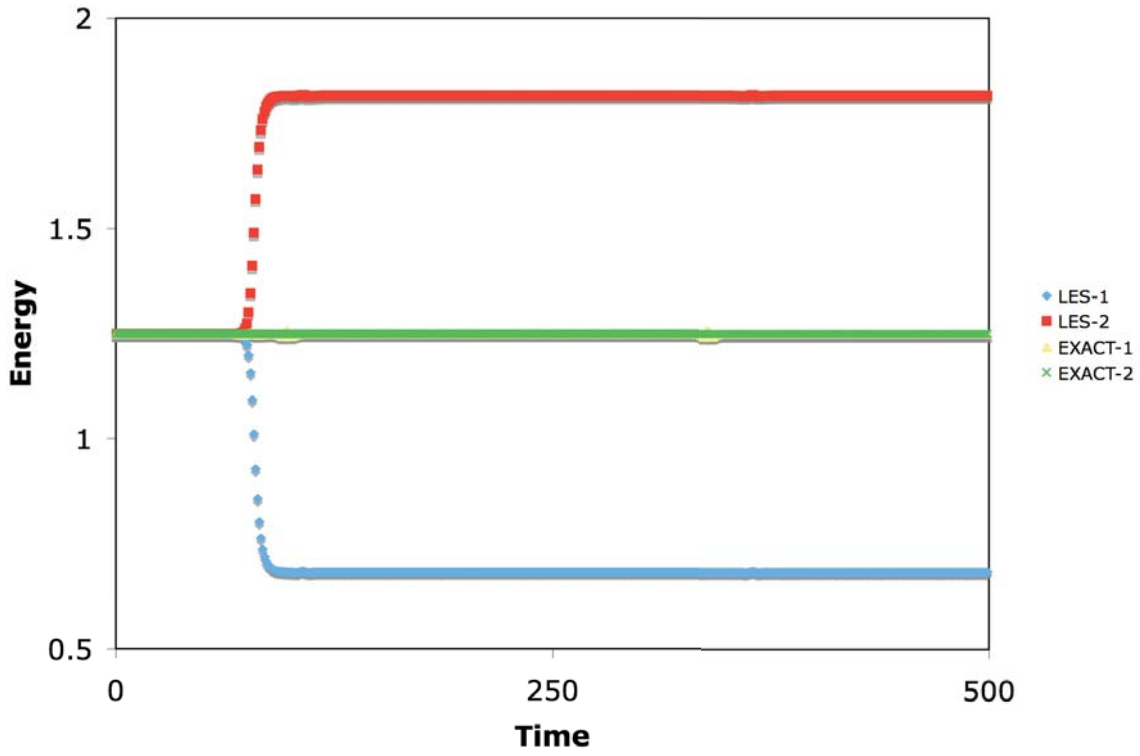
Figure 5.5 a.) (top) Distance (unitless) between colliding atoms in the first system described in Figure 5.4 using conventional MD, LES, and EXACT. The conventional and EXACT results are similar (but differ only near the end of the course) and the LES result differs. b.) (bottom) Distance between the same pair during the second trajectory, where no collision occurred. The LES result differs because the bath atom is affected by the first trajectory.

the simulations. In Figure 5.5a, the distance between one of the pairs of colliding atoms is plotted. There are clear differences between the trajectories generated using the different simulation techniques. The atoms in the conventional MD simulation collided after approximately 100 steps. The EXACT simulation's collision between the atoms was almost simultaneous with those of the MD simulation. In fact, the EXACT trajectory mirrored the conventional MD trajectory except near the end, differing only because of the accumulated effect of small differences. The LES simulation was starkly different. The collision was noticeably delayed, and the trajectory differs greatly from the MD trajectory. Looking at the second trajectory depicted in Figure 2.5b, we see that the conventional and EXACT trajectories are very similar, showing little interaction between the atoms. The LES trajectory is different only because the bath atoms are affected by the results from the first trajectory (represented by Figure 5.5a). The most notable difference between the LES results and the other simulations is shown in Figure 5.6. This plot shows the total energy of the system plotted with time. At the time of the collision, energy is removed from the first trajectory and put into the second. This effect dramatically demonstrates that the trajectories interact strongly, allowing simulations to explore areas they otherwise would not explore in conventional molecular dynamics simulations because of the effects of energy sharing between the copies. This point is illustrated more directly in previously published work.[10, 12, 14]

5.6 CONCLUSIONS

This contribution shows how point transformations may be used in the context of classical molecular dynamics computer simulations to derive equations of motion

Figure 5.6 Energy of the system during both trajectories for LES and EXACT simulations. The EXACT simulation maintains a nearly constant energy, but the LES simulation demonstrates a great degree of energy sharing between the two trajectories. This is a dramatic difference between the two techniques.



describing a collection of structures identical to the original system. Restricting the transformation vectors to be orthogonal leads to a unique, analytical solution for the transformation coefficients. When holonomic constraints are applied to a subset of the transformed coordinates, equations of motion for a new family of MD simulations—multiple-copy, mean field methods—emerge. An existing multiple-copy, mean field method, locally enhanced sampling (LES), represents one limit where bath particles of every structure are constrained to the same location throughout the simulation, whereas the opposite limit, where bath particles of each structure evolve independently of every other structure, represents independent MD simulations of

multiple copies. An algorithm called ensembles extracted from atomic coordinate transformations (the EXACT approximation) has also been developed that allows interpolation between the two limits. A simple numerical example shows that trajectories evolved under the EXACT approximation may closely approximate a conventional MD trajectory, whereas previous publications[10] show that the approximation can also retain the ability of LES to sample geometries that are inaccessible to conventional MD simulations. Thus, these results illustrate a use of several concepts from classical mechanics in a simple, yet practical research application.

5.7 REFERENCES

1. Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*. Clarendon Press: Oxford, 1987.

2. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.

3. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chemical Reviews* **1993,** 93, (7), 2395-2417.

4. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

5. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

6. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

7. Simmerling, C.; Miller, J. L.; Kollman, P. A., Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid. *Journal of the American Chemical Society* **1998,** 120, (29), 7149-7155.

8. Huang, Z. Design of new molecular dynamics global minimum search protocols for mapping energy landscapes and conformations of folded polypeptides and mini-proteins. University of Oklahoma, Norman, 2005.

9. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

10. Hixson, C. A.; Chen, J.; Huang, Z. N.; Wheeler, R. A., New perspectives on multiple-copy, mean-field molecular dynamics methods. *Journal of Molecular Graphics & Modelling* **2004,** 22, (5), 349-357.

11. Goldstein, H., *Classical Mechanics*. Addison-Wesley: Reading, MA, 1950.

12. Hixson, C. A.; Wheeler, R. A., Practical multiple-copy methods for sampling classical statistical mechanical ensembles. *Chemical Physics Letters* **2004,** 386, 330-335.

13. Zheng, W. M.; Zheng, Q., An analytical derivation of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1997,** 106, (3), 1191-1194.

14. Hixson, C. A.; Wheeler, R. A., Rigorous classical-mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles. *Physical Review E* **2001,** 64, 026701.

CHAPTER 6

Evaluating the role of varying pressure in finding energy minima during simulated annealing of polymer models

6.1 INTRODUCTION

Polymer systems are generally difficult candidates for computational study, because their multiple conformations are responsible for a high degree of difficulty in finding low energy structures. This manifests as a glassy potential energy surface which complicates application of conventional computational methods[1] typically employed in studying the structural and thermodynamic properties of these systems, such as molecular dynamics (MD)or *Monte Carlo* (MC).[2] Glassy potential energy surfaces are characterized by having a great number of accessible low energy states. Generally MD and MC are supplemented by techniques such as simulated annealing to improve the quality of the results.[3-9] This is especially true in structural studies, particularly when searching for low energy states of such systems.[6-8] Alternative methods are available and are commonly used,[10, 11] since no one class of optimization methods is suited to all problems. Presented here is a technique that modifies simulated annealing by adding pressure as an additional control parameter. This extra parameter is tested to assess its utility in the special case of frustrated, glassy polymer systems. Frustrated systems are characterized by having several conditions that are impossible to simultaneously fulfill. First we review optimization techniques employed in the study of complicated systems, such as the polymer case we study

here. Then we explain the advantages of the simulated annealing approach, and show how the addition of pressure as a parameter modifies the method. Finally, we present results that demonstrate the value of our proposed approach.

6.1.1 Popular Molecular Dynamics Based Optimization Methods

One of the most important pieces of information useful to understanding chemical phenomena is structure. Understanding the spatial arrangement of the various atoms contained in a chemical system may seem rather basic, but it is necessary to provide insights into the reactivity, thermodynamics, and spectroscopic behavior of the system. Of course, determining the structure of chemical systems can be accomplished experimentally using X-ray crystallography,[12] NMR,[13] and other methods.[14] Quite commonly now, chemical structures are determined computationally. Using the quantum approach,[15] structures of small or even medium sized molecules are determined routinely to great accuracy. For larger systems, which includes condensed phase systems, accuracy is traded for computational efficiency and molecular mechanics methods are employed.[16] Molecular mechanics methods, for example, take the AMBER forcefield[17] and use optimization methods traditionally used in mathematics, such as the conjugate gradient algorithm,[18] to find minima of the energy function. This is a useful approach, but is of limited utility for a variety of reasons, including that the search is generally limited to finding local minima "near" the starting structure. More commonly used is molecular dynamics or *Monte Carlo*,[2] coupled with some optimization algorithm. Some popular algorithms, including simulated annealing,[4] a popular algorithm fundamentally important to this contribution, are discussed below.

6.1.1.1 Locally Enhanced Sampling

Locally enhanced sampling (LES)[19] is a technique first proposed in 1990 (the original idea was taken from another method),[20] designed initially to improve the search for diffusion paths of small ligands inside a protein matrix. It has since become a fairly popular optimization method, useful when the structure of a small part of the system is needed in relation to the remainder. This small part of the system is copied several times. None of the copies directly interact with each other, but instead directly interact only with the remainder of the system, generally referred to as the bath. Atoms in the bath interact normally with each other, but interact with the average of each of the copied systems. It is often claimed that this algorithm allowed the interaction between the copied parts of the system and the bath to be smoothed. Specifically, it has been conveyed that any barriers resulting from this interaction would be decreased proportionally to the number of times the smaller part of the system was copied.[21] We do not feel this to be strictly true. Based on the results of Chapter 2, it is more likely that LES allows a system to surmount barriers more easily because the energy of individual systems is not conserved. Rather, in the LES approximation, the combined energy of the totality of the systems is conserved. Individual systems frequently possess more energy than they started with, allowing a sort of "tunneling" behavior. Each copy can be given a greater allotment of energy than in traditional molecular dynamics, and so can better explore the potential energy surface, if only until the energy is redistributed to the other systems via the bath interaction.

The method has been used to explore potential energy surfaces in several situations.[11, 21-25] The trajectory mapping application used in the original paper is an important example. There have been other applications of LES to finding low energy structures.[11, 21-25] Although not as commonly used as the method described in the next section, LES is still a relatively popular method and has been implemented in a variety of molecular dynamics packages in common use.

6.1.1.2 Replica Exchange

The replica exchange molecular dynamics simulation method is quite popular. First proposed in 1999,[10] the method has a relatively straightforward implementation. The entire system under study is replicated a number of times. Each replica is independent, but is held at a different temperature, spanning a range of temperatures. The systems are allowed to evolve using molecular dynamics for a set number of time steps, and then the temperatures of two systems are exchanged subject to a test: a Boltzmann-like probability factor based on the two systems' temperatures and energies. Since the probability of acceptance is low if the temperatures vary too much, only exchanges between systems of neighboring temperature are attempted.

The method has shown great utility in mapping low energy structures of peptide systems. Generally described as finding free-energy surface minima, replica exchange has been used to study a variety of systems, including polypeptides, proteins, and polymers.[10, 26-31] Many other examples can be found in the literature. This method owes a large part of its existence to simulated annealing, as the methods

are quite similar in approach, though simulated annealing has been in use for a much longer time.

6.1.2 Simulated Annealing

The idea that a state possessing a given energy is populated with a calculatable probability is one of the most fundamental ideas in statistical physics and is immortalized (at least for the canonical ensemble) in the Boltzmann distribution. Given this link between such a readily obtainable quantity (the energy) and population it should come as no surprise that an optimization strategy, simulated annealing, could be formed from it. Simply put, simulated annealing[4] is the process where an ensemble of states of a system are generated corresponding to high-energy conditions (via raising the temperature) followed by allowing the system to evolve toward a lower-energy state. As cooler states are generated, the system tends to settle into areas of low energy, because at cooler temperatures these areas are more likely to be populated. Raising the temperature initially widens the search area by increasing the volume of phase-space available to be populated because the systems have a greater probability of surmounting barriers separating regions of phase-space. Then, lowering the temperature traps the system in wells with a probability that depends on their energy and phase-space volume.

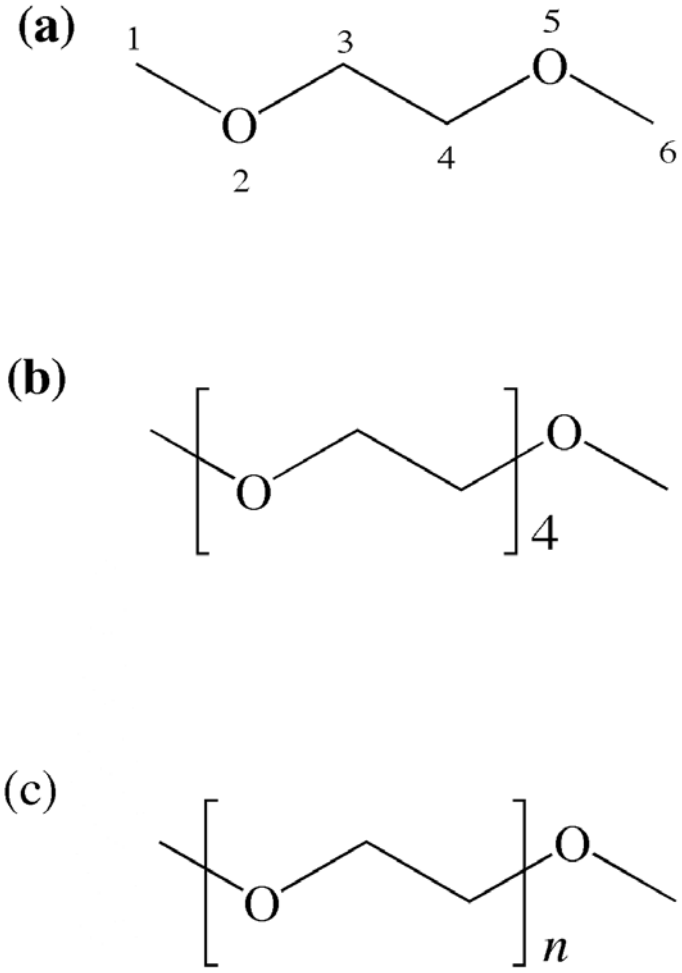
One surprising aspect of simulated annealing is that despite its specific link to statistical physics, it has shown its worth in a variety of fields as a general-purpose optimization technique.[4] Such general use can be contemplated because there are a wide variety of problems that allow a cost function analogous to the energy to be

defined. In one famous example, the traveling salesman problem, the object is to find the quickest path connecting two points along some complicated linkage structure sharing features found in road maps. The energy is represented by a cost function generally chosen to be a function of the number of nodes the salesman must cross. This function is optimized, generally, using a *Monte Carlo* procedure where each newly generated state is compared with the previous one and either accepted or rejected according to the Metropolis criterion.[2] The temperature is a fictitious parameter, but by increasing it, the algorithm accepts trial moves with a greater frequency allowing higher "energy" states to be visited. Lowering the temperature biases the simulation into lower "energy" states. This behavior is exactly analogous to the role of temperature in condensed phase systems. Similar methods have been used to solve problems in electrical engineering to design circuits, and in signal processing to process images and sounds.[4] Finally, the most traditional area of application is in finding energy minima in condensed systems either using the Metropolis/*Monte Carlo* method outlined above, or by using molecular dynamics as the ensemble generation engine.[2]

6.1.3 Role of Pressure

Complimenting simulated annealing with pressure annealing to improve the speed which low-energy geometries can be located has the potential to improve further the quality of simulated annealing searches. As described above, one role that temperature plays in a simulated annealing optimization calculation is in (a least partially) defining the phase-space area within which the search is to be conducted. If the system is located in an area of phase space separated cleanly from another by a

Figure 6.1. Structures of the polymer models used in this work are shown. These structures include a.) monoglyme, b.) tetraglyme, and c.) the polymer polyethylene oxide.



large barrier, the separated area will only rarely be considered within the calculation. This situation requires a particularly long simulation to obtain an accurate result. Raising the temperature ever higher, it becomes more likely to surmount such a barrier. Practically, however, raising the temperature can become counterproductive by causing the simulation to become unstable or increasing the length of simulation time. Keeping this in mind, making more sensible changes to other simulation variables is an avenue worth exploring. For example, changing other variables which allow a greater area of phase space to be considered within the simulated annealing

calculation would be desirable. One such variable is the pressure. Unlike temperature, pressure is not directly related to the energy (at least as directly as in the Boltzmann distribution). Adjusting the pressure indirectly affects the optimization, however, because the pressure determines the volume (and the density) of the system, which in turn affects the amount of phase space available to be explored. It increases the fraction of explorable phase space using a different approach than using temperature alone. The effects of such changes are explored below.

6.1.4 Objectives of This Work

Several systems are tested here to illustrate the benefits that using pressure as an optimization control parameter can add to a simulated annealing optimization strategy. First the method was tested on a system composed of Lennard-Jones particles.[32] This system is uncomplicated, and is used here are a control. Also tested were two models for the polymer polyethylene oxide (PEO).[33-35] Monoglyme is an oligomer containing one unit of the PEO polymer, and tetraglyme contains four units (see Figure 5.1). These systems have been previously studied,[33-35] and PEO is interesting because of its potential use as the matrix of polymer-ion batteries,[36, 37] among other applications. Polymers and polymer models are expected to receive quite a bit of assistance from pressure in the simulated annealing context because the glassy nature of their energy surfaces makes conformational trapping a serious issue.[1]

6.2 THEORY OF "PRESSURE ANNEALING"

Simulated annealing is among the most used optimization methods employed in the field of chemistry for a variety of reasons. First, the method makes a great deal of intuitive sense and can be explained almost entirely using basic appeals to logic.

Thus, the fact that it works is something that almost anyone with a basic grasp of chemistry can understand. On the other hand, simulated annealing has actually been formally proven to be an optimization technique, and is on a firm mathematical basis.[5, 38] It has even been shown that when using a proper cooling schedule the method is guaranteed to find the global minimum of the system.[38] Of course, even though the conditions required to realize completely the promise of simulating annealing are impossible to achieve in practice, the fact that the method could in principle attain success is quite appealing. Perhaps this combination of being easily understood, effective, and firmly rooted in theory explains why simulated annealing is one of the most popular optimization methods in use, not only in chemistry, but also in a wide variety of other fields. What follows is an abridged summary of simulated annealing, along with how a theoretical grasp of simulated annealing helps to understand the pressure-temperature annealing procedure proposed in this work.

6.2.1 Simulated Annealing

A simple, qualitative way of explaining why simulated annealing works starts by noting that low temperature states corresponding to low energy should be populated with greater probability than higher energy ones, while at higher temperatures the populations are less strictly related to the energy of the state, because of the flattening of the probability distribution observed at higher temperatures. In the high temperature phase of the optimization, the system wanders and explores phase space with relatively lax restrictions, and freezes into the states of higher probability at lower temperatures. Repeated application of heating cooling cycles is guaranteed to find the lowest energy states possible, eventually, a fact which has been proven,[5, 38]

but whose explanation is beyond the scope of this contribution. On the other hand, it is instructive to note that simulated annealing is a method that finds the minimum of E by manipulating β in the following expression:

$$\mathrm{Q}(N,V,T) = \frac{1}{C}\int_{-\infty}^{\infty} e^{-\beta E}\Omega(N,V,E)dE$$

where T is the temperature (β is the inverse of the product between the temperature and the Boltzmann constant), V is the volume, C is a normalization constant, and N is the number of particles. Q, is the canonical partition function; Ω, the microcanonical partition function; and E is the energy. In this framework, simulated annealing can be interpreted as finding the lowest energy for a system containing N particles in a volume, V.

6.2.2 Pressure Annealing

Using pressure as a variable in an optimization problem at first glance may not seem significant. However, this procedure can be interpreted to produce a well-defined result. Looking at the canonical partition function above, we can write a similar expression for a related ensemble. Since knowing that in the microcannonical ensemble, the pressure divided by the temperature is conjugate to the volume,[39] we know that if βp (where, p is the pressure of the system) is fixed allowing the volume to fluctuate, we can write a partition function for that ensemble as:

$$M(N,\beta p,T) = \frac{1}{C}\int_{-\infty}^{\infty} e^{-\beta pV}\Omega(N,V,E)dE\,.$$

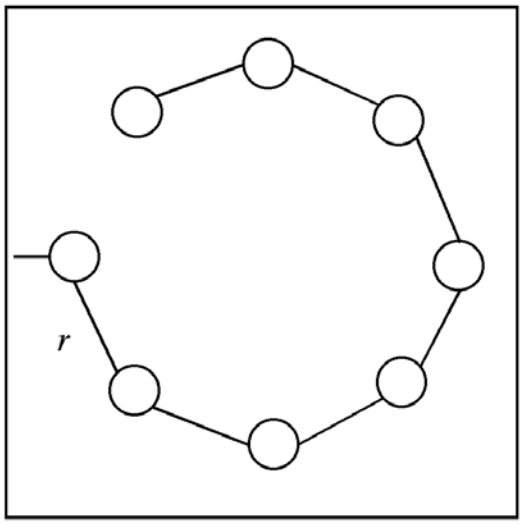
All the variables are the same as described in the previous section, and M is the partition function for the NPE ensemble. This expression for the partition function is reasonable since $\Omega \sim V^N$ is weighted by $e^{-\beta pV}$ as V gets large. It is also apparent that

as the pressure increases, the volume decreases. Conversely, decreasing the volume increases pressure. When the pressure vanishes, the volume becomes unbounded. By manipulating βp as the temperature was manipulated in the simulated annealing example, the function V is minimized. This can be interpreted as finding the smallest volume able to contain N particles with an energy, E.

6.2.3 A Benefit to Including Pressure In Simulated Annealing Optimizations?

In the isothermal-isobaric ensemble (NPT), the partition function contains both the pressure and the temperature as fixed variables. Using this knowledge, we constructed an optimization strategy designed to take advantage of pressure as a parameter alongside temperature in a pressure-temperature simulated annealing strategy. From the analysis in the previous two sections, it is clear that using pressure as an optimization parameter does not act to optimize the energy. What is true, though, is that the pressure does control features of a simulation that can be exploited to aid in simulated annealing optimization calculations. By fixing the pressure to a small value, the density of the system decreases, causing the volume of the system to increase. For some systems, this fact is not particularly helpful from the perspective of designing an optimization strategy. However, for bulky systems that are hindered at the density of interest, increasing the available volume can be quite helpful. By allowing the molecules to separate, intramolecular energies can be minimized more easily. Then, by increasing the pressure, the molecules can be efficiently compressed into a low energy state optimizing the volume, using the procedure explained in the previous section. Using these principles, a pressure-temperature simulated annealing optimization strategy can be designed consisting of four steps. The first step takes the

Figure 6.2. A simple two dimensional model representing a coiled molecule. The molecule exists inside a box whose walls are treated as large potential energy barriers, each side of which has length *L*. The molecule is composed of beads connected by springs, whose equilibrium length is *r*. When the length of the side of the box is greater than 7*r*, the molecule can change coil orientation without need to surmount any barrier, illustrating that increasing the available volume can make conformational searches easier.



initial state and allows an expansion to occur by fixing the temperature and pressure at low values. After the expansion, the volume is fixed, and traditional simulated annealing is performed on the expanded system. After simulated annealing, the pressure is fixed to a large value, as the temperature remains fixed at the final simulated annealing temperature. Finally, after the compression, the system's energy is minimized. The totality of the method allows two complications of simulated annealing to be addressed. First, in condensed phase systems of molecules with hindered rotations, inducing intramolecular conformational changes can be difficult.

Also, for the similar reasons, changing the way the system packs can be a challenge as well.

Understanding how lowered density might benefit a simulation is not difficult. Imagine the simple two dimensional system depicted in Figure 6.2. This system consists of a string of eight beads connected to each other in a coil by springs of length *r*. The beads do not interact except via the springs, and can be considered hard spheres. The entire coil is contained in a box, with side of length *L*. If an optimization algorithm seeks to characterize the low energy states of this system, moving from the counterclockwise coil depicted to the clockwise coil would be an important transition. The height of the lowest barrier between these two states determines the difficult of making the transition. If the length of the box is 8*r* or greater, no barrier to the transition exists because the system can align as a straight line while maintaining each of the springs at its equilibrium length. If the box is smaller than the radius of the semicircle, the system cannot even exist in the low energy state. Systems complicated enough to be of chemical interest not only must account for this, but also packing arrangements. Both are dealt with by the pressure annealing procedure.

6.3 MODEL SYSTEMS AND PROCEDURE

Several systems were used to compare pressure annealing with simulated (temperature) annealing. These systems range from a relatively simple model Lennard-Jones system to oligomeric models of polyethylene oxide. In each case, two sets of simulations were performed. First, traditional simulated annealing was

performed. In these simulations the system was elevated in temperature starting from 25 K to 1025 K over the course of a simulation on the nanosecond time scale. After the simulated annealing simulation, the system was further minimized using a conjugate gradient method. The minimized structure and energy were reported. Secondly, pressure-temperature annealing was performed. These simulations can be divided into three distinct steps. The first step is heating while allowing the system to expand slowly. The second phase allows the system to expand dramatically while maintaining the hottest temperature of the annealing cycle. Finally the system was compressed to its original volume while being cooled. The system was then subjected to the same minimization procedure as in the simulated annealing simulation and the final minimized structure and energy were reported.

6.3.1 Particles with Lennard Jones interactions only

This model system contained particles of argon with a mass of 40 amu. Here, the force between the particles is the (12, 6)-Lennard Jones model for van der Waals forces which can be written as

$$V(r) = 4\varepsilon\left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}}\right).$$

The parameters for argon were taken from the literature.[40] The simulation was performed using NAMD,[41] with a time step of 1 fs using a non-bonded cutoff of 12 Å. Each simulation involved 540 particles which were randomly placed inside the box, and then minimized for 5000 conjugate gradient minimization steps. For the simulated annealing calculations the volume was fixed at 30 Å per side in a cubic box. The temperature ranged from 25 K to 1025 K in 100 K increments with 50 picosecond simulations at each temperature. After reaching the highest temperature,

the simulation was run for 200 ps, and then cooled by reversing the heating schedule. After the system was returned to 25 K, a final equilibration was performed for 50 ps. The final structure and energy was obtained after a 5000 step conjugate gradient minimization. The pressure-temperature annealing simulations started from the same configuration as the corresponding simulated annealing run. It differed in the following ways. During the heating phase, the pressure was fixed at 0.1 bar, and the system was allowed to expand. During the first 50 ps the system was held at 1025 K and the pressure remained at 0.1 bar. During the next 100 ps at 1025 K, the system was allowed to expand dramatically, as the pressure was reduced to 0.025 bar. By the end of this period, the box generally expanded to approximately 40 Å on each side. During the final 50 ps the system was held at 1025 K. During the rest of the cooling phase, the pressure was increased to as much as 1000 bar, until the system returned to its original volume, at which time the volume was fixed. The final 50 ps equilibration and 5000 step minimization was then performed, and the final energy and structure was recorded.

6.3.2 Monoglyme

Monoglyme (Figure 6.1) is the smallest oligomeric analogue for the polymer polyethylene oxide (PEO), and contains a single unit of the glyme repeat unit, capped with a methyl group. The simulations were performed using NAMD and a forcefield previously derived for PEO analogues.[33] Each of the simulations were performed after placing the 120 molecules in a random orientation, and then placing them randomly in the box ensuring that no two atoms were closer than 1.9 Å apart. Again, the traditional simulated annealing runs were performed at a fixed volume in a cubic

box having sides of length 26.25 Å. The procedure was exactly as described above for argon, with the only differences described here. During the heating phase of the pressure-temperature annealing procedure, the pressure was set for 0.05 bar, but lowered to 0.025 bar to expand the system. The final box size was approximately 110 Å. The system was compressed during the cooling phase with a pressure of as much as 400 bar until the original volume was restored, though as the volume neared its original size, the pressure was slowly reduced to avoid overcompressing.

6.3.3 Tetraglyme

Tetraglyme (Figure 6.1) is an oligomeric analogue for the polymer polyethylene oxide (PEO), containing a four repeat units. The simulations were performed using NAMD and the same force field used for monoglyme and intended for PEO analogues.[33] Each of the simulations were performed after placing the 50 molecules in a random orientation, and then placing them randomly in the box ensuring that no two atoms were closer than 1.9 Å. Again, the traditional simulated annealing runs were performed at a fixed volume in a cubic box having sides of length 26.5 Å. The procedure was exactly as described above, with the only differences described here. During the heating phase of the pressure-temperature annealing procedure, the pressure was set for 0.1 bar, but lowered to 0.025 bar to expand the system. The final box size was approximately 100 Å. The system was compressed during the cooling phase with a pressure of 500 bar until the original volume was restored, though as the volume neared its original size, the pressure was slowly reduced to avoid overcompressing.

6.3.4 Procedure

For each of the systems described above, a number of random structures were generated. For the Lennard-Jones system, fifty random configurations were generated. For the polymer models, twenty-five configurations were generated. For each of these random structures, both simulated annealing and pressure annealing simulations were performed, and the final energies and structures were recorded.

6.4 RESULTS AND DISCUSSION

The data collected in the simulations described in the previous section demonstrate that the use of pressure as a coordinate in optimization simulations has some demonstrable effect on the quality of the structures obtained. For 52% of the Lennard-Jones simulations, 84% of the monoglyme simulations, and 68% of the tetraglyme simulations, the energy of the state produced by the pressure annealing procedure was lower than that produced from the same starting structure but by following the traditional simulated annealing procedure. Each of the final states was characterized by a variety of structural measurements including radial distribution functions, radii of gyration, mean square end-to-end differences, and characteristic ratios. Torsion triad analysis was also generated for the polymer model systems. Each of the structures produced structural data consistent with published[33-35] results. Following is a more detailed look at the data, along with some discussion of its significance.

6.4.1 Energetic Results

The final energies from each simulation started from a random structure are listed in Tables 6.1-6.3. They reveal two interesting features. First, the fifty simulations

Table 6.1. Final energies determined by simulated annealing and pressure annealing algorithms for each of the fifty Lennard-Jones systems. Each pair of simulations started from the same randomly generated initial structure. Neither method seems to be superior to the other, as lowest energy was found with equal likelihood by both methods.

Final Energy (in kcal mol-1) for Lennard-Jones systems

	Simulated Annealing	*Pressure Annealing*		*Simulated Annealing*	*Pressure Annealing*
1	-953.54	-959.72	26	-955.89	-954.60
2	-955.53	-958.15	27	-954.38	-957.08
3	-951.82	-941.17	28	-957.22	-957.02
4	-963.88	-958.96	29	-955.10	-964.38
5	-959.86	-957.37	30	-953.17	-954.55
6	-955.72	-962.54	31	-953.00	-954.04
7	-958.82	-960.63	32	-953.45	-953.38
8	-958.23	-952.44	33	-942.91	-955.55
9	-957.33	-957.85	34	-956.65	-949.70
10	-952.84	-955.26	35	-955.83	-955.21
11	-960.52	-958.35	36	-957.34	-959.68
12	-944.61	-951.79	37	-960.03	-959.50
13	-955.15	-957.99	38	-956.19	-948.28
14	-960.71	-947.03	39	-957.45	-960.63
15	-956.07	-951.94	40	-961.74	-956.73
16	-955.63	-960.72	41	-952.40	-962.08
17	-960.70	-953.16	42	-953.11	-961.18
18	-958.62	-956.65	43	-958.97	-960.29
19	-956.99	-957.29	44	-959.52	-961.30
20	-959.05	-953.37	45	-959.73	-959.84
21	-959.97	-956.79	46	-952.60	-958.94
22	-957.59	-955.11	47	-945.38	-958.56
23	-957.30	-954.24	48	-953.92	-963.10
24	-960.38	-945.11	49	-953.77	-956.70
25	-958.37	-952.55	50	-958.79	-955.04

Table 6.2. Final energies determined by simulated annealing and pressure annealing algorithms for each of the twenty-five monoglyme systems. Each pair of simulations started from the same randomly generated initial structure. The pressure annealing algorithm preferentially gave the lowest energy approximately 80% of the time.

Final energies (in kcal mol^{-1}) for monoglyme systems		
	Simulated annealing	**Pressure Annealing**
1	-734.8209	-741.6008
2	-722.3525	-741.6008
3	-721.0294	-741.6008
4	-735.0073	-741.6008
5	-737.5804	-741.6008
6	-729.8086	-741.6008
7	-728.0307	-741.6008
8	-741.7752	-741.6008
9	-727.7178	-741.6008
10	-730.5577	-741.6008
11	-729.5520	-736.1320
12	-729.5520	-737.8802
13	-729.5520	-738.9512
14	-729.5520	-740.6107
15	-729.5520	-725.5618
16	-729.5520	-734.7025
17	-729.5520	-717.1497
18	-729.5520	-732.1472
19	-729.5520	-733.4708
20	-729.5520	-741.2944
21	-744.3963	-715.6028
22	-737.9864	-751.5224
23	-719.8457	-742.0861
24	-734.7466	-737.3870
25	-728.8936	-733.7409

Table 6.3. Final energies determined by simulated annealing and pressure annealing algorithms for each of the twenty-five tetraglyme systems. Each pair of simulations started from the same randomly generated initial structure. The pressure annealing algorithm preferentially gave the lowest energy approximately 70% of the time.

Final energies (in kcal mol^{-1}) for tetraglyme systems

	Simulated Annealing	Pressure Annealing
1	-105.0748	-103.9599
2	-98.6349	-99.8905
3	-100.5817	-107.2664
4	-105.0766	-105.9834
5	-106.7726	-98.3929
6	-102.2849	-101.1396
7	-114.2732	-102.1396
8	-107.0713	-107.6627
9	-105.9728	-107.8258
10	-98.4871	-100.2481
11	-81.9171	-91.4357
12	-118.1657	-107.4743
13	-91.2899	-98.9836
14	-96.3824	-124.9147
15	-83.3597	-110.0198
16	-100.6137	-116.6695
17	-110.3598	-101.1773
18	-87.3115	-108.4409
19	-101.0185	-113.5207
20	-103.8985	-109.7915
21	-94.2803	-108.7402
22	-89.5132	-94.8436
23	-108.8482	-100.1544
24	-87.5746	-96.8615
25	-97.6049	-94.4017

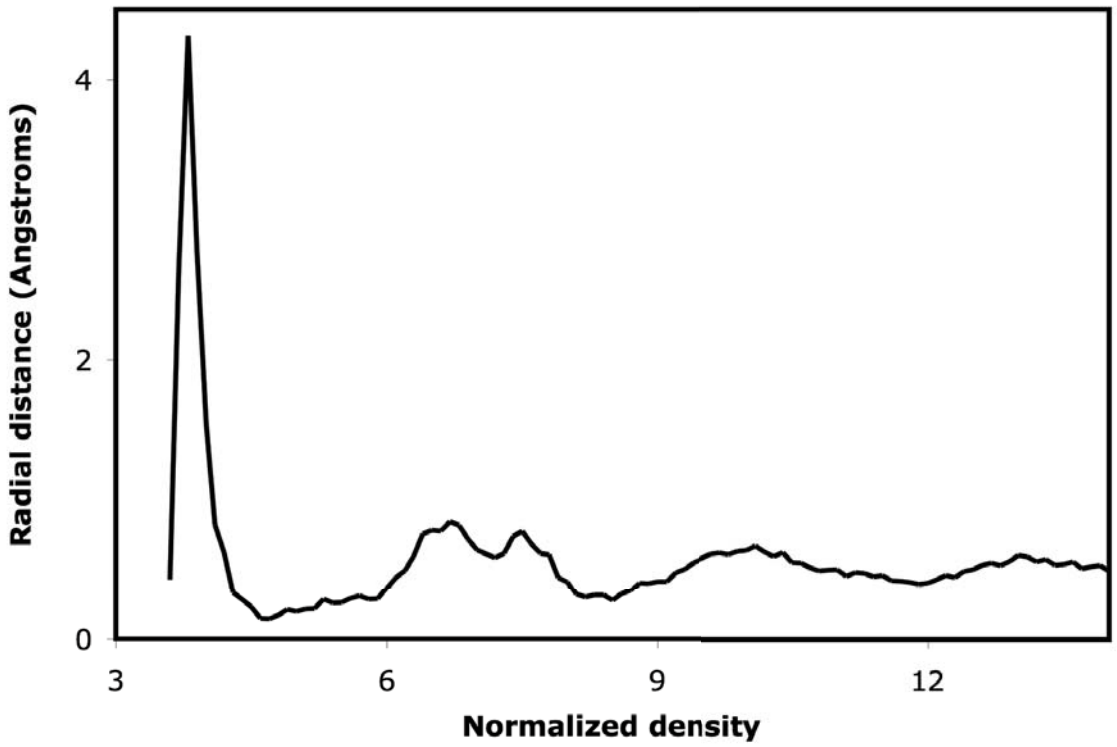
performed on the Lennard-Jones system, 52% of the random structures yielded a lower energy during the pressure annealing simulations. This indicates that pressure does not have a significant effect on the results of the simulation, since the result is randomly distributed within the two simulation techniques. For the other two types of

simulation the effect seems more pronounced. For the monoglyme simulations, the pressure annealing method generated the lower energy 84% of the time. For the tetragylme simulations, pressure annealing generated the lower energy 68% of the time. Each of these models' data was accumulated over the course of twenty-five random structures. While the Lennard-Jones system shows no apparent bias towards pressure annealing, the bulkier models seem to demonstrate noticeably better results when using the pressure annealing method. This supports the hypothesis that the expansion of the system under low pressure, followed by later compression allows the system to pack itself more efficiently. The point particles represented in the Lennard-Jones simulations have little packing complexity, beyond forming a relatively simple lattice. Each of the polymer models has not only intermolecular packing concerns, but intramolecular packing to deal with as well. Expansion frees the system to perform intramolecular reorganization, and the compression helps find an optimal intermolecular arrangement. Based on these results, using pressure as an optimization parameter seems to benefit the search for low energy structures of hindered systems.

6.4.2 Structural Results

It is important to verify that each of the states generated correspond to relevant structures that have been previously observed to ensure that the data is of high quality. Towards this end, each of the states generated by these methods were characterized in a variety of ways, and checked against previous results. For the Lennard-Jones simulations, pair-radial distributions functions were generated and compared with those of similar systems. For the polymer models, radii of gyration

Figure 6.3. Radial distribution function formed from the data taken from the lowest energy structure found in the Lennard-Jones simulations. It is representative of the results, and compares well with previously obtained results.



and mean square end-to-end distances were calculated and compared. These values allowed calculation of the system's characteristic ratio, an additional check. Finally, the torsion angles along the backbone of the polymer models were analyzed using a technique called torsional triad analysis, which can be further compared with previous results. In general, the states generated during the course of this work match well against previous results, and imply that calculated average structures are similar to those reported previously.[33-35]

6.4.2.1 Radial Distribution Functions

One of the most basic measures of structure of condensed phase systems is the radial distribution function. Having importance both theoretically and experimentally, this

function is used here as a quick look at the arrangement of atoms in the various systems understudy. The function can be represented in an algorithmically suitable form, which was used to generate the figures included below:

$$g(r) = \frac{V}{N^2}\left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle .^2$$

Here, r is the distance between atom pairs, V is the volume, N is the number of particles in the system, and r_{ij} is the distance between a specific atom pair. The fact that the radial distribution function can be used to indicate the structure of a system is apparent: is counts the number (density) of atom pairs that match a given criteria within a certain distance window from each other relative to the density of a bulk fluid. For example in Figure 5.3, the pair-radial distribution function representing the lowest energy Lennard-Jones system found in this work is shown. This radial distribution function is typical of fluids with a relatively high degree of order. The function is zero until approximately 3.9 Å (the distance of closest approach due to the interatomic potential function). The high initial peak results from nearest neighbor contacts, and the smaller but clearly defined peaks and dips that extend further out show a higher than average probability of atoms located in those spots, which indicates the presence of additional order. This figure is typical of the results generated in this work for these systems, and is a good comparison with previous work performed on like systems.[40]

6.4.2.2 Radius of Gyration, Mean Square End-to-End Distance, and the Characteristic Ratio

The remaining tests in this section are used to characterize the structures of the two oligomeric systems studied here. The three measurements described in this section

are interrelated and measure bulk properties of the molecules under study – specifically they describe the arrangement of the atoms in relation to the center of mass or the ends of the molecule. The first measurement, the radius of gyration, is defined as

$$S^2 = \frac{1}{N}\sum_{k=1}^{N}\left\|\mathbf{r_k} - r_{mean}\right\|^2,$$ [42]

and measures the extent to which the polymer's mass deviates from its center of gravity. It gives some indication as to how tightly the polymer arranges itself. Of course, the chemical and physical nature of the polymer determines this value, which varies according to the physical conditions in which the polymer exists. However, previous studies showed that for tetragylme, the value is approximately 20 Å^2.[34] Published data for monoglyme are unavailable. The mean square end-to-end distance, R^2, of the polymer is quite easily defined. It is simply the distance from one end of the polymer to the other, squared,[42] and then averaged over the number of polymer molecules in the sample. This gives a good measure of whether the polymer is stretched (or coiled) or compact. Again, this measurement depends directly on the polymer in question, and the physical conditions in which it exists. In tetraglyme a value of approximately 140 Å has been reported. Finally, the characteristic ratio[42] is an important measurement, because it relates the radius of gyration in a unitless form that is easier to compare between oligomers of different lengths. Additionally, according to theoretical calculations, the value has significance in determining the amount of flexibility the polymer has. For a free-jointed polymer of infinite length, the characteristic ratio should be 1. In the freely rotating chain model, the value can be used to determine the bond angle.[42] The characteristic ratio is defined as

Table 6.4. Structural information for each of the runs involving monoglyme is presented. Radii of gyration, mean squared end-to-end distances, and characteristic ratios are compiled from the final structures from each run for both the simulated annealing and pressure annealing simulations.

Structural data for monoglyme systems

	Simulated Annealing			Pressure Annealing		
	$\langle R^2 \rangle (Å^2)$	$\langle S^2 \rangle (Å^2)$	C_n	$\langle R^2 \rangle (Å^2)$	$\langle S^2 \rangle (Å^2)$	C_n
1	29.95462	12.64294	4.131672	30.05138	12.63144	4.145018
2	29.25339	12.642	4.03495	30.05138	12.63144	4.145018
3	29.04504	12.68264	4.006212	30.05138	12.63144	4.145018
4	29.71162	12.62215	4.098154	30.05138	12.63144	4.145018
5	29.33647	12.65295	4.04641	30.05138	12.63144	4.145018
6	29.25582	12.67993	4.035286	30.05138	12.63144	4.145018
7	29.19647	12.64122	4.027099	30.05138	12.63144	4.145018
8	29.50217	12.6627	4.069265	30.05138	12.63144	4.145018
9	29.27499	12.73361	4.03793	30.05138	12.63144	4.145018
10	29.5682	12.64571	4.078372	30.05138	12.63144	4.145018
11	29.42254	12.64617	4.058281	29.56905	12.65642	4.07849
12	29.42254	12.64617	4.058281	29.81467	12.69759	4.112368
13	29.42254	12.64617	4.058281	29.93132	12.67364	4.128458
14	29.42254	12.64617	4.058281	29.8647	12.66215	4.119269
15	29.42254	12.64617	4.058281	29.47143	12.64929	4.065025
16	29.42254	12.64617	4.058281	29.97904	12.63597	4.13504
17	29.42254	12.64617	4.058281	29.2091	12.69603	4.028841
18	29.42254	12.64617	4.058281	29.59777	12.72194	4.082451
19	29.42254	12.64617	4.058281	29.86696	12.67191	4.119581
20	29.42254	12.64617	4.058281	29.80467	12.64283	4.110989
21	29.33573	12.68304	4.046308	29.23696	12.67301	4.032684
22	29.78779	12.65052	4.108661	29.82505	12.68439	4.1138
23	29.23872	12.62334	4.032927	30.34298	12.72182	4.185239
24	30.01462	12.66012	4.139948	30.06649	12.6111	4.147102
25	29.26337	12.65458	4.036327	29.70253	12.61566	4.096901

Table 6.5. Structural information for each of the runs involving tetraglyme is presented. Radii of gyration, mean squared end-to-end distances, and characteristic ratios are compiled from the final structures from each run for both the simulated annealing and pressure annealing simulations.

Structural data for tetraglyme systems

	Simulated annealing			Pressure Annealing		
	$\langle R^2 \rangle (Å^2)$	$\langle S^2 \rangle (Å^2)$	C_n	$\langle R^2 \rangle (Å^2)$	$\langle S^2 \rangle (Å^2)$	C_n
1	124.1425	12.95637	4.280776	129.1815	12.76893	4.454534
2	118.7149	12.88302	4.093617	130.1666	12.89379	4.488503
3	132.4862	12.90877	4.56849	129.0914	12.898	4.451428
4	129.6162	12.86335	4.469524	133.3446	12.976	4.59809
5	124.0264	12.92801	4.276772	123.1806	12.94306	4.247607
6	119.7428	12.95901	4.129062	138.1344	12.9213	4.763255
7	130.5432	12.86378	4.50149	128.7697	12.90611	4.440334
8	135.4648	12.82505	4.6712	132.9224	12.91239	4.583531
9	132.8103	12.90405	4.579666	125.0199	12.84688	4.311031
10	116.7909	12.90094	4.027272	128.7571	12.91853	4.4399
11	124.7005	12.93324	4.300017	128.4394	12.86183	4.428945
12	133.3978	12.79128	4.599924	134.0236	12.93111	4.621503
13	117.9688	12.89667	4.06789	138.3031	13.00522	4.769072
14	119.3006	12.87578	4.113814	125.1847	12.86455	4.316714
15	123.0498	12.9339	4.243097	139.8334	12.81507	4.821841
16	128.7497	12.9282	4.439645	134.7138	12.84843	4.645303
17	135.5878	12.84452	4.675441	140.6227	12.95821	4.849059
18	130.0819	12.98611	4.485583	136.3271	12.9139	4.700934
19	129.7357	12.87325	4.473645	141.1341	12.86061	4.866693
20	131.3619	12.96661	4.529721	148.0451	12.96339	5.105003
21	131.6316	12.92065	4.539021	131.696	12.83847	4.541241
22	118.5199	12.95	4.086893	134.5525	12.88816	4.639741
23	127.7803	12.85471	4.406217	125.419	12.94454	4.324793
24	134.0905	12.92728	4.62381	137.0178	13.00921	4.724752
25	110.0911	12.90026	3.796245	129.1935	12.95532	4.454948

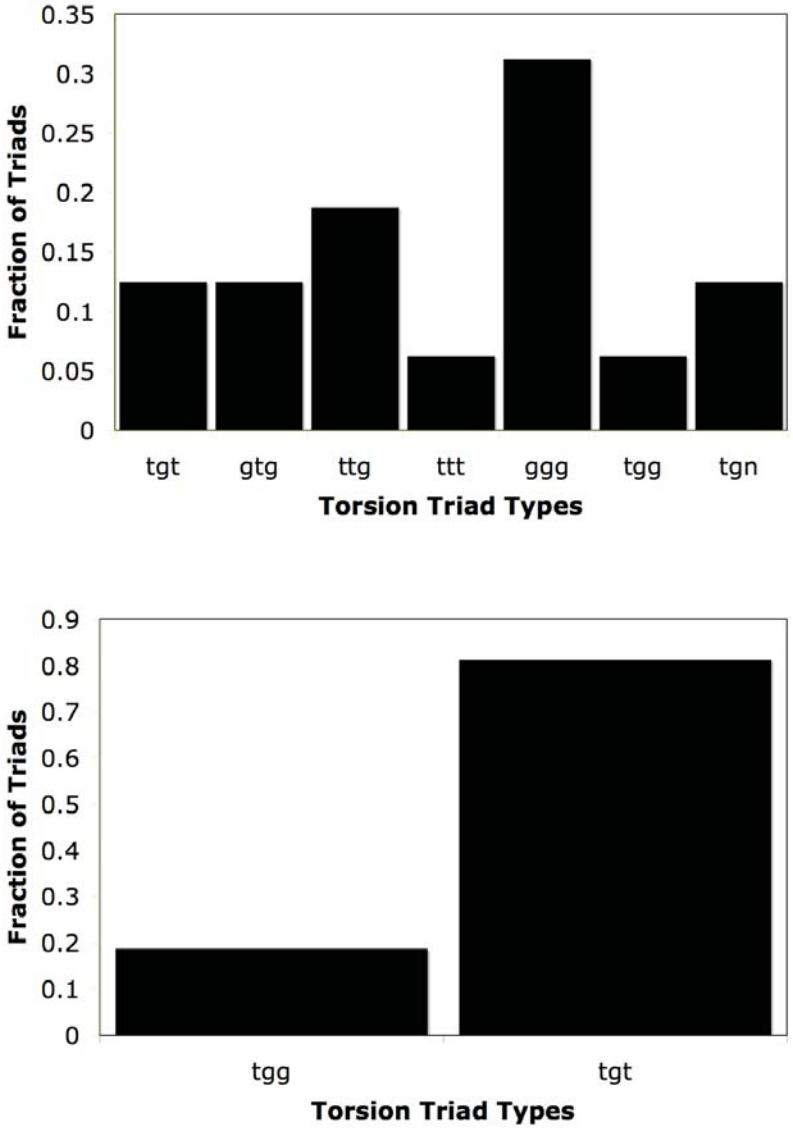
$$C_n = \frac{R^2}{nl^2}.$$

where n is the number of backbone bonds in the oligomer and l is an average bond distance in polymer system and has a value of approximately 1.5 Å for the glymes (based on published geometries, see reference 33, for example). C_n for tetraglyme has been reported to be 4.9.[33, 34] Tables 6.4 and 6.5 list the radii of gyration, mean squared end-to-end distances, and characteristic ratios for the oligomer simulations performed here. The mean squared end-to-end distances for tetraglyme are within 10 Å of that previously reported, and the characteristic ratio is within a few tenths of an Ångstrom. The radii of gyration are uniformly lower than previously reported, and perhaps indicate a difference due to temperature effects, or the fact that these represent a single minimized structure, while the published value is taken from an ensemble average at 300 K.

6.4.2.3 Torsional Triad Population

This measurement[33-35] is specific to oligomers, as it can only be defined for a given backbone atom sequence of three bonds. Similar measurements can, of course, be defined for other polymers. Triad analysis for PEO requires measuring a series of three torsion angles each time they occur, classifying the conformations as trans (labeled t, incorporating angles from 120 degrees to 240 degrees), gauche (labeled g, incorporating angles from 0 degrees to 120 degrees), or gauche minus (labeled n, incorporating angles from 240 degrees to 360 degrees). The triads are characterized by three letters. The combination ttt, for example, indicates that all three angles in the triad are trans. Finally, the number of times each combination occurs is counted and used to generate the figures described below. As shown in Figure 5.1, the three

Figure 6.4. A pair of histograms detailing the population of triad types found in a representative monoglyme system a.) before the simulation was performed and b.) after the simulation was performed. In the initial state, a wide variety of triad combinations existed in the system. After the optimization procedure, only two existed, both of which represent low-strain structures as determined by previous work.

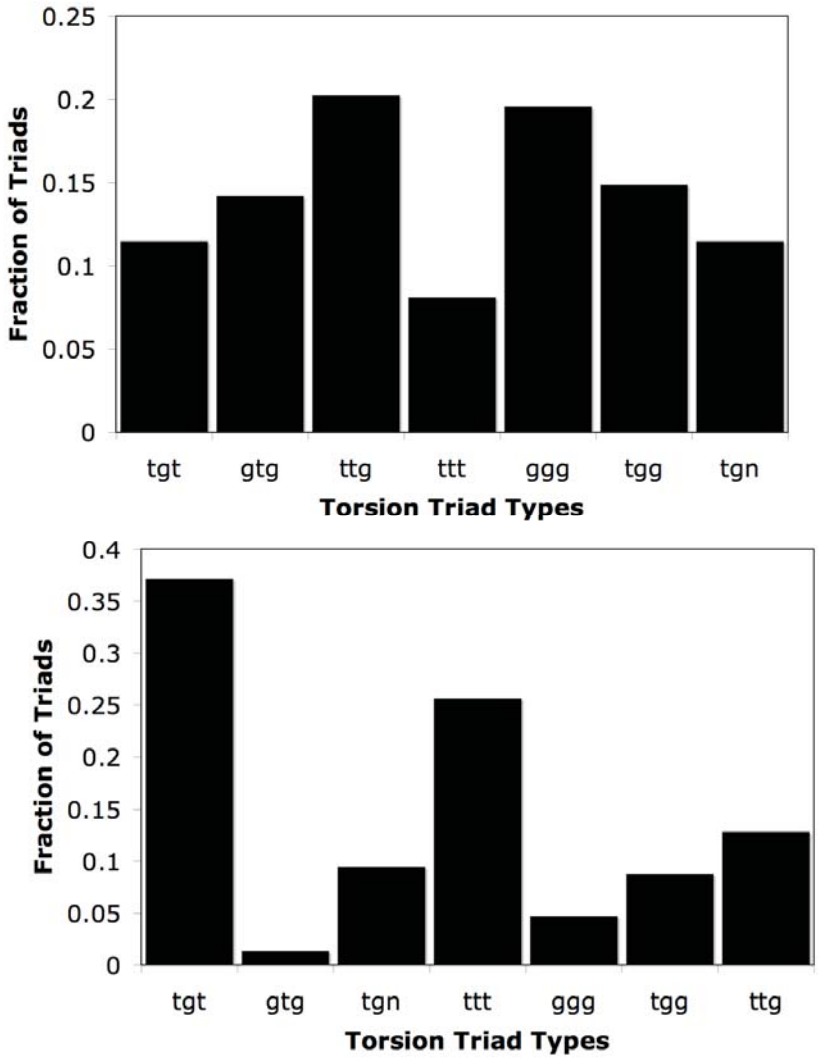


angles of interest are first the torsion angle around the bond between the atoms labeled 2 and 3, the bond between the atoms labeled 3 and 4, and finally, between 4 and 5. Monoglyme thus contains a single triad. Tetraglyme, on the other hand, contains four triads in each molecule. Figure 6.4 shows how the distribution of triads changed between the random initial state and the final state for the monoglyme simulation that produced the lowest energy. Initially, the triad distribution was randomly distributed, containing a wide variety of triad combinations. After the optimization was performed, only two types of triads remained, corresponding to low energy states of each individual strand. The tgt triad dominated, as expected since it is a dominant triad in glyme systems. [33-35] A similar situation occurred in the tetraglyme simulation which yielded the lowest energy, depicted in Figure 6.5. Again, the initial state contained a relatively high variety of traid combinations, but after the simulation was performed more dominant triads emerged. Two triads in particular, tgt and ttt each became more populated, which is consistent with previously reported results.[33-35] Also, states such as gtg which are only minimally populated in previous work dropped in population considerably.[33-35]

6.5 CONCLUSION

We have shown that pressure annealing can help to find low energy structures of systems complicated by steric hindrance. We hypothesize that this is due to the effect that lowered density (due to the lowered pressure) has on a bulky molecule's ability to rearrange itself. Pressure annealing was tested here by allowing such conditions to be realized. Additionally, the final compression phase helps to pack the well-folded molecules onto each other. We showed that bulky molecules gave a lower energy

Figure 6.5. A pair of histograms detailing the population of triad types found in a representative tetraglyme system a.) before the simulation was performed and b.) after the simulation was performed. In the initial state, a wide variety of triad combinations existed in the system. After the optimization procedure, the populations shifted. Triad combinations shown to be common in tetraglyme systems increased in population, while those rarely represented in previous results decreased.



nearly 70% of the time (and has high as 80% in monoglyme) during pressure annealing simulations when compared to conventional simulated annealing. In contrast, for the simpler Lennard-Jones model pressure annealing performed better only about 50% of the time and thus gave no appreciable benefits. For the systems studied, the radial distribution functions, radii of gyration, mean squared end-to-end distances, characteristic ratios, and torsion triad populations generated by the method compare well with published results. Pressure annealing shows promise and should be further studied by applying it to other interesting systems. In particular, applications of the technique to studying the low energy conformations of peptides and small proteins might be pursued. This method, especially in simulations with explicit solvent, would seem to be a very promising approach.

Finally, different heating and cooling schedules can have a significant effect on the results of simulated annealing studies and an extensive literature exists discussing various possible heating/cooling schedules. Although we have chosen one particular cooling (and expansion/compression) schedule simply to illustrate the utility of pressure annealing, many others exist and their effects should be investigated systematically in subsequent work. Another avenue that deserves further exploration includes incorporating the ideas of this work into the replica exchange framework. Exchanging replicas at different pressures as well as different temperatures could perhaps yield a robust optimization technique, particularly in systems containing explicit solvent.

6.6 REFERENCES

1. Malandro, D. L.; Lacks, D. J., Volume dependance of potential energy landscapes in glasses. *Journal of Chemical Physics* **1997,** 107, (15), 5804-5810.

2. Allen, M. P.; Tildesley, D. J., *Computer Simulations of Liquids*. Clarendon Press: Oxford, 1987.

3. Tsallis, C.; Stariolo, D. A., Generalized simulated annealing. *Physica A* **1996,** 233, (1-2), 395-406.

4. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983,** 220, (4598), 671-680.

5. Catoni, O. In *Simulated annealing algorithms and Markov chains with rare transitions*, Seminaire de Probabilites, Strasbourg, 1999; Stringer-Verlang: Strasbourg, 1999.

6. Nilges, M.; Gronenborn, A. M.; Brunger, A. T.; Clore, G. M., Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Engineering* **1988,** 2, (1), 27-38.

7. Hohl, D.; Jones, R. O.; Car, R.; Parrinello, M., Structure of sulfur clusters using simulated annealing: S2 to S13. *Journal of Chemical Physics* **1988,** 89, (11), 6823-6835.

8. Wilson, S. R.; Cui, W. L., Applications of simulated annealing to peptides. *Biopolymers* **1990,** 29, (1), 225-35.

9. Cohn, H.; Fielding, M., Simulated annealing: Searching for an optimal temperature schedule. *SIOPT* **1999,** 9, (3), 779-802.

10. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999,** 314, 141-151.

11. Koehl, P.; Delarue, M., Mean-field minimization methods for biological macromolecules. *Current Opinion in Structural Biology* **1996,** 6, (2), 222-6.

12. Ladd, M. F. C.; Palmer, R. A., *Structure Determination by X-Ray Crystallography*. Springer: 1994.

13. Cui, F., *Distance-based NMR strucutre determination and refinement*. ProQuest/UMI: 2006.

14. Townes, C. H.; Schawlow, A. L., *Microwave Spectroscopy*. Dover: 1975.

15. Pulay, P.; Fogarasi, G., Geometery optimization in redundant internal coordinates. *Journal of Chemical Physics* **1992,** 96, 2856.

16. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H., Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *Journal of the American Chemical Society* **1989,** 111, (23), 8551.

17. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.

18. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T., *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press: New York, 1986.

19. Elber, R.; Karplus, M., Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **1990,** 112, (25), 9161-75.

20. Gerber, R. B.; Buch, V.; Ratner, M. A., Time-Dependent Self-Consistent Field Approximation for Intramolecular Energy-Transfer .1. Formulation and Application to Dissociation of Vanderwaals Molecules. *Journal of Chemical Physics* **1982,** 77, (6), 3022-3030.

21. Simmerling, C.; Miller, J. L.; Kollman, P. A., Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy To Locate the Experimental Structure of a Nonhelical Nucleic Acid. *Journal of the American Chemical Society* **1998,** 120, (29), 7149-7155.

22. Simmerling, C.; Lee, M. R.; Ortiz, A. R.; Kolinski, A.; Skolnick, J.; Kollman, P. A., Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* **2000,** 122, (35), 8392-8402.

23. Roitberg, A.; Elber, R., Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* **1991,** 95, (12), 9277-87.

24. Zheng, Q.; Rosenfeld, R.; DeLisi, C.; Kyle, D. J., Multiple copy sampling in protein loop modeling: computational efficiency and sensitivity to dihedral angle perturbations. *Protein Science* **1994,** 3, (3), 493-506.

25. Stultz, C. M.; Karplus, M., On the potential surface of the locally enhanced sampling approximation. *Journal of Chemical Physics* **1998,** 109, (20), 8809-8815.

26. Wickstrom, L.; Okur, A.; Song, K.; Hornak, V.; Raleigh, D. P.; Simmerling, C. L., The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure. *Journal of Molecular Biology* **2006,** 360, (5), 1094-1107.

27. Larios, E.; Pitera, J. W.; Swope, W. C.; Gruebele, M., Correlation of early orientational order of engineered lambda6-85 structure with kinetics and thermodynamics. *Chemical Physics* **2006,** 323, (1), 45-53.

28. Gnankaran, S.; Nussinov, R.; Garcia, A. E., Atomic-level description of amyloid beta-dimer formation. *Journal of the American Chemical Society* **2006,** 128, (7), 2158-2159.

29. Furlan, S.; La Penna, G.; Perico, A.; Cesaro, A., Conformational Dynamics of Hyaluronan Oligomers in Solution. 3. Molecular Dynamics from Monte Carlo Replica Exchange Simulations and Mode Coupling Diffusion Theory. *Macromolecules* **2004,** 37, (16), 6197-6209.

30. Sikorski, A., Properties of Star-Branched Polymer Chains. Application of the Replica Exchange Monte Carlo Method. *Macromolecules* **2002,** 35, (18), 7132-7137.

31. Yamada, Y.; Udea, Y.; Kataoka, Y., Replica exchange Monte Carlo simulations for folding of di-block polyampholyte. *Journal of Computer Chemistry* **2005,** 4, (3), 127-130.

32. Liboff, R. L., *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*. J. Wiley: New York, 1998.

33. Dong, H.; Hyun, J.-K.; Durham, C.; Wheeler, R. A., Molecular dynamics simulations and structural comparisons of amorphous polyethylene oxide and polyethylenimine models. *Polymers* **2001,** 42, 7809-7817.

34. Dong, H.; Hyun, J.-K.; Rhodes, C. P.; Frech, R.; Wheeler, R. A., Molecular Dynamics Simulations and Vibrational Spectroscopic Studies of Local Structure in Tetraglyme:Sodium Triflate Solutions. *Journal of Physical Chemistry B* **2002,** 106, 4878-4885.

35. Smith, G. D.; Yoon, D. Y.; Jaffe, R. L.; Colby, R. H.; Krishnamoorti, R.; Fetters, L. J., Conformations and Structures of Polyoxyethlyene melts from molecular dynamics simulations and small-angle neutron scattering experiments. *Macromolecules* **1996,** 29, 3462-3469.

36. Dias, F. B.; Lambertus, P.; Veldhuis, J. B. J., Trends in polymer electrolytes for secondary lithium batteries. *Journal of Power Sources* **2000,** 88, 169-191.

37. Tarascon, J. M.; Gozdz, A. S.; Schmutz, C.; Shokochi, F.; Warren, P. C., Performance of Bellcore's plastic rechargable Li-ion batteries. *Solid State Ionics* **1996,** 86-88, 49-54.

38. Hajek, B., Cooling schedules for optimal annealing. *Mathematics of Operations Research* **1988,** 13, (2), 311-329.

39. McQuarrie, D. A., *Statistical Mechanics*. University Science Books: Sausalito, CA, 2000.

40. Hixson, C. A.; Wheeler, R. A., Practical multiple-copy methods for sampling classical statistical mechanical ensembles. *Chemical Physics Letters* **2004,** 386, 330-335.

41. Phillips, J. C.; al., e., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** 26, 1781-1802.

42. Flory, P. J., *Statistical Mechanics of Chain Molecules*. Interscience Publishers: New York, 1969.

CHAPTER 7

Summary and Conclusions

Shown in the previous chapters are several methods developed to improve the conformational sampling of molecular dynamics (MD) computations in a variety of applications. The locally enhanced sampling (LES) approximation was explained using a conventional physical explanation. This explanation was then used to construct a new technique, called the ensembles extracted from atomic coordinate transformations (EXACT) approximation that provides a scalable link between LES and conventional MD. In addition to advantages of the method itself, computational tests of the EXACT approximation demonstrates the validity of the theoretical work presented here. In addition, a method that extends simulated annealing by using an additional optimization parameter, the pressure, was described and demonstrated. Intended for use on systems who's density limits the effectiveness of purely temperature-based optimization attempts, the method allows molecules greater freedom to do a conformational search without being limited by lack of space due to intermolecular interactions. Though different in approach, the methods described here each serve as a useful improvement of molecular dynamics as a research tool.

Though locally enhanced sampling was originally developed by appealing to a Liouville density operator approach, we showed that the method works by simply adding holonomic restraints which cause "bath" particles to be held in the same positions across several copies of the entire system. We then used this knowledge to

provide alternate explanations for several observed behaviors common to locally enhanced sampling simulations. Specifically, we explained why hot particles fail to cool in LES simulations and showed exactly why locally enhanced sampling succeeds as an optimization strategy.

Using our understanding of locally enhanced sampling, we developed an alternate simulation method, called the EXACT approximation. This method, in one limit, mimics LES. In the other limit, it mimics conventional MD. It can also generate trajectories intermediate between these extremes. It works by incorporating the parts of the dynamics that LES ignores when the contribution of a particular part (not clear) exceeds a user-defined threshold. We also demonstrated that the method can produce trajectories intermediate between LES and conventional MD in systems including argon clusters and in the molecule melatonin. The results generated by the method confirm that our understanding of locally enhanced sampling is correct.

Finally, we developed an optimization strategy that incorporates pressure alongside temperature as an analogue to simulated annealing. The method was applied to finding low energy structures for condensed phase systems of argon and the polyethers monoglyme and tetraglyme. The strategy was compared directly to simulated annealing and in the argon simulations, both methods provide comparable results. However, for both of the glyme systems, the simulations incorporating pressure variation produced lower final energies by a significant margin than those using conventional simulated annealing. There is great potential in pairing this method both with other cooling schedules and also in incorporating the ideas developed in this work into the replica exchange method.

Though finding energy minima of condensed phase systems remains a significant challenge for MD simulations, our work has contributed to the field in several ways. Firstly, we have provided theoretical insights into the relatively popular method, LES. We have also provided a rigorously derived alternative, the EXACT approximation, which both illustrates the approximation inherent in LES and provides a method that allows some of the advantages inherent in conventional MD and in LES to be retained. Finally, we illustrate how parameters other than temperature can affect the efficiency of annealing simulations, by demonstrating that pressure is a beneficial optimization parameter when studying hindered systems. The breadth of our work demonstrates the flexibility of molecular dynamics simulations as an important basis for potential energy surface optimization strategies.