EXAMINING THE PHASE BEHAVIOR OF COMPLEX FLUIDS WITH FREE

ENERGY METHODS BY COMPUTER SIMULATION

By

Lili Gai

Dissertation

Submitted to the Faculty of

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

CHEMICAL ENGINEERING

August 2014

Nashville, Tennessee

Approved:

Professor Clare McCabe

Professor Peter T. Cummings

Professor M. Douglas LeVan

Professor Jamey D. Young

Professor M. Shane Hutson

DEDICATION

This dissertation is dedicated to my encouraging parents and supportive friends at

Vanderbilt University.

ACKNOWLEDGEMENTS

Finally, I want to thank my family and friends for maintaining me in good conditions both physically and mentally. To all those I met at Vanderbilt, you are great people, especially those in Department of Chemical and Biomolecular Engineering. All of you made the life abroad free of feeling lonely, and family-like. To my closest friends and family that I don't need to name individually, I cannot thank you all enough for your support and accompany. All people have to move on, and I am happy and lucky to among those with precious and best memory.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

WL   Wang-Landau

STMC   Statistical temperature Monte Carlo

STMD   Statistical temperature molecular dynamics

GCMC   grand canonical Monte Carlo

GPU   graphics processing unit

LAMMPS   Large-scale Atomic/Molecular Massively Parallel Simulator

LJ   Lennard-Jones

MC   Monte Carlo

MD   molecular dynamics

DPPC   dipalmitoylphosphatidylcholine

DPD   dissipative particle dynamics

ODT   order-disorder transition

OP   order parameter

CER   ceramides

HOOMD-blue Highly Optimized Object-oriented Many-particle Dynamics –Blue

Edition

1D   one-dimensional

2D   two-dimensional

CMC   critical micelle concentration

CHAPTER 1   INTRODUCTION

Computer simulation has become increasingly important in engineering, especially for studying phase transitions and critical behavior that are both important in natural processes or industrial applications.[1-4] During these processes, conformational changes occur in response to changes in the environment or as part of a functional process that correspond to changes in the free energy of the system. Free energy, in general corresponds to the total amount of energy that can be converted to do work and is minimized when a system reaches equilibrium. Free energy is the central quantity in thermodynamics for the characterization of physical equilibrium,[5-7] e.g., a wide range of processes as shown in **Table 1.1** are all characterized by the free energy calculation, i.e., the comparison of the relative stabilities among systems is given by the comparison of the free energies of the relevant systems, and almost always the states of interest are those with the lowest free energy (i.e., stable states).[5, 7]

Table 1.1 A list of example systems for which free energy calculation is applied for comparisons of relative stability, from Kofke *et al.*[5]

Phase equilibrium: vapor vs. liquid vs. solid vs. etc.	Reaction equilibrium: reactant vs. product
Protein structure: folded vs. unfolded	Self-assembly/nucleated vs. dispersed

Hence, to fully understand and examine the phase behavior of complex fluids, a detailed and explicit calculation of the phase transition behavior based on the underlying free energies is necessary.[5, 6, 8] However, calculation of free energy remains a challenge in traditional molecular dynamics (MD) or Monte Carlo (MC) simulations[6, 9, 10] as a result

of the direct relation between free energy and the partition function (not a mechanical property, i.e., not expressible in terms of molecule positions and momentum), which is prohibitively difficult to be calculated from traditional simulations. Specifically, complex fluid systems frequently have rough energy landscapes, often exhibiting many minima close to each other separated by large energy barriers.[11] These features can cause problems for the numerical exploration of partition function since for example, some energy states may not be visited due to the energy barriers, and the transitions between states are thus rare events.[12] Typically, traditional MC and MD methods sample only the lower energy regions, while rarely visiting the higher energy regions, and thus can lock the simulation into a particular region of phase space, and fail to sample some of the states of interest. Additionally, very low energy states may not be visited if an energy barrier exists. Thus, traditional methods are typically unable to sample the entire accessible phase space, or even just the important states, within a reasonable amount of computational time, even for rather simple systems, leaving the partition function unsolved.[13] To circumvent this problem, various methods based on equilibrium or quasi-static simulations have been developed by focusing on the differences in free energy between states, such as thermodynamic integration,[13-15] free energy perturbation,[6, 16, 17] and Jarzynski's nonequilibrium fast growth method.[18-20] These work-based methods rely on tailoring a chain of configurational energies that connect the target and the reference states whenever their energy regions have poor overlap (as shown in **Figure 1.1**b and **Figure 1.1**e). This can result in perceived inefficiency where many "uninteresting" states are required to be sampled.[6, 21] These methods can be nontrivial for even simple systems, let alone complex chemical and biological systems, which are frequently characterized by

inhomogeneity, molecular flexibility, and strong long-range interactions.
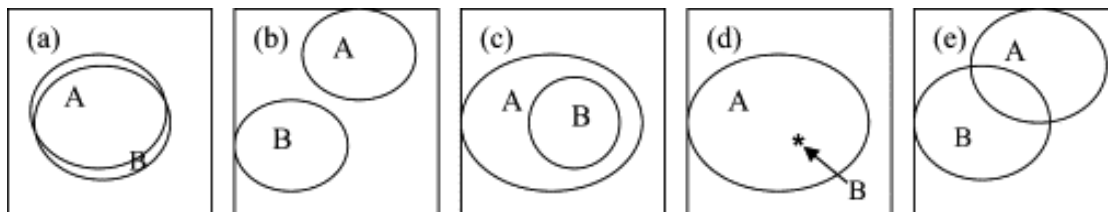
Figure 1.1 Schematic representation of the relation between two systems of interest with their important configurations, from reference,[5] (a) A and B coincide, (b) no overlap at all between A and B, (c) B is a subset of A, (d) B is a point of A, a special case of (c), (e) parts of A and B are overlapped.

To mitigate these problems and improve the energy exploration, advanced sampling methods that go beyond standard MD or MC have been developed, such as the generalized ensemble technique,[22] multiple histogram reweighting,[23] multicanonical or entropic sampling,[24, 25] replica exchange or parallel tempering,[26-28] and the WL random walk[4, 29] and its extended algorithms.[30, 31] The common feature of these methods is the calculation of the density of states (DOS). The temperature independent DOS, $g(E,V,N)$, which is the number of possible states that occupy the same energy level E, is a central quantity of interest in thermodynamics. If $g(E)$ is known with high accuracy for all energies, the canonical partition function over a wide range of temperature can be constructed with a single simulation, and all thermodynamic quantities of interest, including the free energy changes and thus phase behavior, can be determined. Among them, the WL method[4, 29] is frequently cited as among the most efficient ones, performing a random walk over the whole energy range, allowing sampling of conformations at all relevant temperatures in a single simulation. The $g(E)$ is updated at every MC move and determined finally by an iterative process. Compared to other flat-histogram methods, this dynamic update of $g(E)$ on the fly can provide quicker exploration of phase space[30,]

[31], and it can be implemented very efficiently in parallel (by multiple walkers running concurrently in the energy space), making WL suitable for efficient free energy calculation. A detailed discussion of WL sampling is provided in Section *3.1.1*. However, there are two factors limiting the further application of the WL method to large and complex systems; (1) the large increase in the number of energy levels as the system size increases due to discrete representation of DOS, and (2) the lacking of MD sampling for complicated systems beyond the scope of effective MC moves. Thus, based on the general idea of the WL algorithm, the statistical temperature molecular dynamics (STMD) method[30, 31] was developed to overcome the problems listed above, which improves the sampling efficiency and enables efficient simulations of large systems, with a detailed discussion in Section *3.1.4*. The parallel efficiency of STMD is due part to the existence of efficient parallel methods for MD simulations.

In this dissertation, two outstanding scientific problems that are still beyond the capacity of current simulation techniques were selected and studied using sampling techniques based on the WL ideas; the self-assembly of lipids and phase transitions in nano-confined fluids. This dissertation presents results from several studies aimed at elucidating the phase behavior of these complex systems, with explicit free energy, heat capacity calculations and order parameter analysis. Additionally, as an extension of the WL and STMD methods, we propose a novel hybrid MC/MD method for various ensemble studies beyond the canonical one.

Chapter 2 outlines the background related to the phase behavior of the two systems studied, self-assembly of lipid bilayers and nano-confined induced solidification of molecular fluids. Next, Chapter 3 provides a detailed account of the simulation and

theoretical tools and methods applied throughout this dissertation. Chapter 3 also details the application and extension of these methods, which are used throughout the subsequent chapters.

In Chapter 4, a series of simulations are reported for a simple lipid system based on the model proposed by Drefahl *et al.*[32] to study self-assembly using MC methods, such as Metropolis,[33] WL and statistical temperature MC (an extended WL algorithm, and discussion is provided in Section *3.1.3*), in order to demonstrate the applicability of these methods in understanding the self-assembly process. Results from the WL method show that the phase transition occurs in a two-step process with a stable intermediate phase; the second phase transition was missed in the conventional Metropolis MC simulations. This demonstrates the advantage of being able to perform a complete scan of the whole temperature range with the WL method.

Next, in Chapter 5, two different techniques - replica-exchange Wang-Landau (REWL, the detailed discussion is in Section *3.1.2*) and STMD [30, 31] are applied to systematically study the phase transition behavior of self-assembly lipids as a function of temperature using an off-lattice lipid model, both allowing direct calculation of the DOS with improved efficiency compared to the original WL method. The simulations conducted by both methods can go to very low temperatures with the whole system exhibiting well-ordered structures and several bilayer phases are observed within the temperature range studied, including gel phase bilayers with frozen water, water (i.e., frozen and liquid water), and liquid water and a more fluid bilayer with liquid water. The results obtained from both methods, STMD and REWL, are consistently in excellent agreement with each other, thereby validating both methods and the results.

In Chapter 6, the phase behavior of Lennard-Jones (LJ) fluids nano-confined between parallel solid surfaces is examined by STMD over a range of wall separations, including both ideal and non-ideal ones, as well as various wall-fluid interaction strengths. The free energy landscape is monitored between different states and heat capacities are calculated. Results showed that the transition temperature T_{ODT} is reduced with a sharper, more distinct transitions in heat capacity curve, as the pore size is increased. For non-ideal pore spacings, reduced T_{ODT} are obtained, with oscillatory behavior exhibited in the pore density, i.e., density in non-ideal pores is lower than the ideal ones.

In Chapter 7, we discuss the calculation of the joint DOS (e.g., as a function of temperature and compositions, or temperature and pressure) for continuous models. To study phase behavior in many systems, the 2D JDOS is quite useful and becomes necessary. Thus, we propose a hybrid WL/STMD method to take advantage of the benefits of STMD algorithm, but in a 2D WL context. To demonstrate the idea of the method, we performed a series of hybrid WL/STMD simulations on binary LJ systems, with the second variable being the composition of the particle (N_j) or volume (V). The results obtained are compared to those from the standard methods (i.e., original WL, WL in 2D form, and standard STMD) with independent runs and are consistently in excellent agreement with each other.

Finally, Chapter 8 summarizes the main conclusions of the dissertation and discusses recommendations for future work.

CHAPTER 2   BACKGROUND

2.1   Self-assembly

Self-assembly is a spontaneous process by which disordered molecules associate into stable and structurally well-defined aggregates as a consequence of specific and local interactions of the molecules themselves; no external forces need to be involved. Molecules that undergo a self-organized process often form hierarchical structures at the nanoscale. As such, self-assembly is a crucial capability for future large-scale manufacturing of nanotechnology-based systems. In nature, a variety of integral structures that are important in the fields of chemistry, biochemistry and biophysics, such as membranes, vesicles and virus capsids, are created via self-assembly of amphiphilic molecules.

2.1.1   Self-assembly of amphiphilic lipids

Amphiphilic lipid molecules consisting of a hydrophilic head group and hydrophobic tail groups can self-assemble into a wide variety of structures, such as bilayers, micelles, and vesicles, when mixed with a suitable solvent.[34, 35] For example, at a constant temperature, when the concentration of amphiphile is quite low, the molecules in the bulk solvent can be fully dissolved as independent monomers. When it reaches the critical micelle concentration (CMC), the monomers will form into aggregates, such as micelles, and as the concentration further increases, a variety of other phases such as cylindrical micelles or lamellar/bilayer phases will be observed (see **Figure 2.1**).[36] Among them, lipid bilayers are important structures in both natural and artificial biological interfaces (for

containment and separation),[37] such as the membranes that form the basis of almost all life forms.[38,39] As a result, the self-assembly of these systems has been extensively studied, both computationally and experimentally.[40-42] However, most experiments explore the macroscopic structure of the membrane.[43-45] A detailed understanding of the self-assembly process of the lipids on the molecular level is not yet well known and would provide important insight into nanotechnology applications.
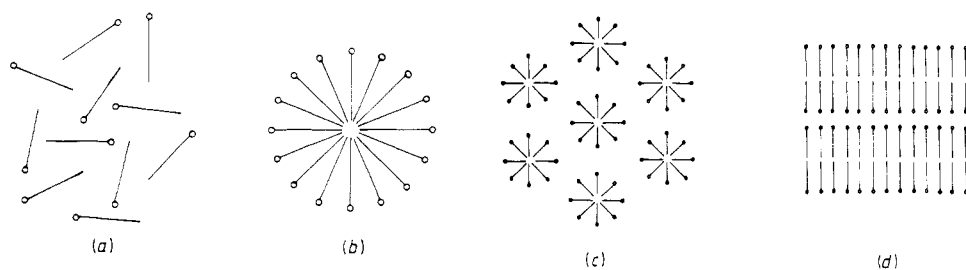
Figure 2.1 A schematic description of amphiphiles at various concentrations, from reference[36] (a)monomers, (b) micelle aggregates, (c) cylindrical micelles, (d) bilayer.

Simulation studies of lipid bilayers emerged in the late 1980's and early 1990's.[46] Initially, due to the complexity of the amphiphilic molecules and the limited resources of computational time, the work mainly utilized simplified on-lattice molecular models to capture the self-assembly behavior. Additionally most studies were limited to micellar systems in dilute solutions due to the rapid increase of computational time for dense systems, see for example, Larson *et al.*[47,48,49], Bernardes *et al.*[39, 50, 51] and Care *et al.*[36, 52] Of the few models that focus on the formation of a bilayer structure, they all adopted the conventional Metropolis MC method to study the self-assembly behavior.[32, 37, 53] For example, Brindle[37, 53] observed that at low temperatures, amphiphiles self assembled into a castellated bilayer structure in a simple cubic lattice (32 × 32 × 32). Based on the prior research, it is generally accepted that lipid self-assembly can be qualitatively explained

8

by the hydrophobic effect,[54] i.e., the monomers form aggregates in a way that minimizes the hydrophobic interaction while maintaining the hydrophilic interaction with bulk water. But complex details such as the necessary forces that direct lipids to assemble into predictable arrangements, or transient intermediate phase behavior during the aggregation process are still the subject of much debate.[55-57] Thus, a systematic study of the temperature dependence on the phase behavior is necessary. More recently, Drefahl *et al.*[32] studied the molecular interactions necessary for the self-assembly of amphiphiles on a cubic lattice. They performed a series of simulations using flexible 3-segment amphiphilic molecules with purely repulsive interactions and found that bilayer structures formed at low temperatures, with a phase transition to clusters observed at high temperatures. They provided the phase behavior as a function of temperature, by performing independent MC simulations at discrete point of temperature at an interval of 0.5 (in reduced units); however, this procedure was likely insufficient to fully disclose the entire phase behavior. Thus, the systematic study of bilayer formation with respect to statistical thermodynamics is still challenging for the traditional methods and remains a longstanding problem that has not been fully resolved.[58]

More recently, many traditional MC and MD simulations have been performed on continuum models (i.e., off-lattice) studying micelle formation from the self-assembly of amphiphiles, but limited studies have been performed on the systematic examination of bilayer phase behavior as a function of temperature.[34, 59-65] For example, Marrink developed the coarse-grained Martini force field and performed a series of work on the phase behavior of lipid bilayers.[61, 62, 66, 67] Previous work based on a simple coarse-grained lipid model by Goetz *et al.*[64] and Fujiwara *et al.*[55] reported the structural

properties of the various aggregates, such as spherical micelles, cylindrical micelles and bilayers, via the variation of the amphiphilic concentration and hydrophobic interactions (see **Figure 2.2**). Of those studies, conventional MD or MC methods, along with temperature quenching, have been used as the main simulation techniques to induce phase transitions. In addition, these studies all relied on visual inspection or calculation of order parameters to support the structural phase transitions observed as a function of the system temperature.
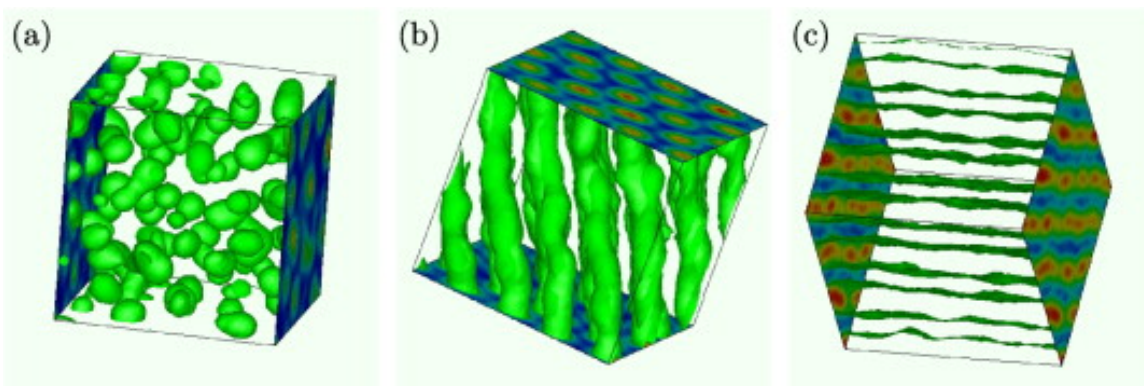
Figure 2.2 Simulation snapshots of self-assembled structures formed by amphiphilic molecules at various amphiphilic concentrations; (a) the isotropic micellar phase (b) the hexagonal phase and (c) the lamellar phase. Isosurfaces of tail particles are shown. The figures are from reference[55].

More recently, the calculation of the heat capacity (C_v) has been used as a more direct signature of a phase transition. The work of Rodgers *et al.* [59] and Nagai *et al.*[60], reported the first attempt to characterize the phase transition of a solvated amphiphilic model by monitoring changes in enthalpy. Specially, Rodgers reproduced the four different bilayer phases that are structurally in agreement with previous experiments utilizing the DPD model, such as subgel, gel, ripple, and fluid, and they studied the enthalpy changes along with them (see **Figure 2.3**). Nagai *et al.* directly calculated the C_p curve for a bilayer system that is composed of 32 dipalmitoylphosphatidylcholine

(DPPC) lipids, see **Figure 2.4**. There are two peaks indicating a pre-transition and main transition of fluid-gel phases, with the three corresponding phases shown in **Figure 2.4**; the titled gel phase is characterized as narrow, thick, and well-ordered, while the fluid bilayer is wide, thin, and disordered. These studies represent promising initial quantitative heat capacity investigations of the structural phase transition behavior. However, these methods are applied with mixed success. Although the simple DPD model given by Rodgers *et al.* [59] describes structurally all the phases of a DPPC bilayer (subgel, gel, ripple, and fluid), the enthalpy change between certain phase transitions was either underestimated or no observable enthalpy changes were detected between some of the structures. In the study by Nagai *et al.*,[60] reference temperatures at 127 points between 283 ~ 390 K are needed to implement valid swapping of structures (replicas are required to be separated by small temperature differences). As the number of degrees of freedom of the system increases, the required number of replicas also greatly increases, and so only a small system of 32 lipids was studied, with only preliminary results presented for a larger 128-lipid system. In addition, the hysteresis between heating and cooling processes is still quite substantial in the systems studied above and may mask the underlying equilibrium phases with heating and cooling processes. Recent work by Guo *et al.*[44] used heat capacity coupled with order parameter analysis to determine the order-order and order-disorder transitions in a fully atomistic ceramide bilayer structure, although it must be noted that the large computational expense necessitated a very rapid heating and cooling sequence.

	Subgel (L_c)	Gel ($L_{\beta'}$)	Ripple ($P_{\beta'}$)	Fluid (L_α)
Experiment				
DPD				
MARTINI				

Figure 2.3 Different lipid phases obtained by Rodgers *et al.*[59], from DPD and MARTINI simulations. DPD simulations obtained structures in agreement with experimental results, while MARTINI missed some intermediate phases.

(a)   (b)
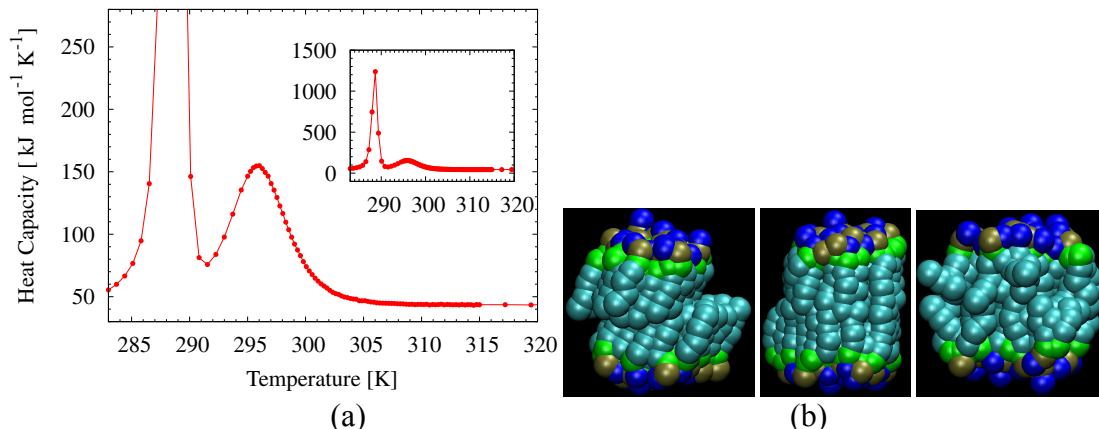
Figure 2.4 Phase behavior of 32 DPPC lipids from Nagai[60], (a) heat capacity (b) with three corresponding phases, from left to right are tilted gel phase, untilted gel phase and fluid phase.

2.2 Nano-confined system

Nano-confined fluids, defined as fluids confined by solid surfaces with separation distances on the order of a few molecular diameters or nanometers, are of great importance both in industrial applications such as lubrication, separations and catalytic processes, as well as emerging areas, such as ionic liquid-based supercapacitors.[8, 68, 69] For example, the industrial separation processes that depend on the relative adsorption of different components in a mixture, and this adsorption always involves the transfer of

12

molecules from a bulk phase into slit or cylindrical nanopores. In lubrication applications where surfaces move with respect to each other at nanometer separations, such as in a hard disk or microscale electromechanical systems, MEMS, and their nanoscale analogs, NEMS, the surface will need to be lubricated continuously under nano-confinement to prevent damage, which means solidification of the lubricant is undesirable.[70] Additional detailed applications are given in the review by Gelb *et al.*[8] Thus, understanding the phase transition behavior of nano-confined fluids is necessary for device design and control.

When molecules are confined within such narrow pores, their phase behaviors can be dramatically different from that of the bulk fluids. This surface-driven phase change can be attributed to a combination effect of fluid–wall interaction[68, 71, 72] and spatial constraint in the pores.[73] This phenomena has been widely examined experimentally,[74-80] with strong evidence of confinement induced solidification, e.g., the observation of a several orders of magnitude increase in viscosity as a function of reduced pore separation,[75, 78-80] although other studies have posited that systems undergo a constant vitrification as the separation is reduced.[76, 81, 82] In either case, it is well agreed upon that the structural and dynamical properties are altered dramatically compared to the bulk as systems undergo nanoscale confinement.[83-85] Molecular simulations have also been performed of nano-confined fluids,[68, 71, 72, 86-88] employing both coarse-grained and atomistically detailed models of the fluids and surfaces. Although many of the studies use hypothetical/simplified model systems (i.e., a united atom approach), such as in the work of Wang,[86] and Ballamudi,[89, 90] useful information can be obtained; however, these early studies ignored important characteristics of the experimental systems, such as a realistic

13

treatment of the wall-fluid interactions. Cui *et al.* [14] was the first to attempt to model the experiments[76, 78, 79, 91] with high-fidelity molecular dynamics simulations of *n*-dodecane nano-confined between mica-like surfaces. In a series of papers they demonstrated a clear first-order transition from a disordered fluid state to a layered and herringbone-ordered state.[92, 93] Recently, with the desire to more accurately represent the mica-wall surface atoms, and in particular the mica-fluid interactions, Docherty and Cummings[68, 72] performed a series of molecular simulations with fully atomistic models of the mica surfaces and confined non-polar molecules (**Figure 2.5**). As in the previous work[90], the confined fluid undergoes a transition to a layered and ordered structure. From **Figure 2.5**b~c, as the separation is reduced, one can see that the molecules undergo a rapid and abrupt transition to an FCC ordered solid-like structure from the fluid-like structure. Overall simulation studies, whether conducted with simplified or fully atomistic models, tend to predict a well-defined transition to a solid phase as the separation between confining surfaces is reduced past a critical value.

Although progress has been made in understanding the behavior of fluids under nano-confinement by preforming simulations, the length and timescales that can be accessed by simulations are still considerably limited compared to experiment, making it possible that structures found in MD or MC simulations are not the equilibrium ones, but are metastable configurations, since the true low-free-energy structures are inaccessible due to large free energy barriers. Furthermore, simulations alone do not provide a clear picture of the free energy landscape, and thus the simulation studies reviewed here were unable to find the precise location of the phase transition. Thus, to examine the nature of the liquid-solid phase transition, a detailed and explicit calculation of the phase transition

behavior based on the underlying free energies is necessary.[8] In this work, we will calculate the free energy of the nano-confined systems and study their phase transition behavior by the STMD method, which will be discussed in details in Section *3.1.4*.
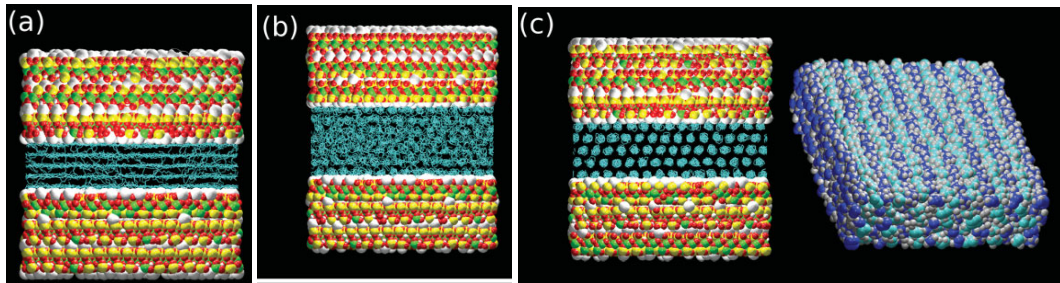
Figure 2.5 The evidence from atomistic simulations of fluids confined between mica walls, from Hugh *et al.*;[68] (a) ordered phase of *n*-dodecane confined between mica sheets, (b) liquid-like structure of cyclohexane at *n*=6 confined by mica sheets, (c) ordered solid-like structure of cyclohexane at *n*=5 confined by mica sheets, with right one shows only the structure of confined fluids.

CHAPTER 3   SIMULATION AND COMPUTATIONAL METHODS

3.1   Free energy simulation methods

In this Chapter, we review the simulation methods used in the reminder of the dissertation.

3.1.1   Wang-Landau Monte Carlo simulation

WL Monte Carlo simulation is applied in Chapters 4 to simulate the bilayer formation on a cubic lattice lipid model, and forms the basis of all other methods employed throughout the dissertation. In contrast to traditional MC methods such as the Metropolis that samples with a Boltzmann factor in order to create configurations distributed according to the canonical distribution,[33] the WL algorithm performs a random walk in energy space with a bias (non-Boltzmann factor). The sampling weight of a specific energy level is proportional to the reciprocal of the density of states, $1/g(E)$, and a flat histogram is generated for the energy distribution. Thus, a proposed move from energy state E_1 to energy state E_2 is accepted or rejected with the probability p given by

$$p(E_1 \rightarrow E_2) = min\left\{\frac{g(E_1)}{g(E_2)}, 1\right\}. \tag{3.1}$$

Initially all $g(E)$ are set equal to unity and determined later by an iterative process. Specifically, each time after an energy level E is visited, the $g(E)$ is updated by a modification factor f (>1.0) via

$$lng'(E) = lng(E) + lnf, \tag{3.2}$$

where $lng'(E)$ with a prime means the update of DOS at energy E at the current simulation step. Here, in order to fit all possible $g(E)$ into double precision mumbers for

the systems studied, $lng(E)$ is used instead of $g(E)$. At the same time, an energy histogram $H(E)$ is also accumulated and when the histogram is "flat", as defined by a pre-set tolerance, $g(E)$ is assumed to converge to the true value with an accuracy proportional to the current f. The modification factor is then reduced to a smaller value ($lnf_1 = 0.5 \times lnf_0$), the histograms reset to zero, and the next level of random walk started. In this new round, $g(E)$ is modified with the new modification factor f_1, and the random walk is terminated when f_1 approaches 1.0 with $g(E)$ converged to the true value. Usually, the natural logarithm of the modification factor f, ranges from 1.0 to 1×10^{-8}, as suggested by Landau et al.[4] The histogram is considered flat when $H(E)$ for all possible E is not less than 85% of the average histogram value, $\langle H(E) \rangle$.

The unique point of the WL method is its equal probability for sampling all the configurations and thus the system will not be trapped by local energy minima, as can happen with conventional MC. Therefore, the WL method is capable of addressing the sampling problem and long relaxation time needed for complex systems with rough energy landscapes.

3.1.1.1 *Application of Wang-Landau MC*

Initially, the WL algorithm was mainly used to study spin and lattice systems.[4, 29, 94-96] For example, the 3D Edwards-Anderson model (EA model) is a spin glass model for magnetism with very rough energy landscape (see **Figure 3.1**).[4] Based on the Boltzmann distribution, the system is very likely to be trapped in local energy minima when simulated by conventional MC methods, resulting in long relaxation time. With the WL method, in order to calculate the order-parameter at any temperature from the canonical average, Wang and Landau[4] performed a random walk in 2D energy-order parameter

space. The internal energy and entropy at zero temperature were estimated up to a lattice size of 203, and the transition temperature was estimated at $T_g = 1.2$.
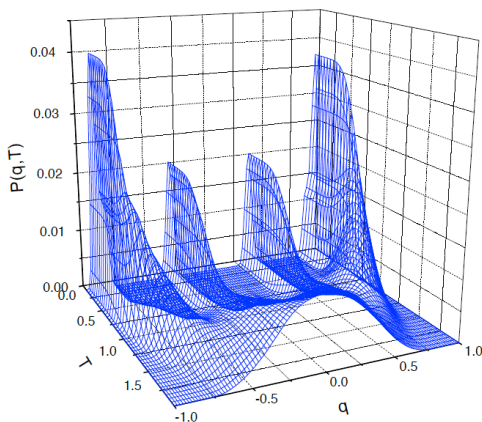
Figure 3.1 Canonical probability of Edwards-Anderson 3D model in the order-parameter space for a single distribution of bonds, from reference;[4]

More recently, researchers have extended the application of WL method to continuous systems, such as the study of collapse transition in fully flexible bead-spring and square well polymers. [97-101] For example, Seaton *et al.*[99, 102] studied the characteristic behavior of a flexible homopolymer with a chain length up to 561. By use of the WL method, they calculated thermodynamic properties over a wide range of temperatures and reached to temperatures lower than accessible in complementary Metropolis MC simulations. From specific heat capacity calculations, they found a sharp peak at low temperature, indicating a solid-liquid phase transition and a weaker coil-globule transition at high temperature. This study shows that the WL sampling algorithm is highly effective for single flexible homopolymer simulations. Similarly, the folding of simple peptide models in vacuum and continuum solvent have also been studied using WL.[11, 103, 104-105] Clearly, these works show the great potential of the WL method to explore systems with rough energy landscapes.

Most WL simulations are carried out at constant density in a *NVT* ensemble, and thus the DOS, $g(E)$, is only a function of the energy variable. However, if we want to determine phase equilibrium, which plays a key role in natural processes and industrial applications, such as vapor-liquid coexistence, the DOS is required to be a function of both the energy and density. This results in the need for calculating 2D DOS. For example, with their own extended WL sampling for random walk in both the energy and the volume space, Yan *et al.*[106] and Shell *et al.*[107] calculated vapor-liquid coexistence curves for small systems of the LJ fluids, and Poulain *et al.* studied LJ clusters.[108] Overall, these simulations require significantly more simulation time than 1D DOS and can quickly become impractical, even for some simple systems.

3.1.1.2 *Method Validation with Ising model*

As an instructive test, the WL algorithm was applied to a classic problem, namely the Ising model.[40] The Ising model consists of discrete variables called spins, s_i, that can be in one of two states, as shown in **Figure 3.2**a. Each spin can interact only with its nearest neighbors, and the energy function is defined as $E = \sum_{i \neq j} s_i s_j$. The Ising model is one of the simplest statistical models showing a phase change, and in the case of zero magnetic field has a complete analytic solution.[109] Thus it is an ideal benchmark for testing new algorithms and theoretical approaches.

Since the proposal of the WL method in 2001, the Landau group has tested the algorithm on the Ising model at different sizes, including 32×32, 50×50 and 256×256.[4, 29] To verify our own implementation of the WL algorithm, we performed calculations on the 2D Ising model with L=256 and obtained good agreement with the

results of Landau *et al.*,[40] as shown in **Figure 3.2**. Here, the WL parameter settings are as follows: initially all $\ln g(E)=0$, $H(E)=0$, flatness criterion of 85%, $\ln f$ changed from 1.0 to 10^{-8}, and the system is run in parallel with 23 sub-energy windows.

(a)   (b)
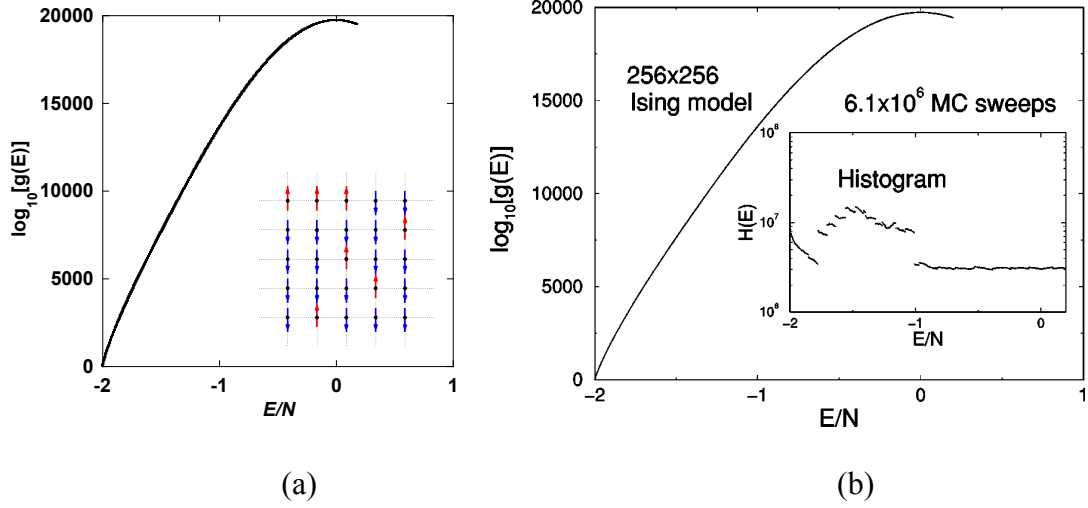
Figure 3.2 Density of states for Ising model; (a) from this work (inset, up in red, +1 or down in blue, -1), (b) from Landau *et al.*[40]

3.1.2   Parallelization of Wang-Landau MC algorithm

The WL algorithm is based on random walks in energy space, which can be implemented in parallel for efficient sampling of the complex systems by parallel "walkers", i.e., multiple side-by-side simulations continually updating $g(E)$ through time.

Generally, there are two routes for the parallelization of the WL algorithm. Firstly, to speed up the simulations, the total energy range can be divided into multiple smaller sub-windows, with each sampled by an independent WL random walker.[110, 111] The strategy of multiple energy windows are frequently used in WL simulations[110, 111] and have been adopted in this work in Chapter 4. In this approach, the total energy range is divided into overlapping energy windows and $g(E)$ is estimated for each window using

independent random walks. The $g(E)$ for the whole energy range is then obtained via a matching procedure in which the overlapping energy regions are *stitched* together. The total simulation time is largely determined by the convergence speed of the slowest walker (i.e., low energy windows, or windows involved with phase transitions). And also for complex system we can introduce unequal distribution of the energy space to faster the convergence of these slow walkers.[112] The details of the parameter setting for running simulation in multiple energy windows is discussed in details in Section *4.2.1*.

Another parallelization scheme is multiple random walkers work simultaneously on the same DOS and histogram (i.e., for the same energy windows), such as the previous parallelization work implemented by GPU,[113] shared memory,[114] and distributed memory.[115] In the work of Yin *et. al*, a parallel WL algorithm on multiple GPUs (Fermi architecture card) are implemented for the simulation of water clusters. Results showed that they obtained an average of about 50 times speedup for a given workload, in contrast to a single thread code on an intel i7-930.

3.1.2.1 *Replica-exchange Wang-Landau*

As stated earlier, the WL method performs random walks in energy space that can be parallelized by using multiple "walkers" and can be broken up into smaller overlapping sub-energy windows that are stitched together at the end. To further speed up the convergence of the simulations, the replica exchange technique was combined with the WL method. We accomplished this in collaboration with Thomas Vogel and David Landau at the University of Georgia at Athens, and this is employed in Chapter 5. As in every MC scheme, a conformational update is proposed at each step that is accepted with a certain probability P_{acc} depending on the total energy E (note reduced units are used and

so E is the reduced dimensionless energy) of the actual (X) and proposed new state (Y) and the statistical ensemble. In this case,

$$P_{\text{acc}}(X \rightarrow Y) = \min\left[1, \frac{g(E(X))}{g(E(Y))} \cdot b\right],$$ (3.3)

where b is a correction factor for any configurational bias that might be introduced by the MC update moves.

To speed up the simulation, the global energy range is divided into individual energy windows[116] with significant energetic overlap. As in the typical replica exchange algorithm, these overlapping energy windows are allowed to swap configurations with neighboring energy windows, which are accepted with:

$$P_{\text{acc}}(X_i Y_j \rightarrow Y_i X_j) = \min\left[1, \frac{g_i(E(X))}{g_i(E(Y))} \frac{g_j(E(Y))}{g_j(E(X))}\right].$$ (3.4)

P_{acc} depends on the total energy E (note reduced units are used and so E is the reduced dimensionless energy) of the actual (X) and proposed new state (Y) and the statistical ensemble. Here, $g_i(E(X))$ is the current estimate of the density of states in energy window i at the energy of its actual conformation X. More generally, we allow for multiple, individual WL walkers in each energy window that merge their density of states estimates at defined stages in the WL run, resulting in a decrease in the statistical error. For a detailed introduction and discussion of the parallel WL framework used, see reference.[112]
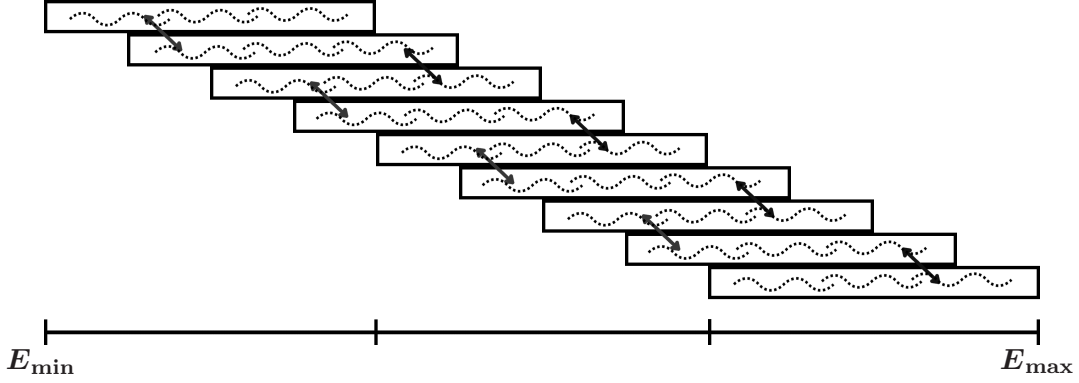
Figure 3.3 Schematic representation of the division of the total energy range into nine individual energy windows with fixed mutual overlap of 75%. Three WL walkers are deployed in each window as indicated by the dotted lines. Arrows illustrate the replica-exchange communication scheme, i.e., walkers can exchange configurations with walkers from neighboring windows.

In this study, nine individual energy windows are created with a mutual overlap of 75% (for a better exchange of the replicas between different energy windows) and three WL walkers per window. This set-up has been found suitable and leads to significant speed-ups compared to single processor WL simulations.[112] **Figure 3.3** provides a schematic overview of the simulation set up.

3.1.3   Statistical temperature Monte Carlo

This simulation algorithm is implemented in Chapter 5. The WL Monte Carlo algorithm has attracted considerable interests due to its ability to generate a flat energy distribution with the dynamic update of $g(E)$; however, the large increase in the number of energy levels as the system size increases limits its application to large and complex systems. Recently, to improve this, Kim *et al*. derived the statistical-temperature Monte Carlo algorithm, which extends the WL method to simulations with larger energy intervals by updating the statistical temperature $T(E)$ instead of DOS.[30, 31] STMC achieves a flat

energy distribution by the systematic refinement of $T(E)$, which is realized by the thermodynamic relationship between $T(E)$ and $g(E)$ via[117]

$$\frac{1}{T(E)} = \left(\frac{\partial S}{\partial E}\right)_{N,V} = \frac{\partial lng(E)}{\partial E}. \tag{3.5}$$

By using the finite difference approximation with Equation 3.5 and combining it with the $g(E)$ update scheme Equation 3.2 (converted to entropy S, $S_j' = S_j + lnf$, where f is the modification factor), a dynamic update of the statistical temperature is realized by

$$T_{j\pm1}' = \frac{T_{j\pm1}}{1 \mp T_{j\pm1} \times (lnf/2\Delta E)}, \tag{3.6}$$

where $\Delta E = E_{j+1} - E_j$ ($j = nint[E/\Delta E]$) (*nint* being a function returning the nearest integer value) is the energy bin size, i.e., the update changes temperature estimates only at the discrete values of $E_{j\pm1}$. For a detailed derivation of the above formulations, the reader is referred to the original references.[30] The temperature is updated in a range $T_l < T_j < T_h$ and at the boundary or beyond set to $T_j = T_l$ or T_h as appropriate, where T_l and T_h are the lower and upper boundaries, respectively. Due to this limited sampling range of $T(E)$, the initial lnf is usually set to a value very close to zero (with the actual value depending upon the temperature range being sampled), in contrast to the WL method in which lnf is initially set to 1.0 in order to cover a large range of energy states.

To implement STMC, all T_j are set to a constant value at the start of the simulation (usually T_h). A proposed move is accepted or rejected with the same acceptance rule as in the original WL method Equation 3.1. After an energy bin E_j is visited, the actual $T_{j\pm1}$ is updated by Equation 3.6, and then integrated by Equation 3.5 to determine the current $g(E)$ for the acceptance probability. The fact that T_j is only defined on the grid points E_j and the sharp change of the integrand at low temperatures makes it

difficult to apply a generic numerical integration, and thus a linear interpolation scheme[31] is used to connect successive bins. For a value of the potential energy E between E_j and E_{j+1}:

$$T(E) = T_j + \lambda_j(E - E_j), \tag{3.7}$$

where $\lambda_j = (T_{j+1} - T_j)/\Delta E$ is the slope of the linear segment connecting the two neighboring bins. The continuum $g(E)$ estimate is then calculated by analytical integration at the end of MD simulations:

$$
\begin{aligned}
ln[g(E)] &= \int_{E_l}^{E} \frac{1}{T(E')} dE' \\
&= \sum_{j=l+1}^{i^*} \frac{1}{\lambda_{j-1}} ln\left[1 + \frac{\lambda_{j-1}(E_j - E_{j-1})}{T_{j-1}}\right] + \frac{1}{\lambda_{i^*}} ln\left[1 + \frac{\lambda_{i^*}(E - E_{i^*})}{T_{i^*}}\right], \\
&= \sum_{j=l+1}^{i^*} L_j(E_j) + L_{i^*+1}(E)
\end{aligned}
\tag{3.8}
$$

where $i^* = i - 1$ for $\bar{E}_{i-1} \leq E \leq E_i$ $(\bar{E}_i = (E_i + E_{i+1})/2)$, and $i^* = i$ for $E_i \leq E \leq \bar{E}_i$, and $L_j = \frac{1}{\lambda_{j-1}} ln\left[1 + \frac{\lambda_{j-1}(E - E_{j-1})}{T_{j-1}}\right]$. By updating the intensive variables T_j located at discrete points, a continuum description of $g(E)$ can be obtained through analytical integration of $T(E)$. Therefore, moderately larger and fewer bins can be used to accelerate the simulation convergence, which still maintains the statistical accuracy.[11]

3.1.3.1 *Validation of STMC by 2D Ising model*

To verify our implementation of the STMC algorithm, we also performed calculations on the 2D Ising model of L=32, with three different energy bin sizes, following the same simulation details as Kim *et al.* As shown in **Figure 3.4**, we obtained good agreement with the results of Kim *et al.* [30]
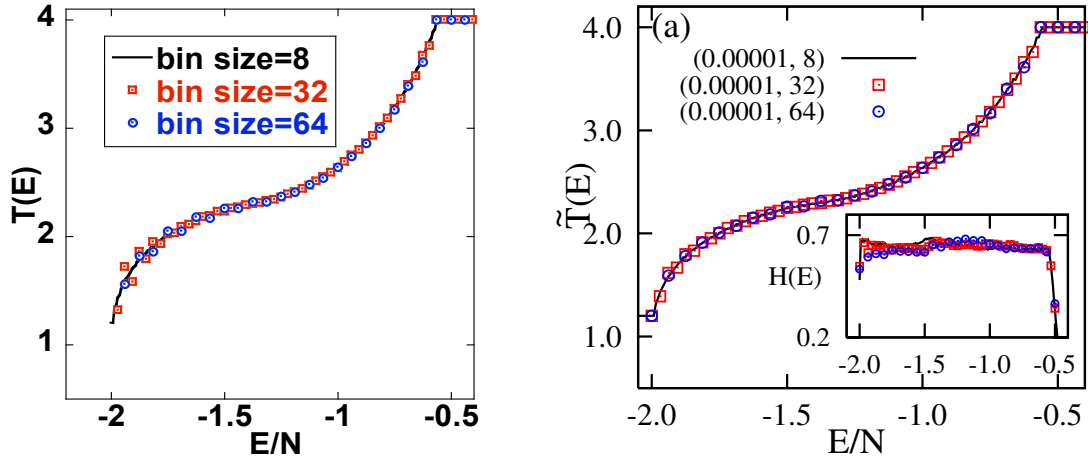
Figure 3.4 Convergent temperature $T(E)$ for 32×32 Ising model; (a) from this work, (b) from Kim *et al.* [30]

3.1.4   Statistical temperature molecular dynamics

3.1.4.1   *Overview of MD simulation*

Based on classical Newtonian equation of motion, MD simulation provides the time evolution of interacting molecules. The dynamics can be directly tracked at the atomic level, making MD especially competitive in molecular biology and chemical areas. Normally, a standard MD simulation follows the steps: initialize the system, compute the forces on all particles, and then integrate Newton's equations of motion through a time step $\Delta t \approx 10^{-15}s$. These two core steps are repeated until we have computed the time evolution of the system for sufficient time (typically, *ns* or longer). When finished, we compute the averages of measured quantities.[33]

Specifically, the time evolution of a set of interacting particles follows Newton's equations of motion, which are solved numerically by calculating the forces on each atom in the system. Due to many-body nature of the problem the continuous equations of motion are discretized and evaluated at regular time intervals, or timesteps. The force

acting on *i*th atom is calculated by the gradient of the potential energy with respect to atomic displacements,

$$F_i = m_i a_i = m_i \frac{d^2 q_i}{dt^2} = -\nabla_{q_i} V(\boldsymbol{q_1}, \dots \boldsymbol{q_N}) = \left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right), \qquad (3.9)$$

where F_i is the force acting upon *i*th particle at time *t* and m_i is the mass of the particle, and $\boldsymbol{q_i} = (x_i, y_i, z_i,)$ is the position vector of *i*th particle, $V(\boldsymbol{q_1}, \dots \boldsymbol{q_N})$, is the potential energy of *N* interacting particles as a function of their positions. Thus, a potential energy function, or force field, needs to be defined for calculating $V(\boldsymbol{q_1}, \dots \boldsymbol{q_N})$ and its derivatives. The force field defines the forces between the particles and potential energies of each conformation. [33]

Compared with other molecular methods, such as molecular mechanics and Monte Carlo, the time evolution of the interacting molecules are explicitly computed in MD simulations, from which the macroscopic functional properties (such as kinetic energy, pressure and temperature) can be computed by time-averaging of the behaviour of individual particles.

3.1.4.2 STMD

Prior implementations of WL sampling have been based on MC. This limits the applicability of the method to more complex systems where effective MC moves are not available. A different approach, termed statistical-temperature molecular dynamics (STMD), was recently proposed by Kim *et al.*,[30, 31] with the objective of generating a deterministic trajectory by MD sampling and to augment the capabilities of the WL method to complicated systems beyond the scope of effective MC moves. To implement the variable *T(E)* into an MD simulation and realize a random walk in energy space, a

generalized ensemble technique[118] is employed in which the energy state is weighted by a non-Boltzmann factor. In the generalized ensemble, the sampling of configurations is based on the weight

$$w(E) = e^{-S(E)}, \tag{3.10}$$

which leads to the effective potential in the canonical ensemble at a temperature T_0 as

$$V_{eff} = T_0 S(E). \tag{3.11}$$

Then the force term $\tilde{f}$ can be solved by the derivative over position q:

$$\tilde{f} = -\frac{\partial V_{eff}}{\partial q}$$

$$= -T_0 \frac{\partial S(E(q))}{\partial q} = -T_0 \frac{\partial S(E(q))}{\partial E}\frac{\partial E(q)}{\partial q}.$$

$$= \frac{T_0}{T(E)} f \tag{3.12}$$

Thus, the sampling weight $e^{-S(E)}$ is realized in an MD simulation with the Nosé–Hoover thermostat by maintaining the kinetic energy at the reference temperature $T_0 = T_h$ and constantly adapting the forces with an energy-dependent scaling factor.

To implement STMD, all T_j are set to a constant value at the start of the simulation (usually T_h). The rest of the algorithm is similar to the implementation of STMC, including the temperature update in Equation 3.7 and integration for $Li(E)$ as in Equation 3.8.

3.1.4.3 *Validation of STMD by LJ fluid*

To firstly examine our implementation of STMD, we studied the LJ systems in the *NVT* ensemble, with the parameter settings matching the work of Kim *et al*;[31] the system consists of 110 particles, with reduced density as ρ=0.88, and the potential cutoff at 2.5σ; temperature range from [0.7,1.8], corresponding to a fluid region, with Nosé-Hoover

thermostat reference $T_0=1.8$. As in **Figure 3.5**, results for all the three different energy bin sizes (2, 4, and 16) agree with the reference.[31]
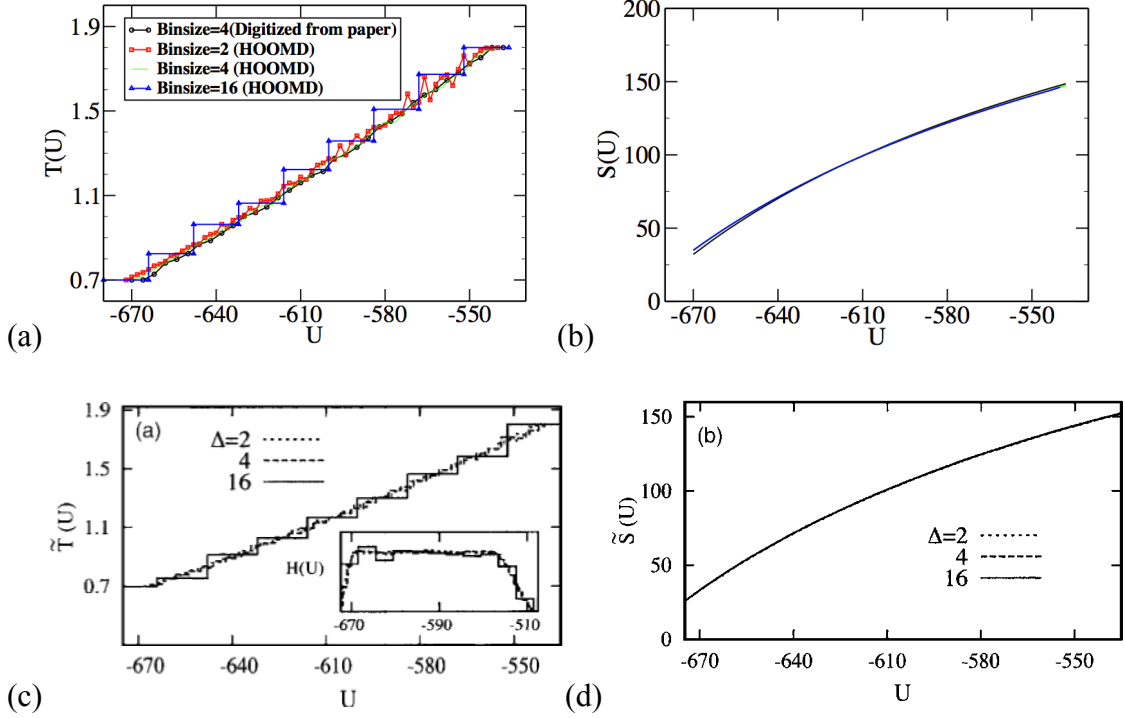
Figure 3.5 Lennard-Jones fluids simulated by STMD; (a) and (b) from this work,(c) and (d) from Kim *et al*.[31]

3.1.5   Hybrid WL Monte Carlo/ STMD method

To calculate the 2D DOS, we proposed the hybrid WL/STMD method, which is composed of original WL sampling and the standard STMD, taking the advantages of both of the algorithms. To illustrate the idea of the hybrid method, we firstly calculate 2D DOS as function of energy and the composition of the systems (i.e., in a semi-grand canonical ensemble), namely $g(E,N)$. Specifically, the simulation is performed in two directions (as shown in **Figure 3.6**). In the vertical direction, standard STMD simulations are performed as a function of energy with fixed system compositions, i.e., the number of all the species are fixed with a total of them as N (see details for the binary LJ system in

29

Section 7.2), and are responsible for calculating $g(E_i,N)$. At intervals, with a fixed energy state E, WL samplings are performed with a random walk of the second variable N_j, which is the number of the particles in one species, as shown in **Figure 3.6**. This idea follows from the ideas of replica exchange MD, but in this case swapping is governed by the WL acceptance criteria (i.e., DOS) rather than the Metropolis weight. Separate sets of histogram are maintained for each direction, e.g., for WL, at fixed energy bin, the simulation needs to be converged over N_j, while for STMD, at fixed N, they are converged over E_i. This allows for different bin sizes to be used for the WL part and STMD part, as appropriate.

$$H_{WL}[E, N_j] = H_{WL}[E, N_j] + 1, \qquad (3.13)$$

$$\ln g_{WL}[E, N_j] = \ln g_{WL}[E, N_j] + \ln f_{WL}, \qquad (3.14)$$

$$H_{STMD}[E_i, N] = H_{STMD}[E_i, N] + 1, \qquad (3.15)$$

After the simulation has converged according to the pre-defined final modification factor, the 2D DOS $g(E, N)$ can be obtained by connecting $\ln g_{STMD}[E_i, N]$ at different N with $\ln g_{WL}[E, N_j]$. Here, the initial $\ln f_{WL} = 1.0$, and $\ln f_{STMD} = 0.00025$, while the final $\ln f_{WL} < 10^{-6}$, and $\ln f_{STMD} < 10^{-8}$, following general setting of the two methods.[4, 31]
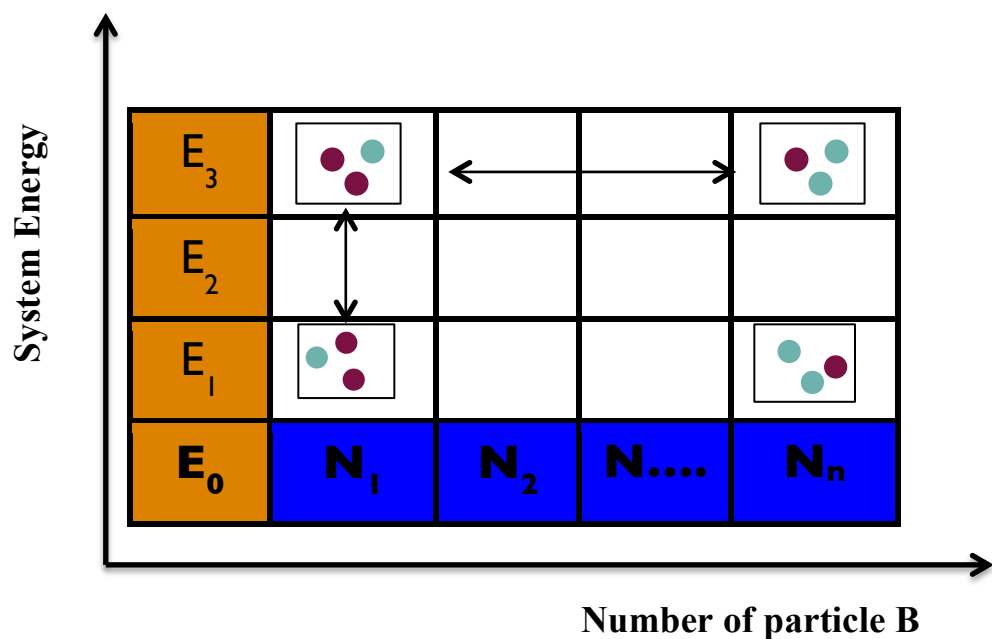
Figure 3.6 Schematic representation of hybrid WL/STMD method

3.2　Analysis methods

A number of methods are applied throughout this dissertation to analyze simulation results. Below is a description of the most important or commonly used methods.

3.2.1　Thermodynamic properties

Thermodynamic properties were calculated throughout the dissertation. After obtaining $g(E)$, the partition function Q can be calculated at any given temperature and other thermodynamic properties determined. The partition function is calculated in terms of the contribution from energy levels,[117]

$$Q(N,V,T) = \sum_E g(E)e^{-E/k_BT} , \qquad (3.16)$$

where k_B denotes Boltzmann's constant, T is the temperature, and $g(E)$ is the density of states, which is defined as the number of states that have the same energy level. The free energy, F, can be calculated from the partition function Q,[119]

31

$$F = -k_B T \ln Q(N, V, T),$$ (3.17)

and the average energy $\langle E \rangle$ and heat capacity C_v can be calculated as,

$$\langle E \rangle = \frac{\sum_E E g(E) e^{-E/k_B T}}{Q},$$ (3.18)

$$C_v = \frac{1}{k_B T^2} [\langle E^2 \rangle - \langle E \rangle^2],$$ (3.19)

where N is the number of molecules, E is the system potential energy and $\langle E^2 \rangle - \langle E \rangle^2$ is the fluctuation in the energy.

3.2.2   Order parameter

3.2.2.1 *Characterization of bilayer*

To quantify the structural properties and transitions within the self-assembled structures, several order parameters have been calculated in Chapter 5. The ordering of the lipid molecules is quantified by using the nematic order parameter, S2,[120] in which the average direction of each lipid is first quantified by calculating the moment of inertia tensor for each molecule:

$$I_{\alpha\beta} = \sum_{i=1}^{N_l} m_i (\hat{r}_i^2 \delta_{\alpha\beta} - r_{i\alpha} r_{i\beta}),$$ (3.20)

where $\hat{r}_i$ is the position vector of each particle i relative to the center of mass of a given molecule, and m_i is the mass of each particle i , $\delta_{\alpha\beta}$ is the Kronecker delta, N_l is the total number of particles in the lipid, and α and β are looping variables that correspond to the coordinate axes (i.e., *x, y, z*). The characteristic vector describing the lipid, $\hat{u}$, is the eigenvector associated with the smallest eigenvalue of the tensor, and is used to construct the nematic tensor:

$$Q_{\alpha\beta} = \frac{1}{N_L} \sum_{i=1}^{N_L} \frac{3}{2} u_{i\alpha} u_{i\beta} - \frac{1}{2} \delta_{\alpha\beta},$$ (3.21)

where N_L is the total number of lipids. The director of the system is the eigenvector associated with the largest eigenvalue, and the nematic order parameter, S2, corresponds to this largest eigenvalue. For a given simulation snapshot two Q tensors, one for each layer, were constructed and the two S2 values averaged. Using the same procedure, S2 for the individual segments of the lipids (the individual bead) is also calculated. All results are calculated as block averages, where quantities are averaged for energy blocks of size $\Delta E = 32$, corresponding to the energy bin size in the STMD simulations.

To quantify the transition of the water to a crystalline (i.e., frozen) structure, techniques based upon shape matching are employed.[121, 122] Briefly, the global structure of the water particles is determined via the super-position of the directions between a particle and its first neighbor shell (i.e., a bond-order diagram is constructed). Systems with a high degree of global ordering demonstrate strong correlations in terms of these "bond" directions. The bond-order diagram is transformed by taking the magnitude of the spherical harmonic transform for harmonics 4, 6, 8, 10, and 12, and concatenating these into a single vector, providing a characteristic "fingerprint" (i.e., shape-descriptor) of the system. To determine the transition, the dot product normalized between 0 and 1, is calculated between this vector and the vector describing a low temperature, ordered crystalline state (determined visually), this procedure is repeated using five different crystalline configurations as a reference to provide a more robust description. The order parameter is constructed such that an ideal match gives a value of unity.

3.2.2.2 Characterization of crystallized structure

In Chapter 6, a layer-by-layer global two-dimensional (2D) hexagonal order parameter (OP) of particles within the nano-pore was constructed by taking the Fourier transform

with frequency $\ell = 6$ of the superposition of the in-plane first neighbor shell surrounding atoms within a given layer; for detailed calculation of the order parameter, refer to references.[121, 122] Note that, unlike the related order parameter of Gubbins,[86] the 2d hexagonal order parameter used here takes into account the global orientational ordering of the system, not just the average local order, providing a clearer separation between ordered hexagonal and disordered states. For a full description of the difference between local and global order parameter construction, refer to Keys *et al.* [121, 122]

3.3   MD simulation package

The STMD algorithm was implemented within HOOMD-Blue[123, 124] simulation package. HOOMD-Blue is highly optimized to run MD simulation on graphical processing units (GPU). By implementing the STMD algorithm within this package, the algorithm can leverage the massively multicore nature of the GPUs for efficient calculation of the DOS. For the lipid system reported in Chapter 5, the STMD simulations run ~4-5 times faster on the NVIDIA GeForce GTX 480 GPUs than those on 8 core CPUs (2.4 GHz Opteron).

CHAPTER 4   LIPID BILAYER SELF-ASSEMBLY ON A 3D LATTICE MODEL

This chapter focuses on the applicability of WL method to the self-assembly of lipid molecules with a simple 3D-lattice model. The lipid models, with different chain lengths and various lattice sizes, are considered. Also conventional Metropolis MC simulations were implemented for comparison. This study provides an important first step towards the overall goal of developing a framework for the studying of phase behavior of complex fluids by density-of-states based methods. This work is published in Gai *et al*.[94]

4.1   Introduction

Amphiphilic molecules consisting of a hydrophilic head group and hydrophobic tail groups can display different fluid phases such as bilayers, micelles, and vesicles through self-assembly when mixed with a suitable solvent.[34, 125] Among them, lipid bilayers are important structures in both natural and artificial biological interfaces,[37] such as the membranes that form the basis of almost all life forms.[39] As a result lipid self-assembly has been the focus of both extensive computational and experimental research.[126, 127] While, it is generally accepted that lipid self-assembly can be qualitatively explained by the hydrophobic effect[54], complex details such as the necessary forces that direct lipids to assemble into predictable arrangements, or transient intermediate phase behavior during the aggregation process, are still the subject of much debate.[55-57] Furthermore, since self-assembly is driven by non-covalent interactions, it is necessary to explore the macroscopic environmental factors that can strongly influence self-assembly, such as concentration, the presence of other amphiphile species, and other state conditions.[39, 127]

During the self-assembly process, conformational changes correspond to changes in the system free energy. Free energy in general corresponds to the total amount of energy that can be converted to do work and is minimized when a system reaches equilibrium. Hence, to fully understand the self-assembly process and study equilibrium configurations, it is necessary to examine the underlying free energy behavior. The free energy, F, can be calculated from the partition function Q,[119]

$$F = -k_B T \ln Q(N, V, T) , \tag{4.1}$$

which in turn can be defined from an enumeration of the states that a system can occupy or in terms of the contribution from energy levels,[117]

$$Q(N, V, T) = \sum_s e^{-E_s/k_B T} = \sum_E g(E) e^{-E/k_B T} , \tag{4.2}$$

where k_B denotes Boltzmann's constant, T the temperature, s the system state with energy E_s, and $g(E)$ is the density of states, which is defined as the number of states that have the same energy level. Therefore, F can be determined either by evaluating the energy of all microscopic states, s, with a complete or at least importance sampling of the configuration space required, or by evaluating $g(E)$ for each possible energy level. The challenge in calculating Q is that traditional molecular dynamics or Monte Carlo simulation methods sample only the lower energy regions, while rarely visiting the higher energy regions, and thus can lock the simulation into a particular region of phase space, i.e., energy minima. Additionally, very low energy states may not be visited if an energy barrier exists. The time needed to ensure convergence of such simulations is prohibitively long and so advanced methods for calculation of the free energy are required.

To calculate F directly, the free energy can also be calculated based on the density of states $g(E)$ (Equation 4.2), which is not dependent on temperature in the canonical

distribution. Thus if $g(E)$ is known with high accuracy for all energies, the canonical distribution at any given temperature can be constructed and all thermodynamic quantities of interest, including the free energy, determined.[128] Unfortunately, this central quantity of interest is not known *a priori* and conventional MC in the Metropolis form does not allow for the direct calculation of $g(E)$. As described in Chapter 3, a breakthrough in this aspect is the WL method,[29] which does allow the direct calculation of $g(E)$. Here we apply both the original WL and STMD methods to calculate the free energy and thermodynamic properties of self-assembling lipid bilayers. We report a series of simulations utilizing the basic model proposed by Drefahl *et al.*[32] in order to demonstrate the applicability of these methods in understanding the self-assembly process.

4.2   Simulation details

4.2.1   Model

Drefahl and co-workers[32] studied on-lattice self-assembly using a lipid model that is described by a chain of 3 beads, with the first bead representing a hydrophilic head group (*h*) and the remaining beads hydrophobic tail (*t*) groups. The molecular interactions are purely repulsive between the non-bonded head and tail beads and between the tail beads and water and are equal in magnitude ($E_{ht} = E_{tw} = +20$ arbitrary units).[32] We note that the lipid model studied here is similar in spirit to the well-studied HP protein lattice model that is composed of a sequence of hydrophobic (H) and polar (P) monomers;[129] however, most HP model simulations are focused on the conformation of a single HP chain,[96, 130] and thus do not include semi-explicit solvent nor employ periodic boundary conditions, as in the current work.

The lipids are embedded in a simple cubic lattice of volume V and dimensions $L \times L \times L$ with each bead occupying one site (i.e., no two beads can occupy the same site), with the remaining lattice sites assumed to be water (w). Periodic boundary conditions are used in all directions. Two beads are considered nearest-neighbors if they are on adjacent sites in the lattice, so each site has six nearest neighbors in a 3D lattice. With these definitions, the volume fraction of lipids is given by

$$X = \frac{N_a \times n}{V},$$ (4.3)

where N_a is the number of molecules and n is the number of beads in a chain.

For the 3-segment chain system studied, three basic move types[35, 48] were used to generate successive configurations; reptation, interchange of tail and head groups and end-bond rotation, as illustrated in **Figure 4.1**, in a 70:20:10 percentage ratio, respectively. Reptation and interchange of tail and head groups provide more flexible changes to the molecular conformation and so a larger percentage of these moves are employed for efficient sampling.

Reptation   **Interchange**   **End-bond rotation**
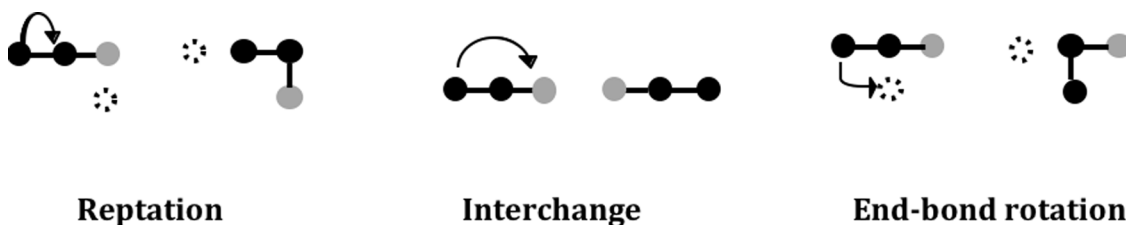
Figure 4.1 Schematic depiction of the Monte Carlo moves used in the simulations. Grey represents the head group bead and black the tail group beads.

In this work we have extended the original model to longer 5-segment chains, which consists of one head group and four tail beads. The free energy surface of longer

molecules is more complicated and thus more elaborate trial moves are required for their simulation. A simple configurational bias MC (CBMC) move for the WL algorithm,[131] which biases the selection of random sites to avoid the overlap of segments, was implemented in place of end-bond rotation in the simulations of the 5-segment system. The simple CBMC move involves cutting any bond between two beads randomly and then re-growing the chain by considering all possible trial orientations and choosing one position from the available trial positions with uniform probability.[132]

4.2.2 Wang-Landau parallel

As noted in Section *3.1.2*, to speed up the simulations, rather than simulating a single energy window, multiple energy windows[110, 111] are frequently used in WL simulations and have been adopted in this work. To ensure a random walk across the energy space, the size of the energy windows were carefully chosen. In general, although the simulations converge faster with smaller energy windows, they cannot be so small that they only cover a narrow energy range and potentially lock the system into a particular region of phase space. There should also be a suitably large overlap between adjacent windows that are to be stitched together. Additionally, since with this approach the random walk is restricted to certain energy ranges, in order to avoid errors in $g(E)$ at the edges of the sampled energy interval, whenever a proposed move is rejected because the energy is beyond the border of a given window the $g(E)$ for the current energy value is still updated, following the work of Schulz *et al.*[92] To determine the optimal energy windows and their overlap, simulations were performed with three different values for the number of energy bins per window (100, 150, 200) and with both 25% and 50% of overlap between adjacent windows on a lattice size of 20. From the calculated $g(E)$ and

heat capacities negligible differences were found between the different sets of simulations; boundary effects were also shown to be negligible for the current system with this approach. Furthermore, the results obtained were found to be reproducible from different initial configurations and agree with the Metropolis MC results (as discussed below). Thus, based on our extensive tests and the additional consideration of simulation efficiency, the results provided herein use moderately small windows (100 energy bins per window), with 25% overlap between adjacent windows and are the average of three independent simulations from different initial conditions. For the Metropolis MC simulations a typical run length was 2×10^7 MC cycles of equilibration followed by 2×10^7 MC cycles of production, while for the WL simulations the number of cycles ranged from 1×10^6 to 3×10^7, with the very low energy range requiring 10^8 cycles.

4.3 Phase behavior of 3-segment lipid model with WL algorithm

In the present work, extensive simulations to study the process of spontaneous bilayer formation of a simple 3D lattice model have been performed. To test the performance of the WL method and determine the appropriate values of the WL parameters, a lattice size of $L = 20$ with 266 3-segment molecules (corresponding to a concentration of 10 vol. %) was first studied and the results compared to those obtained from conventional Metropolis MC simulations. As can be seen in **Figure 4.2**, the results from the WL and conventional Monte Carlo methods are in good agreement. The results from the WL simulations (**Figure 4.2**a) show that the temperature dependence of the energy is smooth both in the high- and the low-temperature regions. In **Figure 4.2**b, a peak is obtained in the heat capacity plot from the WL simulation at $T = 19.62$, indicating a first order phase transition, which is consistent with the range 19.5 - 19.75 observed for the same phase

transition from the Metropolis MC simulations. Here, the transition we are studying is that from micelles to an ordered bilayer.

As noted by Landau *et al.*,[29] in the WL algorithm the accuracy of the $g(E)$ is largely controlled by the modification factor, which also determines the simulation efficiency (i.e., the necessary number of Monte Carlo steps). The ideal final $\ln f$ should be very close to 0.0 and in practice the random walk is usually stopped when $lnf < 10^{-6} \sim 10^{-8}$.[4, 99]
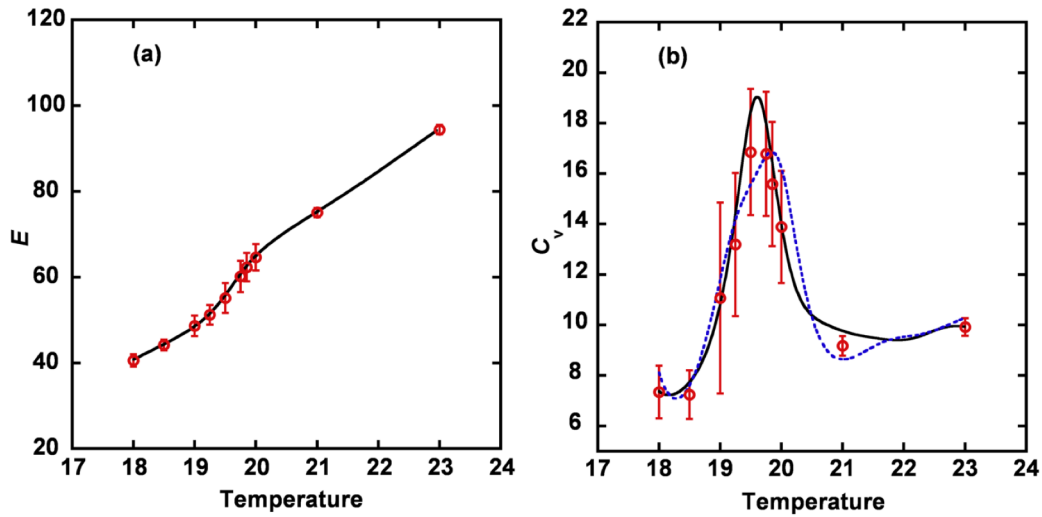
Figure 4.2 Comparison of the (a) energy per molecule, and (b) heat capacity obtained from the Metropolis MC and WL methods for the lattice size *L*=20. The red circles, black solid lines and blue dash lines correspond to results from the Metropolis MC, WL with lnf=10^{-8}, and WL with lnf=10^{-6}, simulations respectively.

We note that it has been shown that the convergence of $g(E)$ to the true value can stagnate for small $\ln f$ (referred to as saturation) and variants of the original WL method have been proposed, with mixed success, to overcome this problem.[89, 133, 134] In the current work, results obtained using $\ln f = 10^{-6}$ and $\ln f = 10^{-8}$ were compared, as

shown in **Figure 4.2**b, with the large difference in results indicating that the saturation point has not been reached thus far. Based on the Metropolis MC results, more accurate results are obtained with $\ln f = 10^{-8}$, which also indicates that the iterative refinement process from 10^{-6} to 10^{-8} is important. Hence, we use the parameter $\ln f = 10^{-8}$ in all of the simulations reported unless otherwise stated.

With our implementation of the WL method tested and optimized, the phase transition behavior for a larger lattice of $L = 48$ with 3686 3-segment molecules was then studied and the results compared to those presented by Drefahl *et al.*[32] using Metropolis MC simulations. With the WL method, the C_v values with a temperature interval of 0.01 have been calculated and a continuous curve with two apparent peaks obtained; one at T=19.68 and a second peak at T=20.40, indicating that a two-step phase transition process occurs (**Figure 4.3**). In the inset of **Figure 4.3**, C_v obtained from the WL simulations is compared to that obtained by Drefahl *et al.* at discrete temperature points.[32] From the figure we can see that while both methods yield similar behavior, only one phase transition is observed between T =19.75 and T = 19.85 in the work of Drefahl *et al.*[32] Clearly, a scan of the whole temperature range with the WL method is advantageous, disclosing more specific information (i.e., the intermediate phase behavior) about the system.
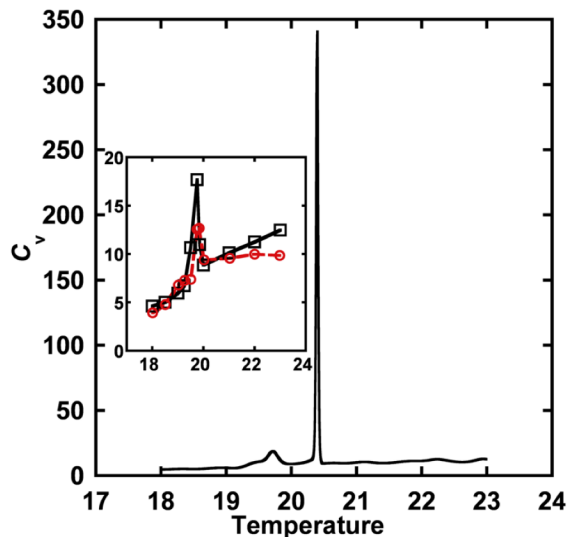
Figure 4.3 Heat capacity as a function of temperature for a lattice size of L=48. Inset shows the comparison between Metropolis MC[32] (red) and WL methods (black).

To demonstrate the phase transition visually, simulation snapshots are shown in **Figure 4.4** and **Figure 4.5** over typical energy ranges (i.e., corresponding to different temperatures). For $L = 48$, at low temperatures one bilayer spanning the whole simulation box is observed and a second smaller bilayer is also found (**Figure 4.4**a). As the temperature is increased, the second bilayer melts into disordered clusters (**Figure 4.4**b) until at higher temperatures further still, the bilayer spanning the whole simulation box also disappears and only clusters are observed (**Figure 4.4**c). For the $L = 20$ system, at low temperatures only one bilayer is formed throughout the simulation box (**Figure 4.5**a), which as temperature is increased transforms to clusters directly without an intermediate phase (see **Figure 4.5**b and **Figure 4.5**c). From the comparison of the behavior observed for the $L = 20$ and 48 lattices, we can draw the preliminary conclusion that the ratio of the number of chains to the number of lattice sites in a plane, i.e., area per lipid, affects the phase transition temperature T_c. At small ratios, when the number of chains is not

sufficient to span the entire lattice plane, only one phase transition from a bilayer to clusters occurs. At higher ratios, when the number of chains is sufficient to form an additional bilayer, the phase transition occurs in two steps; first the small bilayer breaks into smaller clusters, and then the larger bilayer breaks down into clusters. This larger bilayer melting into clusters results in larger energy fluctuations in the system, explaining the very large value of C_v for the second-step phase transition.
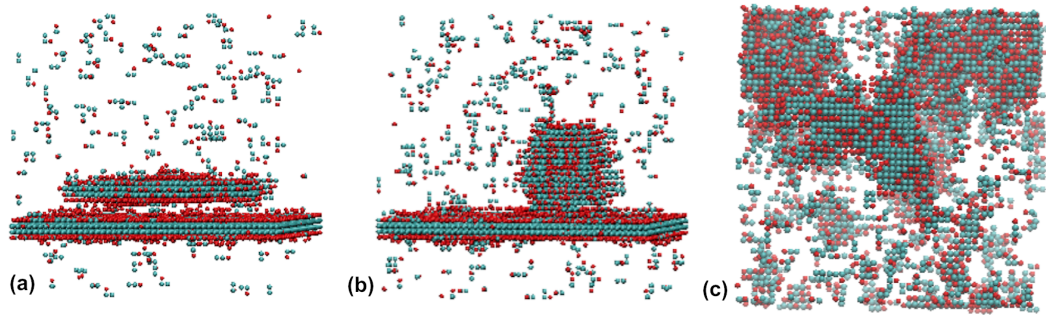
Figure 4.4 Snapshots of system configurations from simulations of a lattice size of $L=48$ illustrating (a) a bilayer structure in the low energy range ($E/N =36.0$), (b) a cluster and bilayer mixture in the middle energy range ($E/N =46.7$), and (c) clusters in the high energy range($E/N =84.7$).
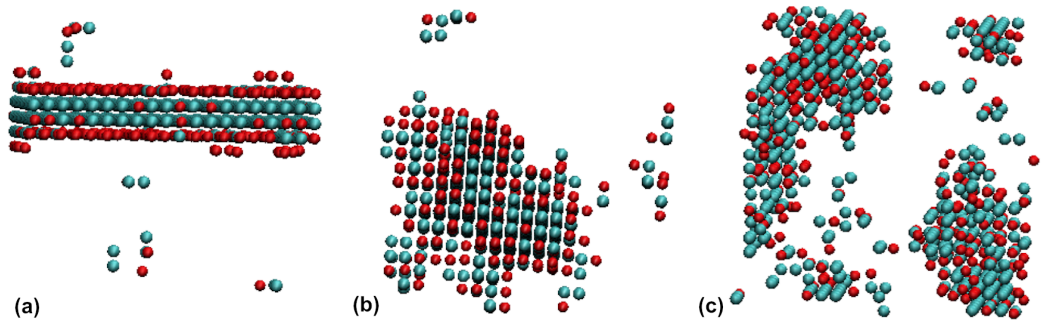
Figure 4.5 Snapshots of system configurations from simulations of a lattice size of $L=20$ illustrating (a) a bilayer structure in the low energy range ($E/N =41.2$), (b) clusters in the middle energy range ($E/N =68.4$), and (c) clusters in the high energy range ($E/N =83.6$).
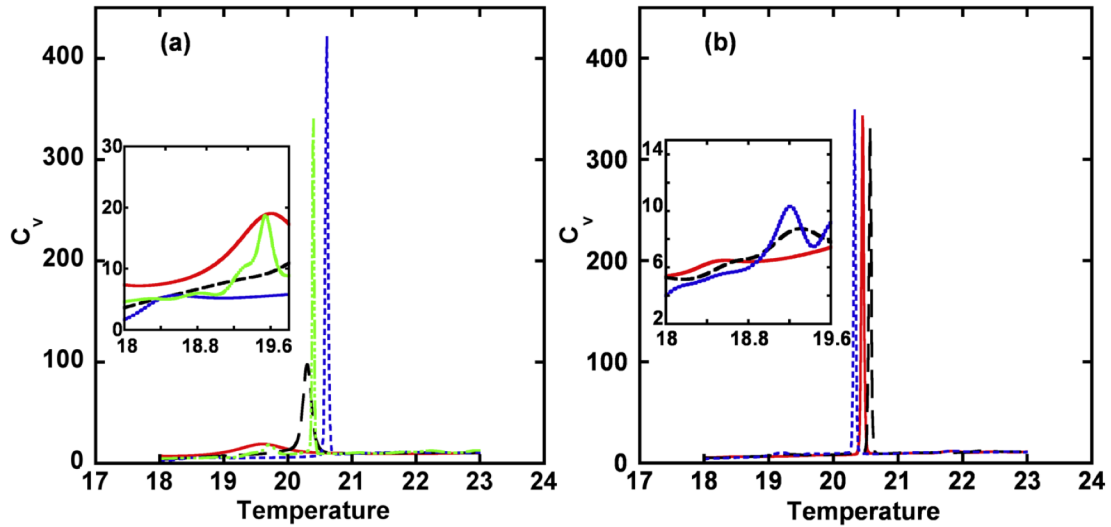
Figure 4.6 Heat capacity as a function of temperature for 3-segment amphiphiles on lattices of different sizes. (a) red, *L*=20; black, *L*=28; blue, *L*=36; green, *L*=48, and (b) red, *L*=40; black, *L*=42; blue, *L*=44.

Based on the observations described above, a system size search was performed to determine the critical ratio at which the two-step phase transition starts. Several different lattice sizes in the range of *L* = 20 - 48 were studied with the same molecule concentration of 10 vol. %. The heat capacity curves obtained are presented in **Figure 4.6**, from which we can see that the phase transition temperature T_c shifts as the system size increases and two peaks begin to appear for lattice sizes of *L* = 42 with 2470 lipids (**Figure 4.6**b) and above, when the ratio equals ~1.4.

4.4   Phase behavior of 5-segment lipid model with WL algorithm

We have also extended the lipid model to simulate longer 5-segment chains using a lattice size of *L* = 20. For the efficient sampling of the longer chains, a simple CBMC move type was implemented as described in section 2.3. To verify the CBMC algorithm, simulations for the 3-segment model and *L*=20 were first performed. As shown in **Figure**

4.7, comparisons between the WL and Metropolis MC simulations with and without CBMC results indicate good agreement.
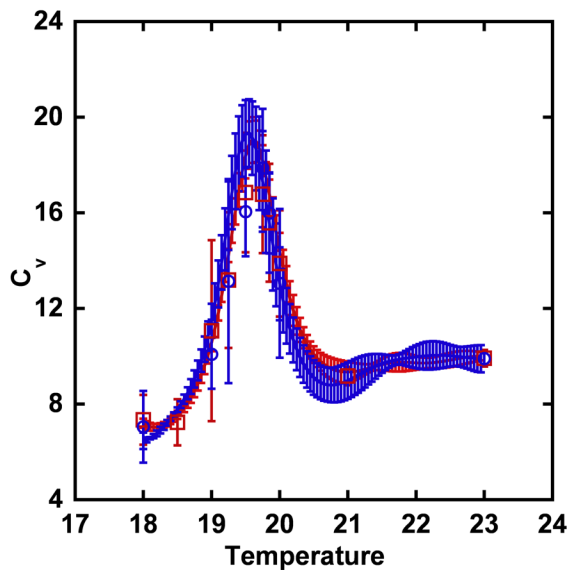
Figure 4.7 Comparison of heat capacities obtained from WL simulation without CBMC (red line) and with CBMC (blue line) and from MC without CBMC (red squares) and with CBMC (blue circles).

Preliminary tests showed that for lipids with longer tail groups using the same concentration of 10 vol% as in the simulations of the 3-segment model, or keeping the same number of lipid chains (N=266, which is 17 vol%), always results in the formation of closed vesicles, even at low temperature, rather than the bilayer sheet of interest. Based on this, and our observations of the effect of the area per lipid, a higher concentration of 31.25 vol% (500 chains) was chosen for study and a one-step phase transition observed as temperature is reduced, as shown in **Figure 4.8**. The peak in the heat capacity curve around $T = 19.16$ indicates the phase transition temperature. Thus, we have successfully applied the WL algorithm to longer amphiphilic chains; however, as the length of the lipid chain and the system size increases, the number of energy bins

required increases due to the extensive nature of the energy and discrete representation of the density of states. In this work this resulted in a factor of 5 increase in simulation time for the 5-segment lipid system over the analogous 3-segment system.
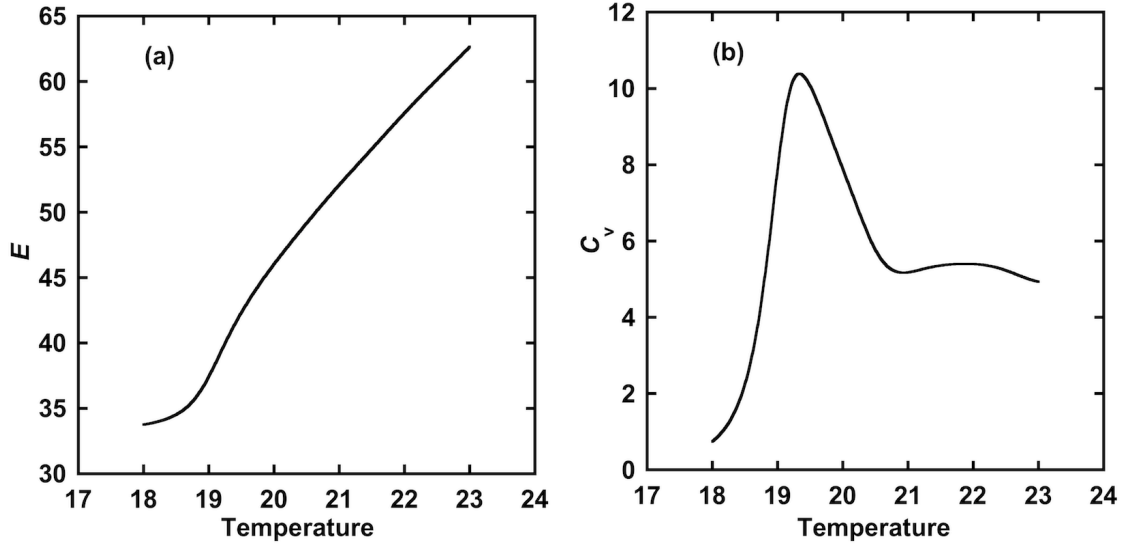
Figure 4.8 (a) Energy per molecule, and (b) heat capacity per molecule as a function of temperature for the 5-segment lipid model in a lattice of size $L=20$.

4.5 3-segment lipid model with STMC algorithm

The STMC algorithm is applied to the 3-segment lipid system on the $L = 20$ cubic lattice for direct comparison to the WL method. To implement the STMC simulations efficiently, the strategy of multiple energy windows is again employed, with four different energy bin sizes considered ($\Delta E = \Delta_{WL}$, $2\Delta_{WL}$, $4\Delta_{WL}$, $8\Delta_{WL}$, where Δ_{WL}, is the interval between neighboring energy levels in the original WL method), and an initial value of $\ln f = 5\times10^{-6}$. The simulations are considered to be converged when the final modification factor is smaller than 10^{-8}.

Similar phase transition phenomenon (shown in **Figure 4.9**) and system structures (not shown) are obtained as that from the WL simulations. In **Figure 4.9**a, we can see that the $T(E)$ values appear to be largely independent of the bin sizes ΔE and all four sets of values have the same shape as those calculated from the WL simulation. With the continuous density of states integrated from the convergent $T(E)$, we can again calculate any thermodynamic properties of interest and present results for the heat capacity in **Figure 4.9**b. While there are small differences between the WL and STMC results at high energies, the simulation results are all within the error bars. While the STMC method helps avoid the increase in the number of energy bins (proportional to the number of sites) as the system size increases seen with the original Wang-Landau method, in practice, the time needed for a STMC step is ~ 3 times as long as that for a WL step on the same system (which can be seen from the algorithm, as more calculations are done in the STMC method at each step). Since the STMC method is capable of utilizing simulations with larger energy bin sizes (i.e., fewer energy bins are required for histogram flatness), the simulations can be converged with fewer MC steps than the WL method and so can compensate for the slower speed of an individual step. For this system, the overall efficiencies are similar for the WL and STMC methods for $\Delta E = 4\Delta_{WL}$, while the STMC method for $\Delta E = 8\Delta_{WL}$ is faster than WL; however, as the system size increases, the simulation time required increases significantly. For example, preliminary results indicate that while the STMC algorithm achieves ~34000 MC cycles/min for $L = 20$ this decreases to only ~ 20000 MC cycles/h for $L = 48$, making the study of the larger lattice via STMC impractical. For comparison the WL method is still robust for $L = 48$ and achieves ~520000 MC cycles /h.
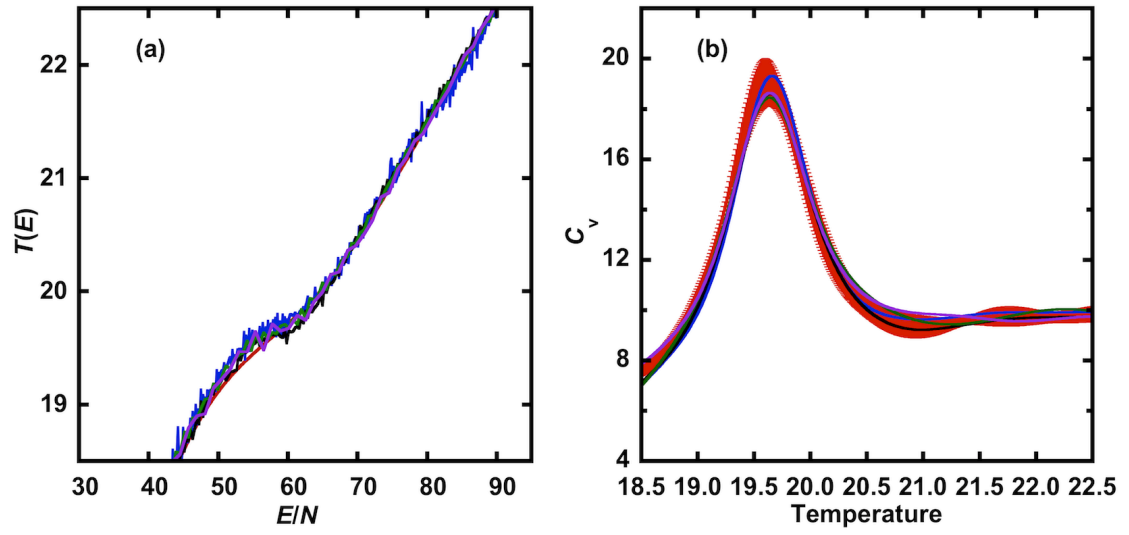
Figure 4.9 Comparison of the WL and STMC algorithms for lattice size L=20 (a) convergent $T(E)$, (b) heat capacity, red, WL; blue, STMC with bin size=Δ_{WL}; black, STMC with bin size=2Δ_{WL}; green, STMC with bin size=4Δ_{WL}; violet, STMC with bin size=8Δ_{WL}.

CHAPTER 5   PHASE TRANSITION BEHAVIOR OF AMPHIPHILIC LIPIDS IN SOLUTION

In this chapter, two different techniques - REWL and STMD[30,31] - were applied to systematically study the phase transition behavior of self-assembling lipids as a function of temperature using an off-lattice lipid model, both allowing direct calculation of the density of states with improved efficiency compared to the original WL method. The phase behavior of the lipid molecules with respect to bilayer formation has been characterized through the calculation of the heat capacity as a function of temperature, in addition to various order parameters and general visual inspection. The simulations conducted by both methods can go to very low temperatures with the whole system exhibiting well-ordered structures. With optimized parameters, several bilayer phases are observed within the temperature range studied, including gel phase bilayers with frozen water, water (i.e., frozen and liquid water), and liquid water and a more fluid bilayer with liquid water. These results provide a more complete understanding of the lipid phase transitions, with improved efficiency. The results obtained from both methods, STMD and REWL, are consistently in excellent agreement with each other, thereby validating both the methods and the results. This work is published in Gai *et al*.[135]

5.1   Introduction

Generally speaking, amphiphilic systems have rough energy landscapes, exhibiting many minima close to each other separated by large energy barriers.[11] These features can cause problems for the numerical exploration of their phase behavior as a function of temperature using standard MD or MC methods, for example, since some energy states

may not be visited due to the energy barriers. Thus, conventional MD or MC simulations can become locked into a particular region of phase space (i.e., sample only small regions), and the results may depend on the initial conditions and thermodynamic pathway of the simulation.[118]

The time needed to ensure convergence of conventional simulations is often prohibitively long and thus various methods have been proposed to improve the rough energy exploration by going beyond standard MD or MC methods.[4, 118] Among them, the as described in Chapter 3, the WL method,[4] which performs a random walk over the whole energy range, allows sampling of the conformations at all relevant temperatures in a single simulation. Compared to other flat-histogram methods, this dynamic update of $g(E)$ on the fly can provide quicker exploration of phase space,[136] and once the $g(E)$ is known, all thermodynamic and structural quantities can be calculated at any temperature. The WL method has been applied to study a number of different systems, focusing primarily on simple models for polymers and proteins.[4, 89, 99, 133, 134] When applying the WL algorithm to more complex systems, convergence problems[30, 31] can be encountered due to the extensive nature of the energy and discrete representation of the density of states on an energy grid, which results in the number of energy bins required increasing substantially with system size.

To improve the sampling efficiency, a parallel replica-exchange framework for the WL sampling scheme, described in Section *3.1.2*, can be employed.[112] In this approach individual simulations with overlapping energy distributions are carried out and replicas are allowed to move between the individual simulations.[26, 137] A different approach, STMD, was recently proposed by Kim *et al.*,[30, 31] with the objective of

generating a deterministic trajectory by MD sampling and to augment the capabilities of the WL method to complicated systems beyond the scope of effective MC moves. The STMD algorithm, describe in Section *3.1.4,* implements the idea of WL sampling into MD simulation using the generalized ensemble *via* the statistical temperature. This is achieved by relying on the thermodynamic relationship between the statistical temperature $T(E)$ and the density of states $g(E)$; STMD achieves a flat energy distribution by dynamic modification of $T(E)$. By updating the *intensive* variable $T(E)$, a continuum description of $g(E)$ can be obtained through analytical integration of $T(E)$; therefore, moderately larger bins, and hence fewer bins, can be used to accelerate the simulation convergence, while still maintaining the statistical accuracy. [30, 31]

Here we have applied both the REWL and STMD methods, which are very different techniques, to calculate the free energy and thermodynamic properties of self-assembling bilayers using an off-lattice lipid model. Since the systems and phenomena being studied are of a complex nature, they hold the potential for revealing hidden algorithmic difficulties, which might not be obvious by using a single method. We report a series of density-of-states simulations by both the REWL and STMD methods to systematically study the phase transition behavior of self-assembling lipid molecules as a function of temperature based on the CG models of Goetz *et al.*[64] and Fujiwara *et al.*[65] From the density of states, we determine a number of thermodynamic properties (average energy and heat capacity) to characterize the phase transition, in addition to visual inspection and order parameter-based analysis. Together they provide a more complete understanding of the lipid phase transitions.

5.2   Simulation details

We studied continuous lipids model solvated in water. The amphiphilic molecules are also modeled as short linear chains composed of a hydrophilic head bead and two hydrophobic tail beads (**Figure 5.1**). The water molecules are represented as hydrophilic beads. The total number of particles in the system is fixed at $N = 1000$ and the number of amphiphilic molecules M is varied. The volume of the simulation box is chosen such that the number density, ρ, is maintained at $\rho = 0.8$.
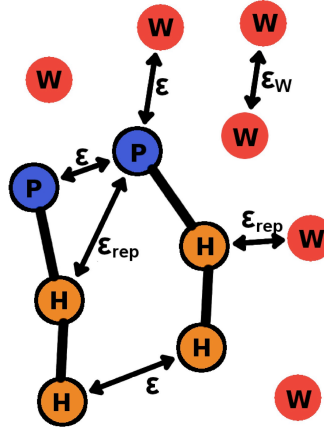
Figure 5.1 Schematic of the model amphiphilic lipid and solvent molecules studied. P: polar head beads, H: hydrophobic tail beads, W: solution (water) beads.

All non-bonded particles of the same type as well as the solvent and hydrophilic head beads interact via the LJ potential:

$$U_{\mathrm{LJ},\varepsilon}(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right]. \tag{5.1}$$

The interaction between tail beads and head or solution particles is purely repulsive:

$$U_{\mathrm{rep}}(r_{ij}) = 4\varepsilon_{\mathrm{rep}} \left(\frac{\sigma_{\mathrm{rep}}}{r_{ij}}\right)^{9}, \tag{5.2}$$

where r_{ij} is the distance between two particles i and j. The van der Waals radii, σ and σ_{rep}, are set to 1.0 and 1.05 respectively, and $\varepsilon = \varepsilon_{rep} = \varepsilon_w = 1.0$, following Fujiwara *et al.*[65] A version of the model is also studied in which the interaction between water

particles ε_w is reduced to 0.7. Bonds between monomers in the lipid molecules are finitely extendible, non-linear elastic (FENE) and are modeled by the corresponding FENE and WCA potentials[138, 139]:

$$U_{\text{FENE}}(r_{\text{b}}) = -\frac{k}{2}R^2\ln\left(1 - \left(\frac{r_{\text{b}}}{R}\right)^2\right) + U_{\text{WCA}}(r_{\text{b}}) , \qquad (5.3)$$

with

$$U_{\text{WCA}}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \varepsilon , & r < 2^{1/6}\sigma \\ 0 , & r \geq 2^{1/6}\sigma \end{cases} , \qquad (5.4)$$

For conformational updates a combination of local displacements of single particles in Cartesian coordinates and reptation moves for the amphiphilic molecules are used. As illustrated in **Figure 5.2**, in the reptation move a molecule and one of its ends are selected at random. Then, from all the neighboring solution particles (i.e., particles within the maximal allowed bond length R, Equation 5.3), one is chosen randomly and connected to the molecule, while the bond to the monomer at the other end is broken (**Figure 5.2**b). Finally, the types of the monomers are restored to create a valid conformation (**Figure 5.2**c). Clearly, this move introduces a bias, as there is, in general, a different number of potential neighbors for the forward and backward moves. To fulfill detailed balance, this bias is corrected by an appropriate weight factor b in the acceptance probability as shown by Equation 3.4.

During one Monte Carlo sweep N individual updates are made, of which *3M/10* are reptation moves.
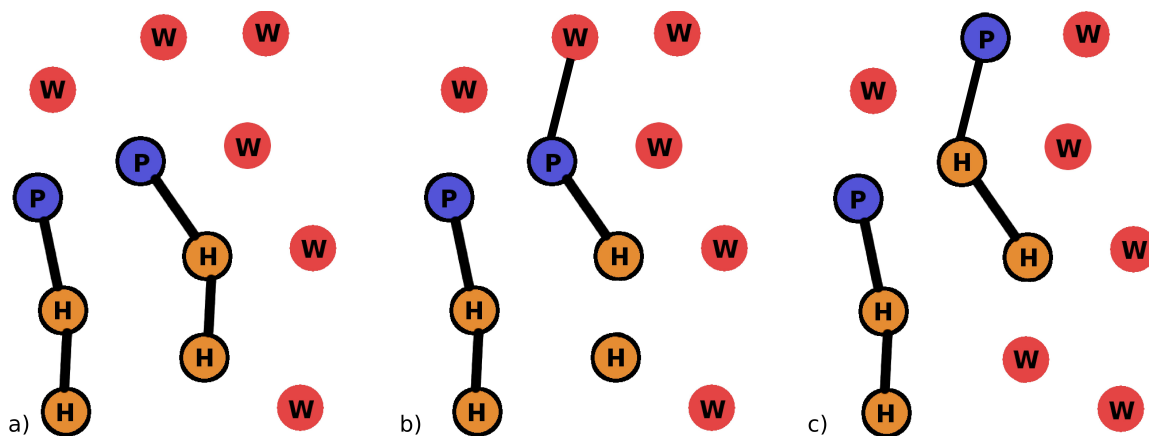
Figure 5.2 Illustration of the applied reptation move for conformational update during the Wang-Landau sampling. a) shows the initial state, c) the final one. Note that the Cartesian coordinates do not change during the trial move.

5.3   Phase transition behavior of system with M = 125

The self-assembly of amphiphilic lipids solvated in water at different concentrations has been simulated by both STMD and REWL MC in this work. As discussed earlier, the model studied is similar to that employed by Goetz *et al.*[64] and Fujiwara *et al.*,[65] however, previous work focused on the structural properties of the various aggregates, such as spherical micelles, cylindrical micelles and bilayers, via the variation of the amphiphilic concentration and hydrophobic interactions using standard MC and MD simulations. Additionally, these previous studies relied only on visual inspection of simulation snapshots and calculated order parameters at discrete temperatures to characterize the different structures. In this work, we investigate the phase transition behavior of the bilayer structures through their thermodynamic properties as a function of temperature in addition to structural and visual analysis. All T_j are set to a constant value at the start of the simulation (usually T_h).

In both the STMD and REWL simulations a wide temperature range from $T_l = 0.4$ to $T_h > 6$ was first tested, in order to cover a sufficient energy range corresponding to both fluid and gel regions. For the STMD runs, an energy bin size of $\Delta E = 32$ was used with an initial modification factor $\ln f = 0.00025$. The entire temperature range was split into three windows, $T \in [0.4, 1.0]$, $T \in [0.6, 4.0]$ and $T \in [3.5, 6.5]$, which were independently simulated with time steps of 0.01, 0.01 and 0.001, respectively. For example, for the system with M=125 lipids, we found that it took approximately 4.7×10^6, 4.2×10^7 and 3.1×10^6 MD steps respectively for the three windows to reach the lowest temperatures for the first time and was then followed by uniform sampling of the temperature range. The total simulation time for each energy window, $T \in [0.4, 1.0]$, $T \in [0.6, 4.0]$ and $T \in [3.5, 6.5]$, was 4.0×10^9, 2.0×10^8 and 4.0×10^8 MD steps, respectively, with a final WL modification factor of $\ln f < 5 \times 10^{-7}$. Even though results from STMD and REWL agree over a wide range of temperature, for the results presented henceforth we focus on temperatures less than $T_h \lesssim 1$, as this range encompasses the bilayer phase transitions of interest.

The evolution of the potential energy E in the first two windows for the lipid system is shown in **Figure 5.3**, each of which displays a walk with frequent sweeps in the energy range, which corresponds to the restricted temperature range applied. Once the simulations are converged, the density of states $g(E)$ is integrated from the convergent statistical temperature $T(E)$.
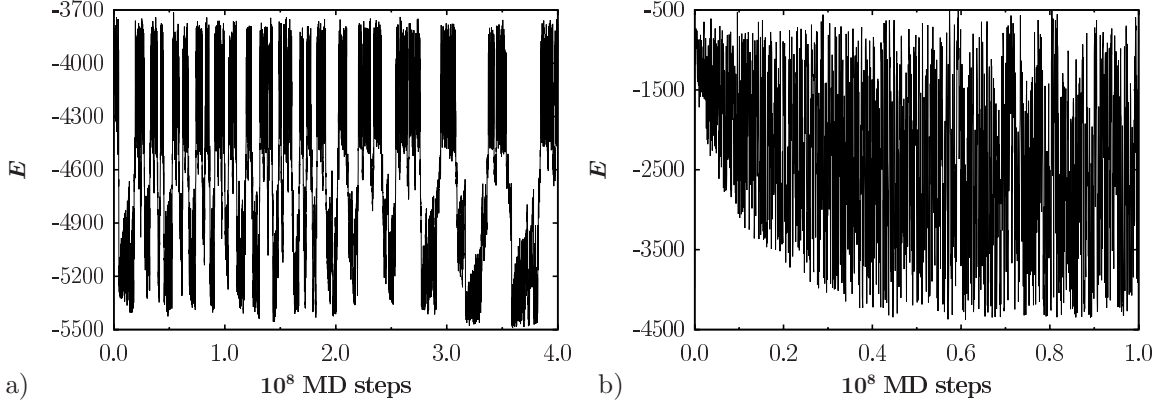
Figure 5.3 Energy trajectory from STMD simulation of $M = 125$ lipid system with ε_w =1.0 for a) $T \in [0.4,1.0]$ and b) $T \in [0.6,4.0]$.

For the REWL simulations, preliminary scans were performed in the energy ranges $E \in [-5200, 3500]$ (which corresponds to a temperature range of approximately $T \in [0.5, 8.5]$) and $E \in [-5400, 0]$ ($T \in [0.4, 5]$). For these preliminary runs, a single walker per energy subinterval and the Wang-Landau flatness criterion proposed by Zhou and Bhatt[140] was used in order to obtain faster results. The energy bin size for these runs was $\Delta E = 5$. Replica exchanges were attempted every 10^5 MC sweeps. **Figure 5.4** shows examples of two energy trajectories and the corresponding walks through energy windows for both of these runs. While the transition from completely mixed and disordered configurations into a fluid bilayer (**Figure 5.4**a) does not present a significant obstacle for the algorithm, **Figure 5.4**b illustrates that the fluid-gel bilayer-bilayer transition is hard to overcome. Following the preliminary scans, multiple independent simulation runs with three independent Wang-Landau walkers per energy subinterval were performed in the energy range $E \in [-5350, -3350]$, corresponding to the temperature range of the bilayer phases ($T \in [0.45, 1.4]$). Furthermore, the originally proposed, stricter 80%-flatness criterion[4, 29] and an energy bin size of $\Delta E=1$ was used. Up to a

maximum of 10^8 Monte Carlo sweeps (one sweep corresponds to one update of each particle in the system, on average) per walker were used to converge the estimator for the density of states, followed by a production run consisting of 3×10^7 MC sweeps for each walker. Error bars for all observables were obtained from 200 iterations of bootstrap resampling[141] of the multiple density of state pieces from the REWL runs. Confirming the quality of the WL weight determination, the estimated density of states at that point and the measured density of states in the production run were within mutual error bars.
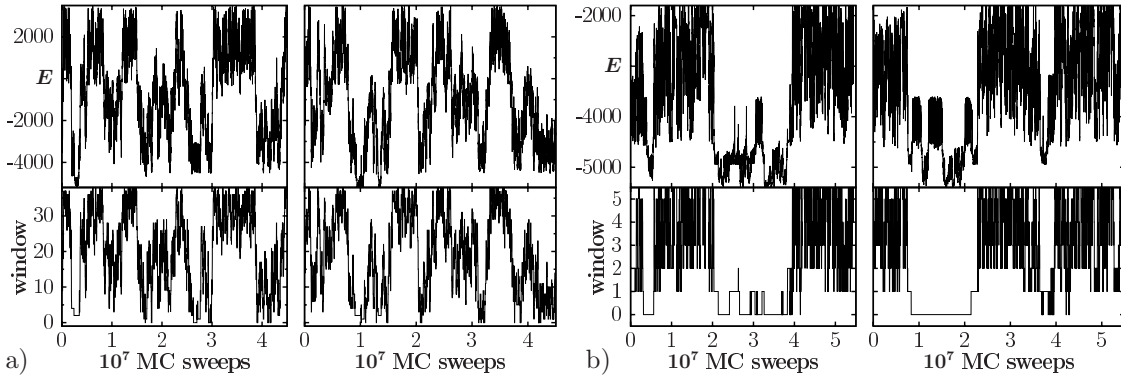
Figure 5.4 Trajectories through energy space (top) and energy windows (bottom) from REWL preliminary scan for the same system studied in Fig. 4. a) Sampling of energy range $E\in[-5200,3500]$, covering the clustering of the lipid molecules and the bilayer formation. b) Sampling of energy range $E\in[-5400,0]$, focusing on the fluid-gel bilayer-bilayer transition. For each case, walks of two single replicas out of 39 (a) and 9 (b) are shown as examples.

From both methods, STMD and REWL, thermodynamic properties such as the average energy and heat capacity at constant volume were then calculated from the obtained density of states $g(E)$, with a temperature step <0.001, providing a systematic analysis of the phase transition behavior as summarized in **Figure 5.5**.

Figure 5.5 Density of states and thermodynamic properties of lipid system with $M = 125$ and $\varepsilon_w=1.0$; comparison of data from STMD (symbols) and REWL (solid line) (a) logarithm of density of states, (b) Energy-Temperature relations, comparison between the convergent energy (dots) and reweighted energy (straight line). The inset shows the comparison between the reweighted energy from STMD and the mean energy calculated from REWL results. (c) Heat capacities calculated from STMD and REWL results. (d) Microcanonical analysis. As an example, $\beta=1.950$ on left scale agrees with the peak temperature of $T=0.513$ in (c) via the relation $\beta=1/T$.
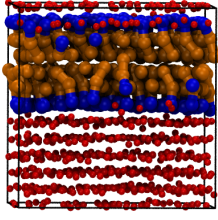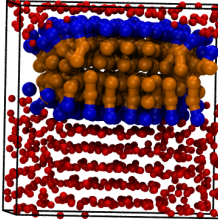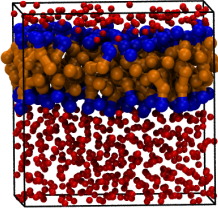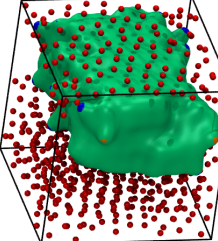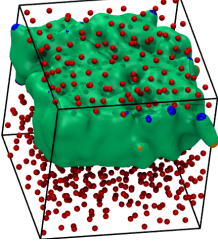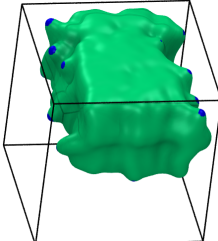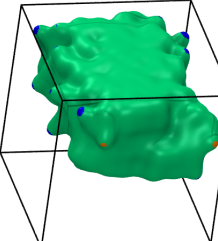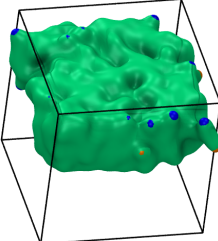
We first note that excellent agreement is obtained for the density of states calculated from both the STMD and REWL methods. Secondly, we find that there are two jumps in the average energy between 0.5 and 0.6 in the temperature (**Figure 5.5**b); we also note from **Figure 5.5**b good agreement between the inverse reweighted average energy calculated from $g(E)$ and the convergent $T(E)$ obtained directly from STMD. The changes in the average energy are better marked quantitatively by the two peaks observed

in the C_v curves shown in **Figure 5.5**c. The small energy jump at lower T manifests in a peak in C_v at $T = 0.513$ and the larger energy jump in the peak at a higher temperature of $T = 0.535$. These peaks in C_v indicate structural phase transitions in the bilayer system. Simulation snapshots are provided in **Table 5.1** that demonstrate the three different bilayer phases observed. Note that at very high temperature $T \gg 1$, the lipid molecules GB1 become more hydrated and increased disorder is seen in the lipid tails (denoted GB2). During the main transition, the gel bilayer phase concurrently with the water undergoes a transition to a fluid bilayer phase (denote FB) in which the tails are disordered without any tilt, concurrently with the water melting. This phase covers a wide temperature range for the current system and is physiologically generally the most important phase of the bilayer.

As both the water and bilayer phase transitions occur at temperatures very close to each other and to further demonstrate the extremely good agreement between the results obtained from the STMD and REWL methods, an additional microcanonical analysis,[142, 143] which is based on the same relation used in STMD and given in Equation 3.5 was carried out on the density of states obtained from REWL. The results are shown in **Figure 5.5**d and compared to the analysis of the heat capacities (**Figure 5.5**c). As can be seen from the figure, the peaks in γ corresponding to the phase transitions are well separated and the microcanonical temperature of the transition corresponding to the lower peak in the heat capacity from STMD at $T = 0.513$ is measured as $T = 1/\beta = 1/(1.950 \pm 0.003) = 0.513 \pm 0.001$. This essentially perfect agreement indicates once more that reliable results are obtained for the simulation of the self-assembly amphiphilic

molecules. The complete comparison for all transition temperatures is discussed and presented separately in **Table 5.2** below.

Table 5.1 Simulation snapshots of the different bilayer structures seen for the $M = 125$ lipid system with $\varepsilon_w = 1.0$. From left to right the phases observed as temperature increases are denoted as follows: gel phase bilayer 1 with frozen water (GB1+FW); gel phase bilayer 2 with mixed water (i.e., frozen and liquid water, GB2+MW); gel phase bilayer 3 with liquid water (GB3+LW); and fluid bilayer with liquid water (FB+LW).

GB1+FW	GB2+MW	GB3 +LW	FB+LW
		-	
		-	
		-	

5.4   Phase transition behavior at different ε

5.4.1   Thermodynamics properties

Using an equal strength for the interaction energy of the water and lipids (i.e., $\varepsilon_w = \varepsilon = 1.0$), the gel-fluid phase transition of the lipids and the frozen-liquid water transition occur almost concurrently in temperature. In order to examine the phase transitions of the lipids and solvent separately, and determine all of the phases that may be present during

the self-assembly process, a lower energy constant for the water-water interaction has also been studied.

Table 5.2 Comparison of transition temperatures obtained from peak positions in the heat capacities (STMD and REWL[a]) and by microcanonical analysis of the density of states obtained by REWL[b]. It is known that for strong first-order like transitions in finite systems, canonical and microcanonical transition temperatures might[144, 145] differ, as can be seen for transition #2, for example. All other temperature values are in good agreement.

Energy constant	Transition	STMD	REWL[a]	REWL[b]
$\varepsilon_w = 1.0$	#1	0.513	0.5147 ± 0.0001	0.5128 ± 0.0009
	#2	0.535	0.5351 ± 0.0001	0.519 ± 0.006
$\varepsilon_w = 0.7$	#3	0.404	0.4032 ± 0.0001	0.401 ± 0.002 0.408 ± 0.002
	#4	0.427	0.4273 ± 0.0001	0.4277 ± 0.0003 0.433 ± 0.002
	#5	0.533	0.532 ± 0.001	0.533 ± 0.01

[a]Transition temperatures obtained from peak positions in the heat capacities (STMD and REWL).
[b]Transition temperatures obtained by microcanonical analysis of the density of states obtained by REWL.

The thermodynamic data including the average energy and C_v obtained from simulations with $\varepsilon_w = 0.7\varepsilon$ are presented in **Figure 5.6**. In **Figure 5.6**a, the logarithm of the density of states calculated from both the STMD and REWL methods is presented and can be seen to agree well. From **Figure 5.6**b, good agreement between the reweighted average energy and convergent $T(E)$ from STMD is again also observed, with a more gradual transition in temperature seen for $\varepsilon_w = 0.7$ than for $\varepsilon_w = 1.0$. Three energy changes are observed, each correlated to a transition temperature, which manifest in C_v peaks at $T = 0.404, 0.427,$ and 0.533. These also agree very well with the transition temperatures obtained from the microcanonical analysis as reported in **Figure 5.6**d and **Table 5.2**. Furthermore, the microcanonical analysis reveals more clearly some pre-transitions that are not easily detected through analysis of the heat capacity peaks as they

are represented as shoulders at slightly higher temperatures than the main transition peaks. To obtain the error bars in C_v from the STMD simulations, three independent simulations were run from different initial conditions.
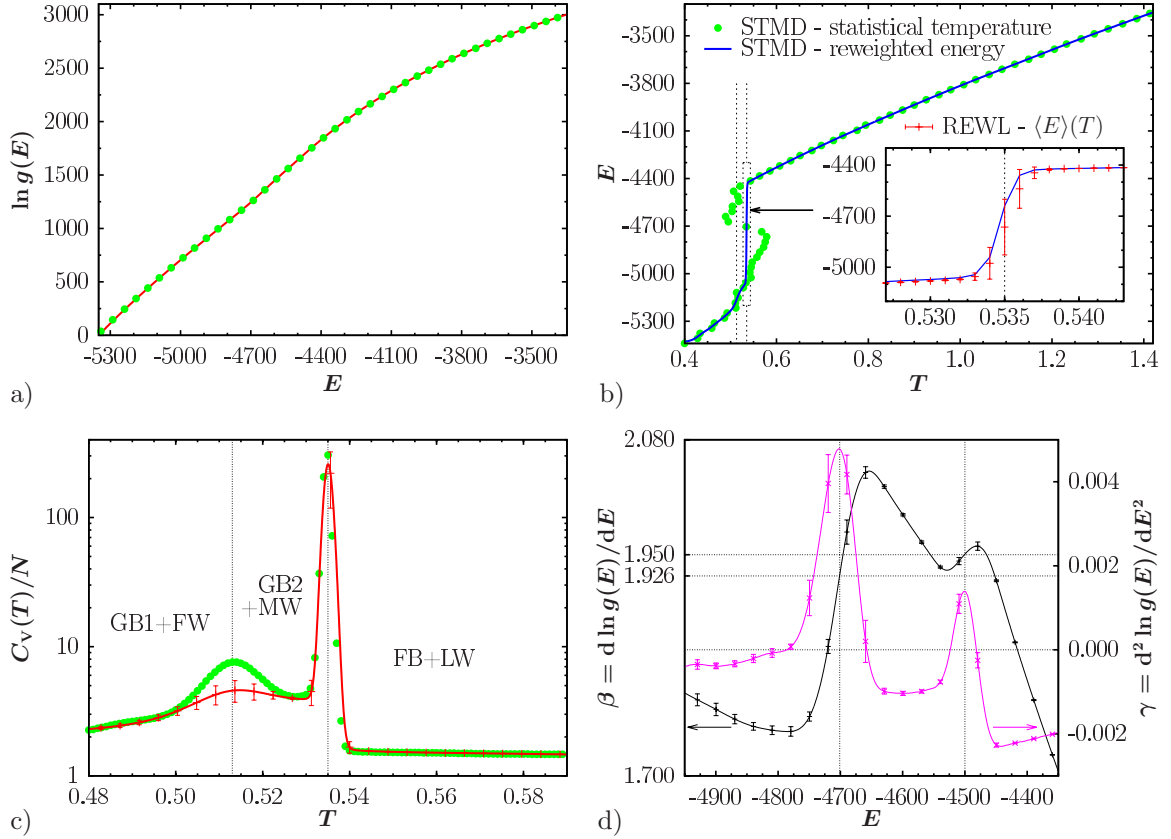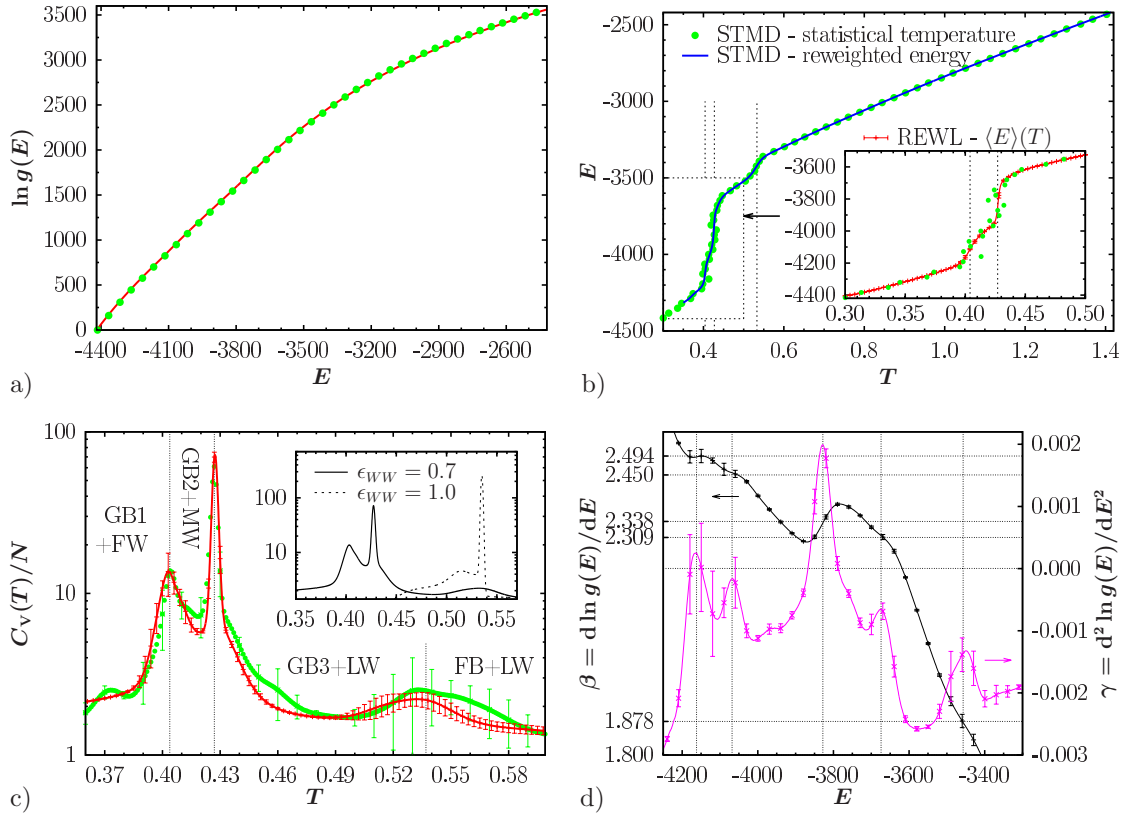

Figure 5.6 Density of states and thermodynamic properties of lipid system with $M = 125$ and $\varepsilon_w = 0.7$; comparison of data from STMD (symbols) and REWL (solid line) (a) logarithm of density of states, (b) Energy-Temperature relations, comparison between the convergent energy (dots) and reweighted energy (straight line). The inset shows the comparison between the statistical temperature from STMD and the mean energy calculated from REWL results. (c) Heat capacities calculated from STMD and REWL results. Inset: Comparison with data shown in **Figure 5.5** for $\varepsilon_w = 1.0$. (d) microcanonical analysis. Values for β on left scale agree with peak temperatures in (c) via the relation $\beta = 1/T$. Generally, error bars are only shown every nth ($n > 20$) data point.
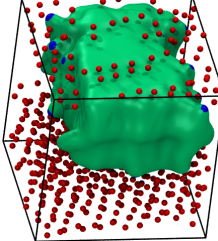
Simulation snapshots are provided in **Table 5.3** that demonstrate the different bilayer phases observed in the simulations. At low temperature, the system is again in the GB1+FW phase (see **Table 5.1**) and as temperature increases the frozen water melts in

the above-mentioned two steps via a pre-transition (exhibiting GB2+MW structure) and then the main transition (to GB3+LW phase). At a higher temperature still, the gel bilayer phase begins to change to the fluid bilayer phase (FB+LW phase).

Table 5.3 Simulation snapshots showing four different bilayer phases seen for the $M = 125$ lipid system with $\varepsilon_w = 0.7\varepsilon$.

GB1+FW	GB2+MW	GB3 +LW	FB+LW
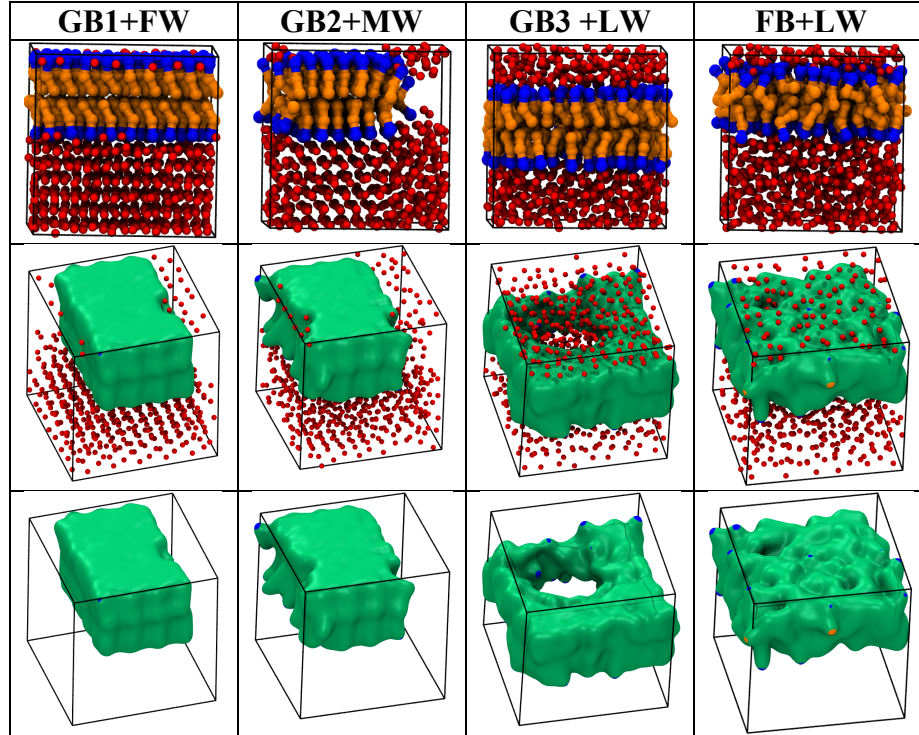
From the comparison of the heat capacities obtained from the two sets of simulations performed for different water-water interactions (i.e., with $\varepsilon_w = 0.7$ and 1.0. see **Figure 5.6**c for a direct comparison), we can see that only two peaks are observed when $\varepsilon_w = 1.0$, since the gel-fluid phase transition of the lipids and the frozen-liquid water transitions occur almost concurrently. This also explains the large peak in C_v at $T = 0.535$ for $\varepsilon_w = 1.0$, which is generated from the energy changes in both the LJ potential energy between water particles and the lipid fluidization process. Thus, the utilization of

a smaller ε_w for the water-water interaction separates the processes of the lipid gel-fluid transition and the solvent frozen-fluid transition, providing a clearer demonstration of the complicated phase behavior of this solvated lipid system.

5.4.2 Order parameter

To provide additional interpretation of the various phases and demonstrate that the peaks in C_v are indeed associated with the structural transitions discussed above, several structural metrics have also been calculated for the system of $M = 125$ with $\varepsilon_w = 0.7$. In **Figure 5.7**a, the average nematic order parameter S2 (see Section *3.2.2.1*) is plotted as a function of temperature. As can be seen from the figure, S2 maps closely to the transitions observed in C_v at $T = 0.404$ and 0.427. The most significant change (i.e., steepest) occurs at $T = 0.427$, resulting in a large increase in the orientational ordering of the lipids; this transition appears to finish by $T = 0.404$, where S2 attains a near constant value with a small standard deviation (i.e., it is in the GB1 phase). Additionally, S2 appears to first start increasing roughly around the transition marked by $T = 0.533$, marking the transition between the GB3 ($T < 0.533$) and the FB ($T > 0.533$) phases.

Additional insight into the transition at $T = 0.533$ can be gleaned by examining the value of S2 for the individual segments of the lipid chain. In **Figure 5.7**b, S2 calculated for "segment 1", containing the head group and first bead of the tail, and "segment 2," containing the two tail group beads is calculated. Overall, segment 1 has a more uniform alignment than segment 2 over the temperature range studied, however S2 is reduced in both cases as temperature increases, indicating more random configurations. The increased uniformity of segment 1 is likely related to the fact that it is located at the planar interface of the bilayer. Additionally, for segment 2, there is a clear transition in

S2 associated with $T = 0.533$, where S2 begins to more rapidly increase as system temperature is reduced, consistent with visual inspection. Segment 1 shows a less obvious transition; as the system is cooled, orientational ordering of this first segment is actually reduced until the temperature is decreased below $T = 0.533$, where it then begins to trend upwards. This loss in ordering is associated with the formation of holes within the bilayers; the formation of the water-filled holes induces curvature at the water-head group interface, reducing the overall global-order.

We observe that the normal to the bilayer does not change significantly during the simulation, which enables us to easily quantify the formation of the aforementioned holes. By projecting the lipids onto the plane parallel to the bilayer, a 2D histogram can be constructed; cells that do not contain lipid particles are considered to be holes. **Figure 5.7**c shows the fraction of the plane occupied by the bilayer, from which we can see that the key changes in fraction match closely to the transitions in C_v, where, in particular, there is a clear plateau of the average fraction in the region bounded by $0.427 < T < 0.533$. Even more telling is the standard deviation, where, from approximately $T = 0.6$ to $T = 0.5$, the error bars, corresponding to the standard deviation, grow significantly, reaching a peak at $T = 0.533$. For state points in this regime around $T > 0.533$, we see a mixture of systems with holes and those that span the box, while for $T < 0.533$ all systems appear to exhibit holes in the bilayer structure. The formation of holes results in a more compact bilayer sheet, as well as decreased hydration as T reduces to the point where water changes from liquid to frozen (i.e., the bilayer hole is not filled by water). These phenomenon result in decreased potential interactions between the head groups and water, which manifests itself in the heat capacity curve.
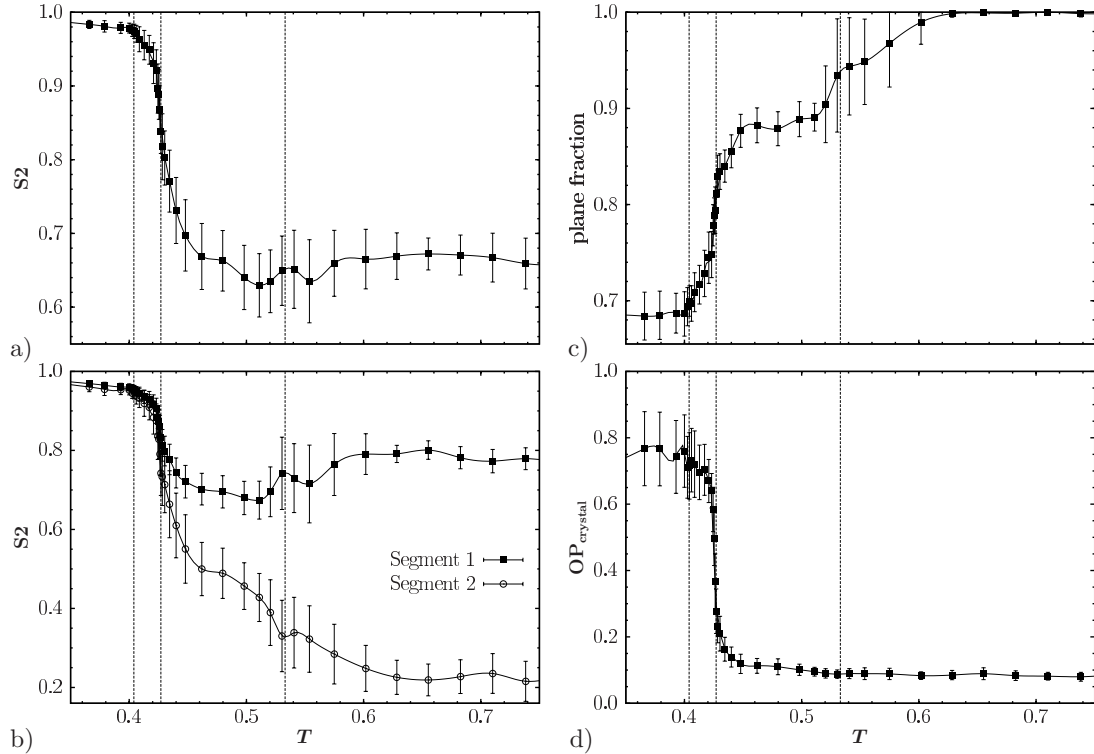
Figure 5.7 Structural metrics, a) global nematic order parameter S2; b) S2 for the individual segments of the lipid chain; c) the fraction of the plane occupied by the bilayer; d) order parameter for crystallization of water particles. Note the structural transitions determined via C_v are plotted as dashed vertical lines for visual aid (T=0.404, 0.427, 0.533)

Table 5.4 Feature summary of the lipid system at M=125 with ε_w = 0.7

T	0.35 ~ 0.404	0.404 ~ 0.427	0.427 ~ 0.533	>> 0.533
Water	Frozen water	Mixture	Liquid water	
Lipids	GB1	GB2	GB3	Fluid bilayer
Holes	Yes, without water		Yes, filled with water	No
Structural feature	Tail highly order	Less order; More hydrated	Less order; Curvature at boundary	Tail random distributed

Finally, the order parameter associated with the crystallization of water particles has been calculated showing a strong increase in the average order parameter starting at T = 0.427 and ending at T = 0.404 (**Figure 5.7**d). This supports the results of the visual

inspection that observed frozen water for $T < 0.404$, a mixture of frozen and liquid for $0.404 < T < 0.427$, and predominantly liquid water for $T > 0.427$.

To give an overall view of the bilayer structures discussed above, the features of the lipid system are summarized in **Table 5.4**.

5.5 Phase transition behavior at different lipid concentrations

A system with a lower lipid concentration of $M = 75$ ($\varepsilon_w = 0.7$) has also been studied to examine the effect of lipid concentration on the phase transition behavior. The calculated heat capacity for this system is provided in **Figure 5.8** and shows that the phase transition behavior is similar to the system with $M = 125$, i.e., three apparent phase transitions are observed but at shifted temperatures. Simulation snapshots for this system are also presented in **Table 5.5**. At low temperatures when the water is frozen, the gel phase bilayer structure given in **Table 5.4** is found to be quite similar to that seen for the system with $M = 125$; however, upon increasing the temperature, when the system is in the fluid phase (bilayer with liquid water, FB+LW), due to the smaller amphiphilic concentration, some head groups are found to wrap around the border of the bilayer which introduces curvature at the interface and prevents the lipid tails from interacting with the water particles. However, the large lipid aggregate will merge with itself at higher concentrations ($M = 125$) in one or two directions and form a stable bilayer.
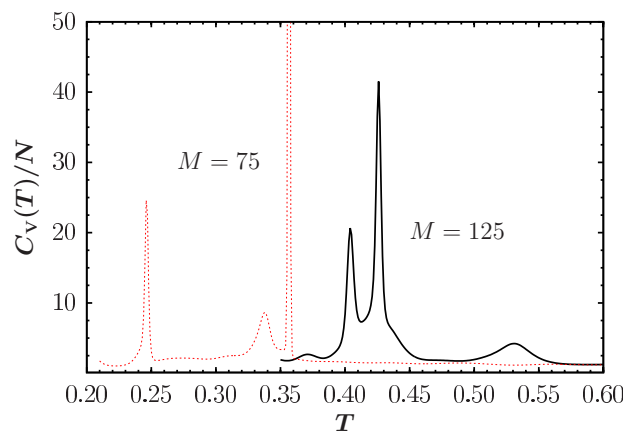
Figure 5.8 Heat capacity at $M = 125$ (black) and $M = 75$ (red dotted) lipid concentrations. $\varepsilon_{\mathrm{w}} = 0.7\varepsilon$ in both cases.

Table 5.5 Snapshots showing the four different bilayer phases seen simulations for $M = 75$.

GB1+FW	GB2+MW	GB3 +LW	FB+LW
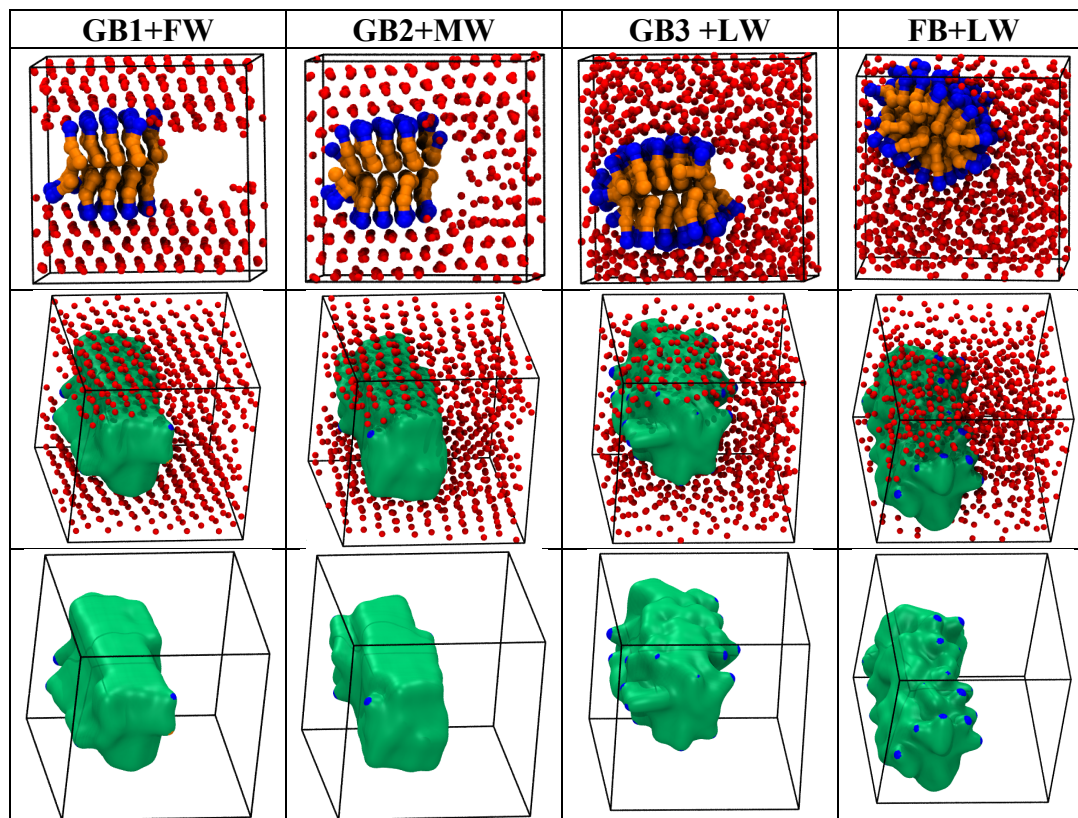
CHAPTER 6   PHASE TRANSITION BEHAVIOR OF NANO-CONFINED FLUIDS

In this chapter, the phase transition behavior of LJ fluids confined by attractive walls was examined, at various pore sizes, wall-fluid interaction strengths and registry of the wall surfaces. The application of STMD enables a detailed and explicit calculation of the underlying free energies and heat capacities, which more fully reveals the nature of phase transition behavior.[86,146,147] Here, we find convincing evidence that the phase transition strongly dependent on the pore sizes and wall-fluid interaction strengths. This work corresponds to a manuscript in preparation by Gai *et al.*[148]

6.1   Introduction

Fluids confined by solid surfaces with separation distances on the order of a few molecular diameters are of great importance both in industrial applications such as lubrication and separation, as well as emerging areas, such as ionic liquid-based supercapacitors.[8, 68, 69] When molecules are confined within such narrow pores, their phase behavior can be dramatically different from that of the bulk fluid. This surface-driven phase change can be attributed to a combined effect of fluid–wall interaction[68, 71, 72] and spatial constraint in the pores.[73] This phenomena has been widely examined experimentally,[74-80] with strong evidence of confinement induced solidification, e.g., the observation of a several orders of magnitude increase in viscosity as a function of reduced pore separation,[75, 78-80] although other studies have posited that systems undergo a constant vitrification as separation is reduced;[76, 81, 82] in either case, it is well agreed upon that the structure is significantly altered as compared to the bulk, and may undergo

disorder-to-order transition, as systems undergo nanoscale confinement. Molecular simulations have also been performed,[68, 71, 72, 86-88] employing both coarse-grained and atomistically detailed models of the fluids and surfaces; these studies have led to an important understanding of the behavior under confinement, specifically the importance of the wall-fluid interactions as dictated by the partial charges on the atoms. However, to examine the nature of the liquid-solid phase transition, a detailed and explicit calculation of the phase transition behavior based on the underlying free energies is necessary.[8]

In early work, Dominguez *et. al.*[146] examined the free energy of the solid and fluid phases by thermodynamic integration in purely repulsive and weakly attractive pores, and located the phase transition point accordingly. However, the method is not extensible to other cases of moderately or strongly attractive wall–fluid interactions, since it is (1) difficult to find a suitable path of integration that is thermodynamically reversible at these conditions, and (2) the calculations did not take into account the heterogeneous nature of the particles in the wall vs. the fluid, as occurs in strongly interacting systems. In other work, Gubbins and coworkers[86] performed grand-canonical Monte Carlo simulations with umbrella sampling on the Lennard-Jones fluid confined between structure-less walls. They calculated the relative free energy differences between the liquid and solid phases under fixed confinement for systems with both weak and strong attractive wall-fluid interactions. From the comparison of the relative free energy, two transitions were observed corresponding to the pre-crystallization (the layers near the two walls freeze first while the rest of the system remains fluid-like) and another transition associated with full solid formation (all the layers are frozen). However, this approach can only provide relative free energy difference between the confined solid and liquid phases, which

makes it impossible to directly compare the free energies between different pore sizes. Furthermore, this relies on the definition of a suitable order parameter for umbrella sampling;[149] this may be non-trivial for complex systems and the wrong choice of order parameter may miss important behavior in the system. Recently, Wan *et al.*[147] examined the phase behavior of LJ fluid system confined by attractive walls modeled by explicit LJ particles using a combination of grand-canonical molecular dynamics (GCMD) simulations and absolute Helmholtz free energy calculations by the Einstein crystal method. This work specifically focused on the behavior of systems over a range of separations, including both ideal and non-ideal separations. It was shown that the general trend was a reduction in the free energy as the pore size was decreased to sizes that can accommodate an integer number of molecular layers; however, when also considering non-integer spacing, considerable oscillatory behavior in free energy is observed as the pore separation was decreased. However, the Einstein crystal method[150] is only applicable to stable solid states, and cannot be extended for determining the free energy of non-crystalline confined states, and thus the phase transition cannot be directly probed.

As an extension of the above work of Wan *et al.*,[147] we examined a similar LJ system using the STMD method,[31] which allows the calculation of density of states, and thus relative free energy over a wide range of temperatures interested, including the phase transition, as described in Section *3.1.4*. We examine the phase transition behavior *via* the calculation of the heat capacity, to determine the effect of pore size, pore registry and the nature of the fluid-wall interaction on the phase transition. By combining these results with aforementioned Einstein crystal method, the offset between the free energy determined by STMD can be determined, allow for comparison of the absolute free

energy between different states, e.g., different pore separations. Here, we see strong evidence of a significant dependence of the phase transitions on the pore sizes and wall-fluid interaction strengths.

6.2   Simulation details

The nano-confined system employed in this work is similar to that used in Wan *et. al.*,[147] where the system is composed of two particle species, mobile fluid particles and static wall particles, with a bulk phase in contact with two identically sized pores, as shown in **Figure 6.1**. Two pores are used as this provides a means to adjust pore size for non-ideal pore sizes while keeping the box dimensions fixed. All particles are modeled as LJ spherical particles with identical size (i.e., $\sigma = 1$), with interactions between mobile fluid particles (i.e., the characteristic LJ energy for mobile molecule-mobile molecule interaction) set as $\varepsilon_{mm} = 1.0$ for all simulations. Interactions between the walls (w) and mobile particles (m) are varied with $\varepsilon_{wm} = (1.0, 2.0, 4.0)$. XPLOR style shifting is used such that the well depth of the potentials are unchanged regardless of cutoff; due to the heterogeneous nature of the system we employ a cutoff of 3.0σ between mobile particles, where XPLOR shifting starts at 2.75σ, and a longer cutoff of 5.0σ between wall-mobile interactions, with XPLOR shifting starting at 4.75σ. System sizes are fixed at 7680 particles, with roughly 25% of the particles existing as part of the pore walls; the number of mobile particles varies slightly as pore size is varied. Wall particles are set in an FCC structure with number density $\rho_{wall} = 1.0$, corresponding to the minimal energy state for an LJ system. In all cases, the box size is set as (11.225, 33.852, 21.996), which for an ideal crystalline pore state with density of 1.0 results in a bulk fluid with number density of 0.85. In this case, the size of the bulk fluid is sufficient that changes in densities in the

pore do not strongly influence the bulk state. While anisotropic *NPT* simulations are typically employed in GCMD simulations, where the long dimension of the box can fluctuate, this approach cannot be used with the standard STMD method. For ideal pore separations, the distance is set to integer multiples of the FCC layer spacing $2^{1/6}\sqrt{6/9}\sigma$ (i.e., 1 Layer=$2^{1/6}\sqrt{6/9}\sigma$).
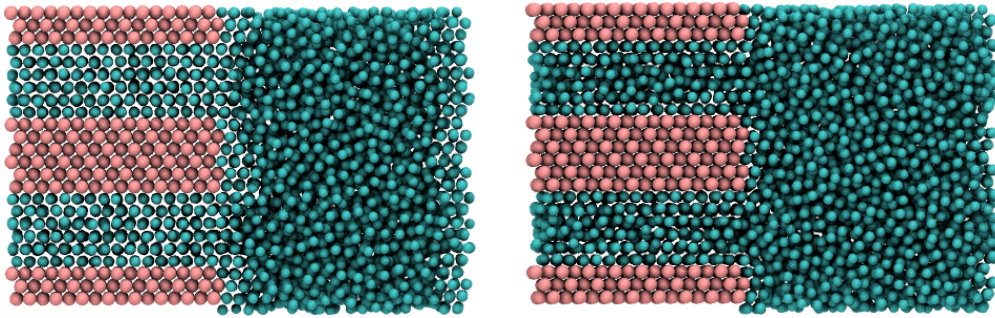
Figure 6.1 Snapshot for illustration of the confined system

6.3 Ideal systems with varied fluid-wall interaction strength

With the STMD method, a temperature range covering the phase transition temperature within the nanopore is given specifically for each system in **Table 6.1**. For the STMD runs, an energy bin size of $\Delta E = 32$ was used with an initial modification factor $\ln f = 0.00025$.[31] To accelerate the simulations, the entire temperature range was split into two separate temperature windows which were merged in postprocessing,[4, 111] e.g., for 4 layers confined with ε_{wm}=1.0, two windows in $T \in [0.65,1.0]$ and $T \in [0.85,2.0]$ are used. For $\varepsilon_{wm} = 1.0$, the total simulation times for STMD were 8.3×10^7 and 5×10^8 MD steps, respectively for each window, to converge with a final modification factor of $\ln f < 5 \times 10^{-7}$. **Figure 6.2** plots the energy as a function of simulation progression, demonstrating a random walk in energy space.

Table 6.1 Temperature range simulated for each system

Confined layers	Separation (σ)	T range		
		$\varepsilon_{wm}=1.0$	$\varepsilon_{wm}=2.0$	$\varepsilon_{wm}=4.0$
4	3.67	0.65~1.9	0.65~2.1	0.65~2.7
5	4.58	0.65~1.7	0.65~1.8	0.65~2.2
6	5.50	0.65~1.6	0.65~1.6	0.65~1.75
7	6.42	0.65~1.5	0.65~1.5	0.65~1.5

a)   b)
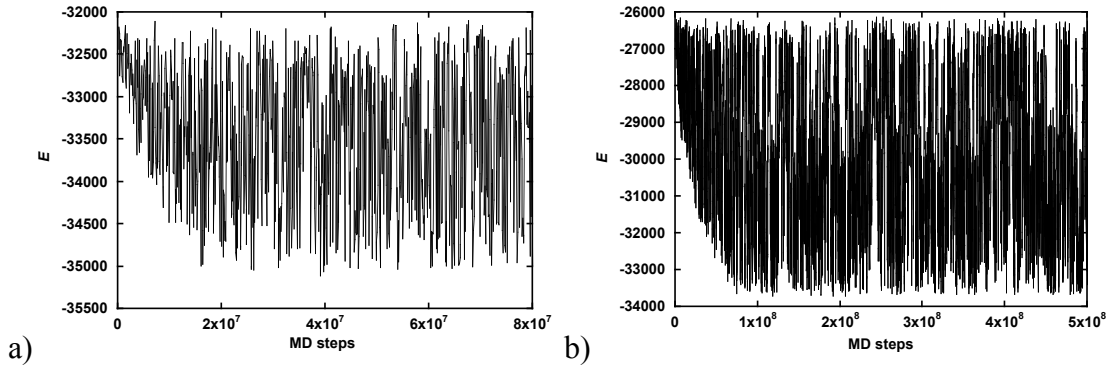
Figure 6.2 Energy trajectory from STMD simulation of pore size $=4L$ system with ε_{wm} $=1.0$, for a) $T\in[0.65,1.0]$ and b) $T\in[0.85,2.0]$.

6.3.1 Heat capacity

With the convergent $T(E)$ (not shown here), $g(E)$ can be integrated and heat capacity calculated to give a robust estimate of the location of the phase transitions. First, for $\varepsilon_{wm}=4.0$, the T_{ODT} is compared with varied pore separations, as shown in **Figure 6.3**. As the size of the confined region increased, we observed a decrease in fluid-solid phase transition T_{ODT}, as expected from prior work,[147] with a sharper and higher peak in C_v. The sharpness of the peak is likely related to several factors. First, the interaction between the central layers of the pore and wall is reduced as the distance between the pore walls is increased. Second, as the pore size increases, the number of particles in the confined region increases, which partially contributes to a larger change in energy upon

crystallization and thus sharper changes in the heat capacity, resulting in more bulk like behavior.
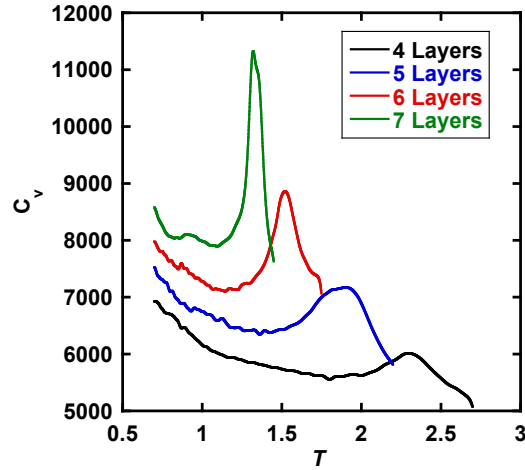
Figure 6.3 Heat capacity of nano-confined LJ system at varied pore distances

In addition, for a fixed confining distance, we compared the heat capacity at different interaction ε_{wm}, shown in **Figure 6.4**. For all the systems, as ε_{wm} is reduced, a lower T_{ODT} is observed, in agreement with the above discussion that the reduced interaction between mobile particles and walls requires a lower temperature to crystallize, and the behavior becomes more bulk like.[86, 151]

Table 6.2 gives a summary of the T_{ODT} estimated from the peak(s) in the C_v curves; this peak corresponds to the midpoint of the transition and thus will overestimate the transition temperature. However, this estimate provides unambiguous and uniform criteria to establish trends. Similar trends in T_{ODT} are observed as compared to Wan *et al.*,[147] where large pores demonstrate more bulk like behavior (i.e., T_{ODT} in the pore approaches that of the bulk). However, as expected, T_{ODT} estimated from STMD tends to be slightly higher than in Wan *et al.*,[147] which relied on order parameter analysis. Additionally, as will discussed below, the C_v curve begins to demonstrate two peaks as

the wall fluid interaction strength is reduced and pore size increased; for consistency, we report both peaks in **Table 6.2**.
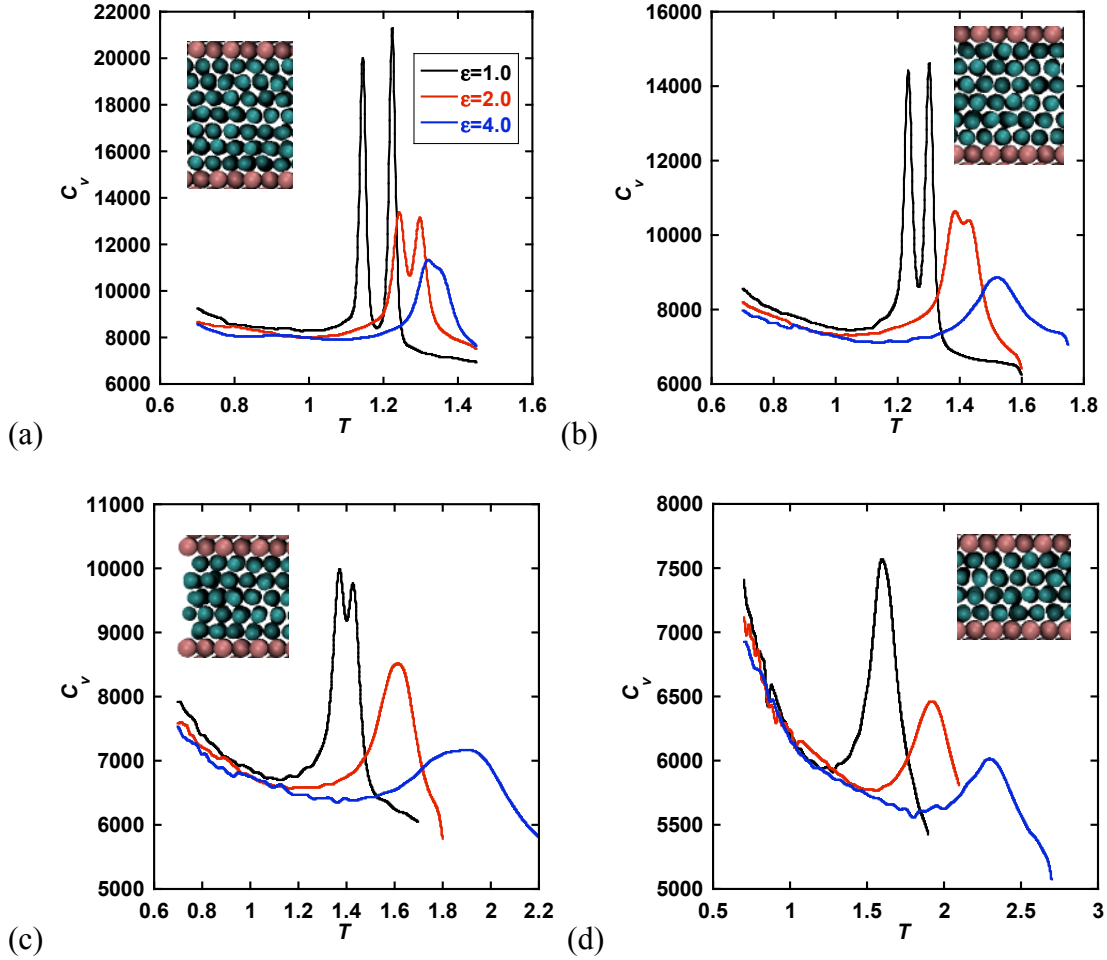
Figure 6.4 Comparison of heat capacity at varied interaction with fixed confined distances, (a) 7 layers, (b) 6 layers, (c) 5 layers, (d) 4 layers

Table 6.2 Summary of T_{ODT} from STMD in this work with resolution of T=0.001, compared to Wan *et al.*[147]

No. of Layers	ε_{wm} =1.0		ε_{wm} =2.0		ε_{wm} =4.0	
	T_{ODT}^{STMD}	T_{ODT}^{147}	T_{ODT}^{STMD}	T_{ODT}^{147}	T_{ODT}^{STMD}	T_{ODT}^{147}
4	1.598	1.4	1.920	1.8	2.297	2.3
5	1.371	1.1	1.614	1.4	1.904	1.7
6	1.234/1.302	0.9	1.385/1.429	1.1	1.521	1.2
7	1.145/1.224	0.8	1.242/1.298	0.9	1.320	1.0

6.3.2 Order parameter

The behavior of the nano-confined system is different from that of the bulk, and the extent of the difference is dependent on the strength of the wall-fluid interaction ε_{wm}, the separation of the confining walls, and the location of a given particle from the walls.[72] The confining walls introduce heterogeneity in the system; even in the absence of attraction, the dynamics of particles along the walls have been shown to be altered compared to bulk.[73] For systems with attractive walls, as considered here, the potential landscape within the pore varies at each parallel separation from the wall, where those particles closest to the wall demonstrate the strongest, least bulk-like behavior. It is also important to note that, in this work, while the contribution from a single wall is fixed regardless of separation, ultimately the potential felt by the particles will depend on the combined effects of both walls; particles in smaller pores experience stronger potential interactions than larger pores at the center layers, as shown in **Figure 6.5**.
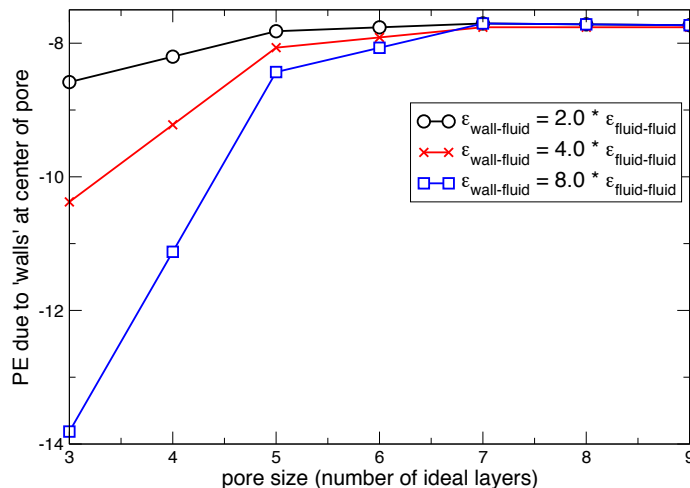
Figure 6.5 Potential energy at the center layer(s) of the pore resulting from the effective walls on each side

To further examine this heterogeneity, the 2D global hexagonal OP[121, 122] is examined as a function of T for different spatial regions in the pore, and simultaneously compared to the heat capacity. **Figure 6.6** plots the OP for each specified layer (shown in **Figure 6.6**a) for a pore that can hold six ideal layers, as a function of T for various ε_{wm} (shown in **Figure 6.6**b-d).

For all interaction strengths, particles in Layer 1 - i.e., those particles in contact with the walls - demonstrate a high value of the OP over the entire T range considered; the changes in the OP are relatively minimal as T is increased beyond the T_{ODT} predicted by the peak(s) in C_v curves, suggesting a highly ordered state even at high T, as has been seen in other works.[86, 147] Layers 2 and 3 demonstrate a more significant drop in OP value as T is increased, coincident with the peak(s) in C_v curves. However, for Layer 3 the magnitude of the change in the OP as the ODT is crossed does not strongly depend on ε_{wm} (in all three cases, OP of Layer 3 is reduced to small values ~0.2), while the change in the OP for Layer 2 becomes smaller in magnitude as ε_{wm} is increased. Specially, in **Figure 6.6**b, the OP of Layer 2 drops to ~0.2, and **Figure 6.6**d shows OP of Layer 2 is still large (~0.5) at the right end of the C_v peak. That is, Layer 2 becomes increasingly ordered as ε_{wm} is increased. As such, the ordering of the central most layers, in this case Layer 3, is ultimately indicative of the formation of a solid phase in the pore for this pore size.

Additionally, the difference in structural ordering with T between the central-most layers (i.e., Layers 2 and 3 for 6 total layers here) appears to manifest itself in the number of peaks seen in the C_v curves. Specifically, for the system with 6 total layers, systems

with weak ε_{wm} demonstrate two distinct peaks (**Figure 6.6**b), steadily reducing in magnitude until only a single peak is seen for the strongest ε_{wm} (**Figure 6.6**d). The increased ordering in Layer 2 at this larger ε_{wm} results in a less significant change in structure as the system is cooled, which manifests itself in a more gradual, less sharply peaked C_v behavior.

As was seen in **Figure 6.4**, the appearance of one or two peaks is also dependent on the pore separation, where larger pores tend to demonstrate two peaks in heat capacity, in correspondence with different ordering behavior of the central most layers. For example, in **Figure 6.4**a, C_v is plotted for a system that can contain 7 ideal layers; even ε_{wm} = 2.0 appears to have two peaks, albeit less well defined than weaker interaction strengths. Compared to a pore size of 5 ideal layers, only two peaks for ε_{wm} = 1.0 are observed, albeit rather weakly differentiated, and only a single, wide peak is seen for ε_{wm} =2.0. Further considering an even smaller pore, e.g. 4 ideal layers, only a single peak in C_v is seen for all ε_{wm}, due to the fact that there are no particles that would exist in Layer 3, i.e., no particles separated from the walls by 2 layers of particles and thus the system is unable to introduce different ordering behaviors as a function of T.

Furthermore, the ordering of the individual layers in the pore, or more specifically, the degree of ordering above the ODT, also appears to be directly related to the height of the C_v peak(s). Systems with weaker ε_{wm} have a more significant structure change, as was seen in **Figure 6.6**b, where both Layers 2 and 3 change from mostly disordered to ordered; in contrast, in **Figure 6.6**d, systems with larger ε_{wm} demonstrate wider peaks of lower magnitude, even though the system size is equivalent, since the structural transition primarily occurs on a smaller subset, namely Layer 3. Thus, it is clear

that the wall-fluid interaction strength will also influence the sharpness and height of C_v in addition to the number of particles in the pore, as stated previously.
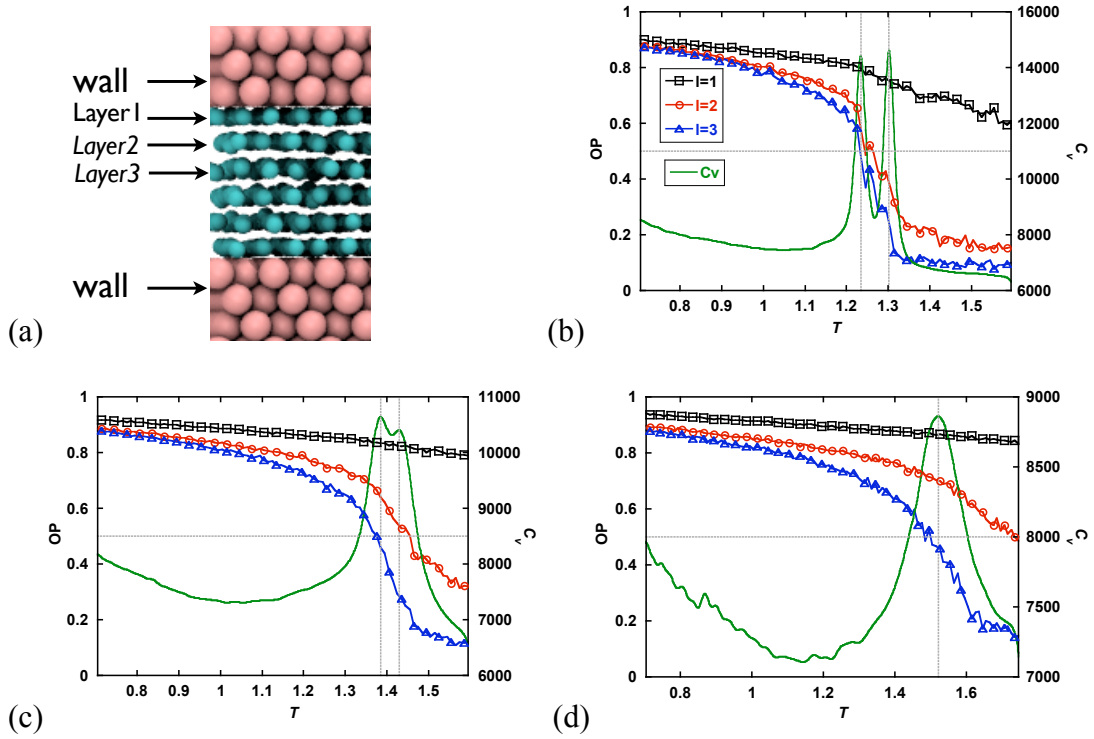
Figure 6.6 Order parameter for a system that can hold six ideal layers, (a) snapshot for the layers (b) ε_{wm} =1.0, (c) ε_{wm} =2.0, (d) ε_{wm} =4.0. Note, OP values are calculated with a resolution of T=0.01 to capture the curve of change, but for better view, symbols are drawn with every three of the values

6.3.3 Free energy calculation

To support the results obtained above, we further calculated the underlying free energy for the nano-confined systems. With the obtained density of states, the partition function can be calculated using[117] Equation 3.16 and then the relative free energy is calculated as[119] Equation 3.17.

To compare the free energy trends with the change of confined distance, we calculated the absolute energy at T=1.0 with the Einstein crystal method (N in the pore region) using three different configurations from STMD for a system with ε_{wm} =4.0; the

states chosen were those in the crystalline solid regime where the Einstein crystal method is valid. These free energies can then be used to offset, or align the relative free energy curves provided by STMD, in the same matter as was done in Iacovella *et al.*[152] using absolute free energy calculations to align free energy perturbations. The advantage of additionally using STMD is that the free energy can be calculated even for a non-crystalline confined state, where the Einstein crystal method would not work, allowing direct investigation of the phase transition. Four different temperatures and their aligned free energies are plotted for different ideal pore sizes, as shown in **Figure 6.7**a. The free energy increases as the pore size increases for each of the temperatures explored. We also compare the free energy changes by normalizing by the values at the estimated T_{ODT}, as shown in **Figure 6.7**b. These plots show that as the confined region becomes larger, the changes in free energy also become larger, further supporting the more significant change in the structural ordering previously discussed in **Figure 6.3** in relation to the sharper change in C_v.

(a)

(b)
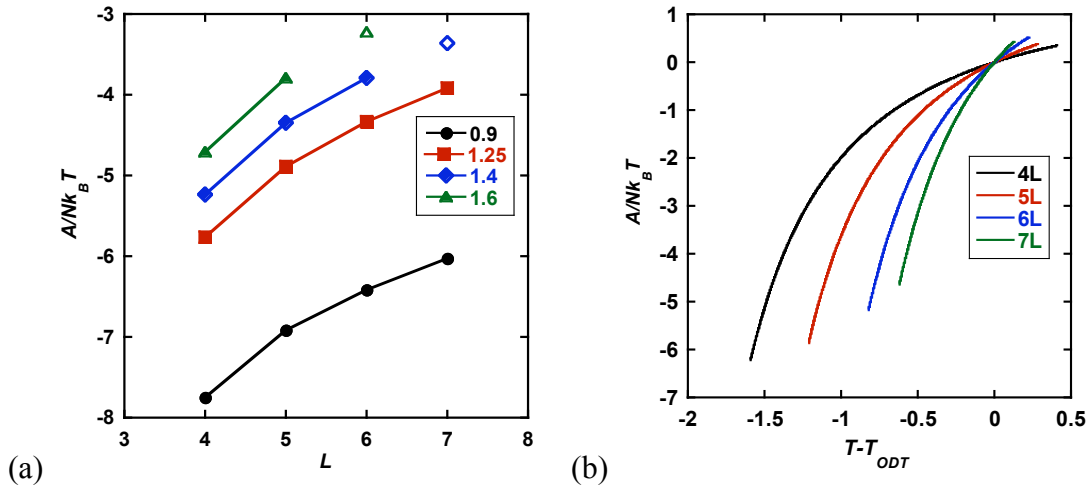
Figure 6.7 (a) Free energy at different layers, solid symbol means solid states, while open symbol means liquid states, (b) free energy changes with ΔT.

6.4 Non-ideal pore size

STMD simulations were also performed for systems with non-ideal pore spacing. The distances between walls studied were 4.5, 5.5 and 6.5 layers, respectively. With the same method as above, the phase transition behavior is demonstrated by heat capacity calculations, shown in **Figure 6.8**. We found that the behavior of these non-ideal spacing systems is quite different from that of the ideal pore spacing systems, with T_{ODT} shifted to much smaller values (T=0.8~1.1). To explain this, we examined the density of the particle in the pores and found considerably oscillatory behavior, i.e., the density in non-ideal pores is lower than that in the ideal pores (**Figure 6.9**a). Generally, compared to the ideal-pore separation systems, the particles tend to be more disordered in the non-ideal spacing pores (due to the extra space), with fewer particles entering the pore region, and thus require a lower temperature to get well ordered structures.
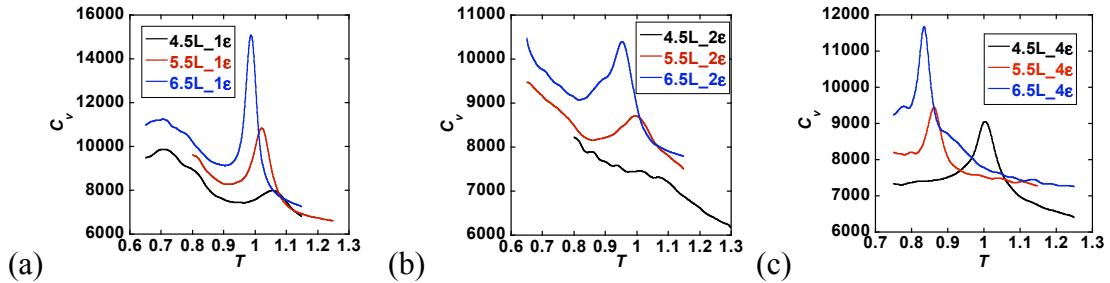
Figure 6.8 Heat capacity for non-ideal pore spacing systems

Additionally, for the non-ideal spacing system, as ε_{wm} becomes stronger, T_{ODT} is reduced, the opposite of what is seen for ideal spacings. The possible reason is that the stronger wall-fluid interaction tries to accommodate more particles into the non-ideal spacing pores, i.e., increased densification, as is seen in the density plot shown in **Figure 6.9**b, where a jump to higher density is seen for the system with strong interactions,

namely ε_{wm}=4.0. Examining the simulation snapshots for the structure in each pore in **Table 6.3**, we found that for the weaker interaction of ε_{wm} =1.0/2.0, in for example, the pore of 6.5 L, only 6 layers of solid particles are found in the pores, thus introducing a lower density in the pores. In contrast for a strong interaction, ε_{wm} =4.0, the upper was found to have 7 layers and bottom pore 6 layers, which also supports the notion of a more disrupted structure formed for strong ε_{wm} in non-ideal spacing pores, resulting in the lower T_{ODT}.

(a)   (b)
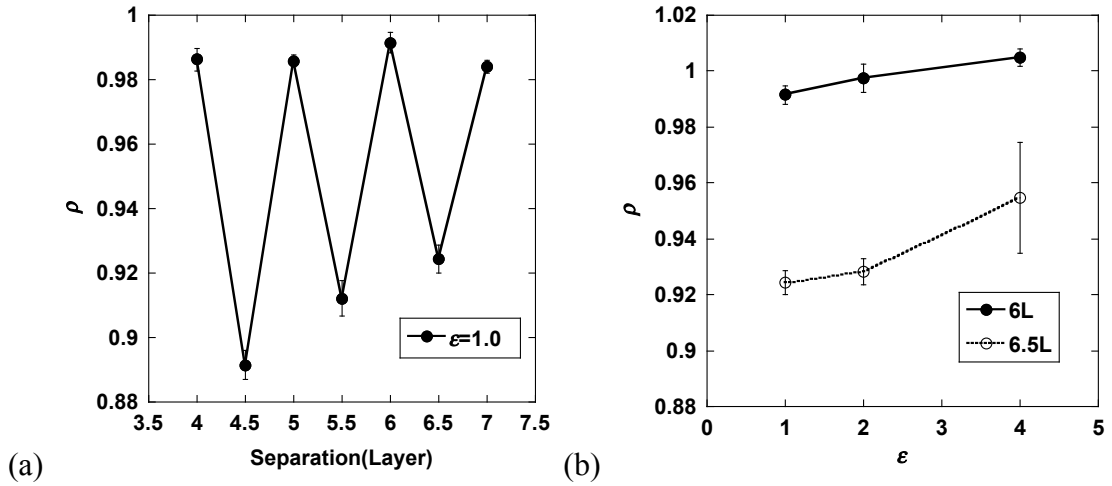
Figure 6.9 Density profile at T=0.8 (a) as a function of pore separation at ε_{wm} = 1.0, (b) as a function of ε_{wm}. Note, density here takes both of the two pores into consideration.

Table 6.3 Snapshot for the confined system with non-ideal seperation

4.5L		6.5L	
ε_{wm} =2.0	ε_{wm} =4.0	ε_{wm} =2.0	ε_{wm} =4.0

6.5 Registry of the surface

The confining walls of the above systems are all perfectly aligned, such that an ideal FCC structure can be accommodated without any stacking faults. However, such alignment cannot be guaranteed in experiment. Thus, it is necessary to examine the effect of the registry of the crystalline surfaces on the layered-structure in the pore. We shift the middle wall separating the two pores with a distance of $2^{1/6}\sqrt{3}/2$ σ, which induced a HCP stacking fault, as shown in **Figure 6.10**a. Similar STMD simulations are performed on these systems. We still observe a transition of fluid to solid as T is reduced, indicating non-ideality in orientation of walls does not impact the formation of a solid. Examining the C_v curve in **Figure 6.10**b more carefully, a slight change in the T_{ODT} compared to the perfectly aligned systems is observed, where in all cases, a loss of registry shifts the transition to lower temperature. In addition, for the fixed confining distance, we also compared the heat capacity at different ε_{wm}, and similar phase behavior is obtained as the above FCC stacking wall, i.e., as ε_{wm} is reduced, a lower T_{ODT} is observed.
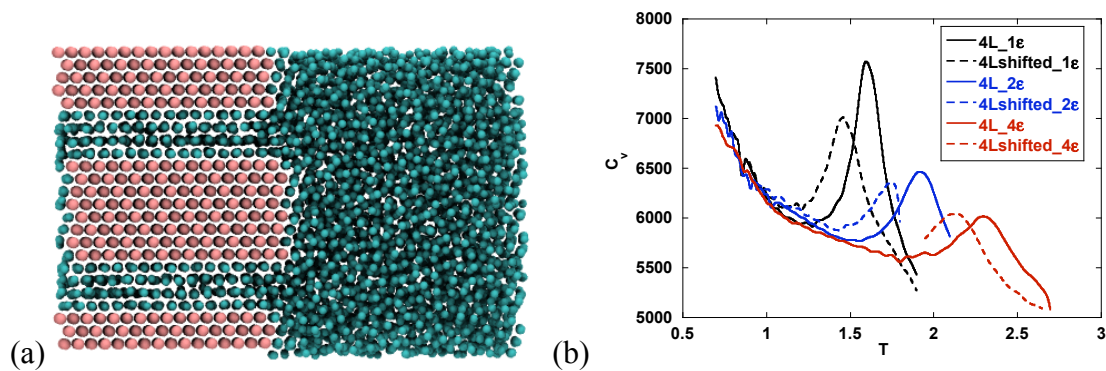
Figure 6.10 (a) Snapshot for the system with unregistered wall sheet, (b) heat capacity

Thus, the study here provides strong evidences that for the ideal pore separation, in general, systems with larger pores and weaker wall-fluid interaction strengths result in a sharper and more distinct fluid-solid transition at a lower temperature.

CHAPTER 7   CALCULATION OF TWO-DIMENSIONAL DENSITY OF STATES

In this chapter, we propose a novel algorithm, the hybrid WL/STMD method to take advantage of the benefits of STMD algorithm, but in a 2D Wang-Landau MC context. The hybrid method enables the calculation of 2D joint density of states, thus extending the original STMD method to various ensembles besides canonical ensemble (constant *N, V, T*), which will be helpful for revealing the entire phase behavior. The idea and feasibility of calculating the continuous 2D DOS by the hybrid method was demonstrated using a binary LJ system here. The hybrid method proposed here enables a more direct way for offsetting the free energy among different simulations/states, compared to the Einstein crystal method employed in the Chapter 6. This work is contained in a manuscript in preparation by Gai *et al.*[153]

7.1   Introduction

The temperature independent density of states (DOS) $g(E)$ is a central quantity of interest in thermodynamics. Many flat-histogram Monte Carlo (MC) methods have been developed for the direct calculation of $g(E)$.[4, 12, 25, 30, 154-157] As stated in Section 3.1.1, the WL method[4] is outstanding and flexible because it can dynamically update the density of states on the fly (at every simulation step instead of between runs), which can provide quicker exploration of the phase space. And the WL method in the original form has been effectively applied to study a number of different systems, focusing primarily on discrete systems such as simple models for polymers, lipids and proteins.[89, 94, 134, 135]

Recently, the WL method has been extended to apply in off-lattice systems with continuous degrees of freedom.[99, 106, 107, 158-161] In these systems, not only does the energy range become significantly larger (random movements can result in unpredictable energy changes in continuous systems),[12, 162] but also it is frequently important to calculate the joint DOS (JDOS)[105] depending on one or more order parameters besides energy. The JDOS of continuous models is useful since the free energy can be calculated not only as a function of temperature, but also as a function of concentration or pressure, etc., which can reveal the entire phase behavior. Nontrivial extensions of the original WL method are required for continuous systems. The de Pablo group[106, 158] has successfully described properties of continuous LJ fluids with extended WL algorithm. They also studied open systems in which random walks were performed for both the energy E and the number of particles N, thus generating an estimate of 2D g(N, E).[106] Shortly thereafter, Shell et al.[107] generalized the continuum WL method to various ensembles, presented the appropriate acceptance criteria, and implemented simulations on LJ fluids determining 2D density of states such as g(E,N) and g(E,V). However, when applying the WL algorithm to more complex systems, convergence problems can be encountered due to the extensive nature of the energy and discrete representation of DOS on an energy grid, which results in the number of energy bins required increasing substantially with system size.[30, 31, 163] When dealing with JDOS, such accumulation problems may occur more easily and become more severe. Thus, these simulations are limited to relatively small systems, or a selected narrow region in the parameter space.[105] To mitigate the accumulation problem, Zhou et al.[105] implemented a kernel function update in the WL method, combined with a global update (referred to "frontier sampling") to partially scale the estimate of the DOS to

higher values in order to visit the unexplored states. However, this method is very sensitive to the system–dependent parameterization and several studies have been reported[108, 164] that do not show significant improvements using the approaches recommended by Zhou *et al.*[105] for continuous systems.

Overall, these aforementioned WL samplings are based on Monte Carlo simulation, which is not easily applicable to more complex systems often due to the inability to create effective MC moves. To overcome these problems, a different approach, termed statistical-temperature molecular dynamics (STMD), was recently proposed by Kim *et al.,*[30, 31] as stated in Chapter 3. This is achieved by relying on the thermodynamic relationship between the statistical temperature $T(E)$ and the density of states $g(E)$.[165] The highly parallelizable nature of MD[166, 167] and independence of bin sizes makes STMD better suited for large, dense, complex systems. STMD sampling has been applied successfully to LJ fluids, liquid crystalline and biomolecules.[31, 135, 168] Despite its advantages, STMD in its current form can only use to calculate DOS as a function of temperature in the canonical ensemble.

Thus, in this work, we proposed the hybrid WL/STMD method (referred to as the 2D STMD method) to take advantage of the benefits of STMD algorithm, but in a 2D WL context. The general idea of the hybrid method is similar to replica exchange STMD,[169] but we perform WL samplings for the exchange of configurations at intervals. To demonstrate the idea of the method, we performed series of 2D STMD simulations on binary LJ systems, with the second variable, the composition of the particle (i.e., a semi-grand canonical ensemble) and to a lesser degree volume (i.e., *NPT* ensemble). In this work, we will put more focus on the former variable (i.e., composition of particle), the

composition change. From the results obtained, we determined a number of thermodynamic properties (average energy and heat capacity) to characterize the phase behavior. The results obtained were verified by comparison to the original WL method (referred to as 1D WL), WL in 2-dimensions following Shell *et. al.* (referred to as 2D WL), and standard STMD (referred to as 1D STMD).

7.2　Simulation details

The particular model studied is composed of a binary mixture (A and B) of LJ spherical particles at the same size. The total number of particles is set as 110 and the range of B particles is changed from 20 to 30. The number density of the system is 0.85 for systems with composition changes. The interaction parameter between particles of unlike species A and B is 0.5, and all others are set to 1.0. The temperature range studied is 0.7 to 1.8.

The STMD simulations are performed within HOOMD-Blue,[123, 124] and the WL sampling is implemented through an in-house code written in C++ with a python interface to enable easy integration with the HOOMD-Blue package. The cutoff of the interaction for all particles is set to half of the box length (2.53σ). Potentials are shifted to zero at cutoff (i.e., for each interaction potential a constant shift is applied to the entire potential such that it is 0 at the cutoff).

7.3　Hybrid method in semi-canonical ensemble with WL identity swap

The hybrid method was first applied to a binary Lennard-Jones fluid with a composition change by WL swap at an interval of 1000 STMD steps. As shown in **Figure 7.1**a, in the temperature range studied (0.7-1.8), systems with different compositions correspond to different energy ranges, but with some shared regions. For the WL part of the algorithm, the fixed energy bin needs to be among this region to enable swapping. In this work, we

used a bin size of $E=20$ or 100, and perform a random walk in the number of particles, as shown in **Figure 7.1**b. For the current studied system, the composition change is not very large, such that there is significant energy overlap. If extending to larger systems or those with larger composition changes, the calculation may need to be broken up into separate overlapping windows, similar to what is typically done in WL and STMD.

(a)   (b)   (c)
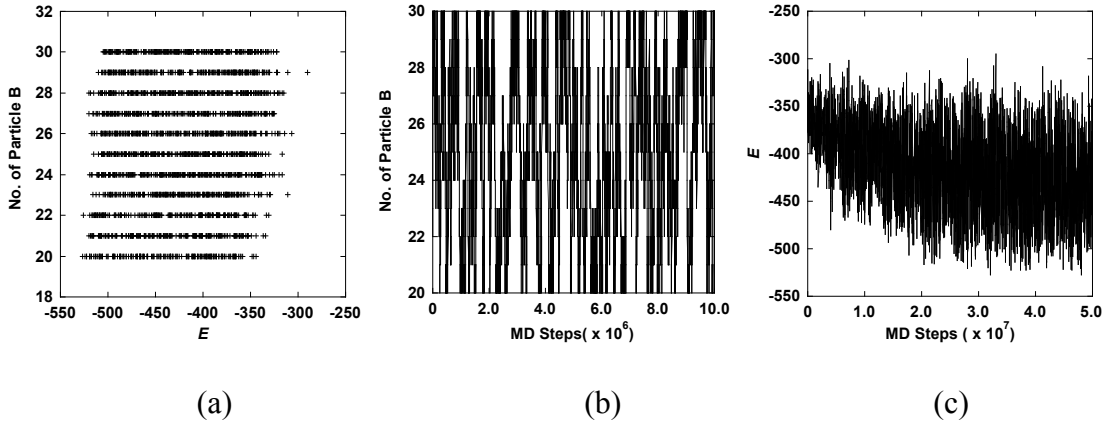
Figure 7.1 (a) Number of particles and the corresponding energy range in T [0.7:1.8], (b) the trajectory of the number of particles, and (c) energy trajectory from 2D STMD simulation

On the other hand, the evolution of the potential energy E at fixed composition in the STMD direction is also shown in **Figure 7.1**c, each of which displays a walk with frequent sweeps in the energy range, corresponding to the given temperature range. Once the simulations are converged, g(E, N_j) is obtained directly from the WL sampling (**Figure 7.2**a), and $g(E_i, N)$ (**Figure 7.2**b) is integrated from the convergent statistical temperature $T(E)$ (not shown here) in the STMD direction. A continuous distribution of DOS g(E, N) can be obtained by connecting the independent $g(E_i, N)$ with the offset from g(E, N_j). This is similar to the method described in Chapter 6, where the offset between different states was determined via absolute free energy calculations; however, absolute free energy calculations are only applicable to crystalline solid states and may present

challenges for complex molecules. For STMD, it took approximately 5×10^7 MD steps to reach the lowest temperatures for the first time in each window and was then followed by uniform sampling of the temperature range. The total simulation time was 2.5×10^8 MD steps, with a final modification factor of $\ln f < 10^{-8}$. At the same time, the WL part was also converged to $\ln f < 10^{-6}$, with 2.5×10^5 MC steps.

(a)   (b)
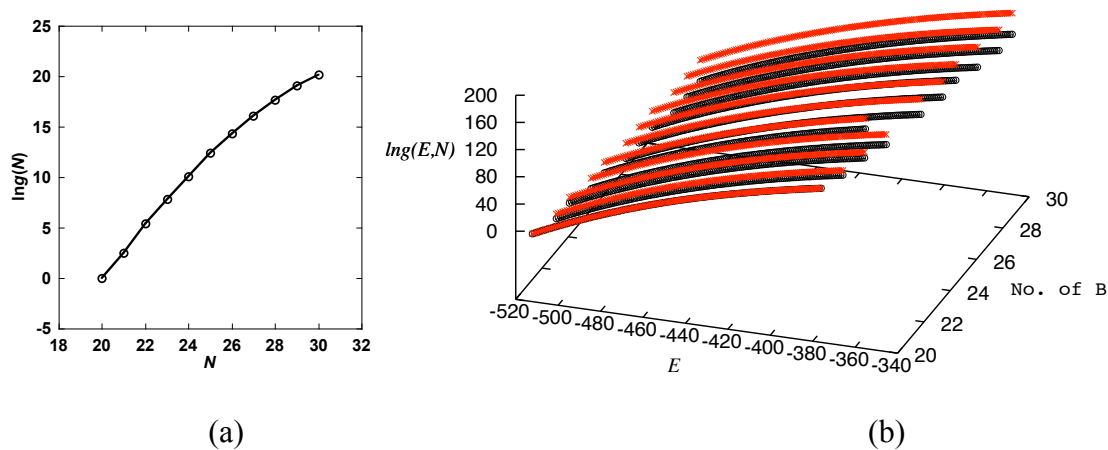
Figure 7.2 Density of states, (a) g(E, N_j) from WL part, (b) g(E_i, N) from STMD direction (black line), and final continuous g(E, N) (red line)

Thermodynamic properties such as heat capacity were then calculated from the obtained density of states, with a temperature step <0.01, providing a systematic analysis of the phase behavior as summarized in **Figure 7.3**. The peak in the heat capacity curve corresponds to the phase separation of the binary mixture, as can be demonstrated by the VMD snapshot in **Figure 7.3**b.

(a)   (b)
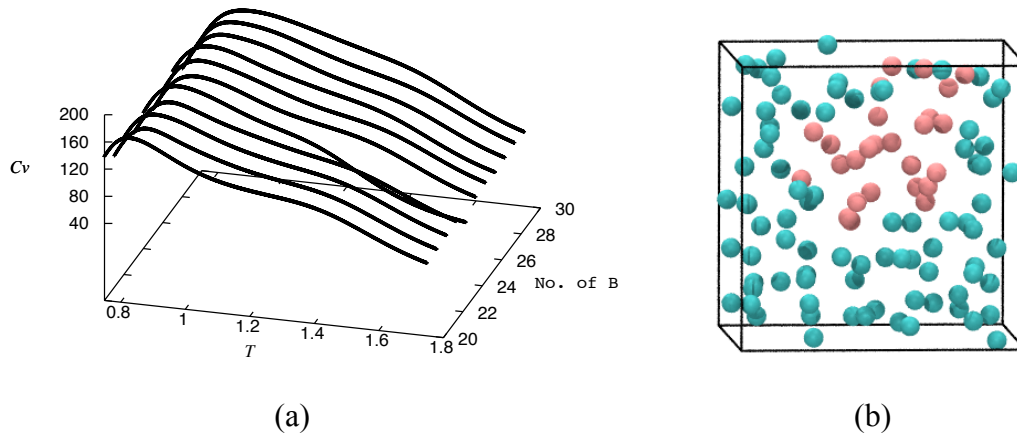
Figure 7.3 (a) Heat capacity calculated in the T range [0.7,1.8], and (b) VMD snapshot showing the state of phase separation of the binary mixture (sphere sizes are reduced to enable visualization)

(a)   (b)
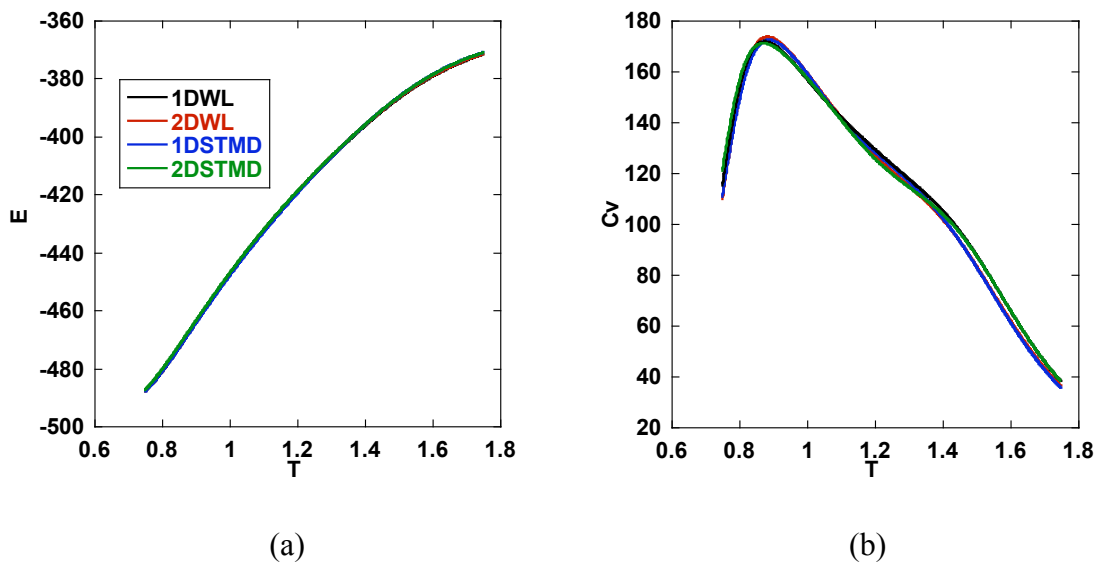
Figure 7.4 (a) Energy and (b) heat capacity from simulations by four different methods

To verify the results obtained here, we also ran the same system with the modified Wang-Landau method given by Shell *et al.*[107] (i.e., 2D WL); and comparisons of the average energy and heat capacity are shown in **Figure 7.4** for a specified system with N_B = 26, selected at random, but representative of the results across the entire range.

Additionally, 1D WL and 1D STMD simulations have also been run for the same system (i.e., $N_B = 26$) and also included in **Figure 7.4** for comparison. Nearly identical behavior is seen for E and heat capacity among the four different methods, verifying the results and proposed methodology.

7.3.1 Swap frequency

Here we examined three different sets of simulations that only differ in the relative frequency of WL composition swaps. Specifically, we examine systems where swapping is attempted ever 500, 1000, and 20000 STMD steps. The heat capacity of these systems as a function of swap frequency is shown in **Figure 7.5**, where we note that the curves are nearly identical for all frequencies. For the hybrid method, it is desirable for $\ln f_{WL}$ and $\ln f_{STMD}$ to both converge simultaneously or as closely to each other as possible. With a very frequent swap (i.e., 500), the convergence of the STMD part of the method will be the bottleneck, while for a very infrequent swap, the convergence of the WL part is delayed. Among the three sets of simulations, it seems that simulations with a swap frequency of 1000 STMD steps exhibit a good balance between the convergence of the two parts for the system being considered. Alternatively, instead of performing one WL swap move each cycle, we can also make multiple WL swap attempts for longer STMD cycles, as another means to balance the convergence of WL DOS and STMD $T(E)$.
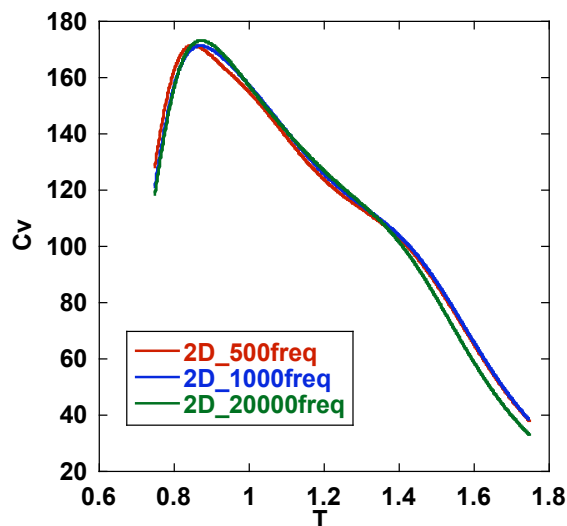
Figure 7.5 Heat capacity from simulations with different swap frequency

7.4   Extension to volume change by Wang-Landau

Similarly, we extended the hybrid 2D STMD method to the simulation in the *NPT* ensemble. Instead of composition swap by WL, volume-scaling moves are attempted each cycle. The system is composed of 110 particles with 26 of type B and 84 of type A, and the volume is restricted from 130 σ^3 to 230 σ^3, with a volume bin size set as $5\sigma^3$. Other simulation details are similar as above. Here we provided the calculated 2-dimensioal $g(E,V)$, and heat capacity for every volume considered, as shown in **Figure 7.6**. Generally, systems within large volumes have a more apparent peak in heat capacity curve, reflecting the larger energy changes in these systems compared to those with small volumes.

For comparison, 1D STMD and 2D WL simulations were also run for comparisons, as shown in **Figure 7.7** for several different volumes. We randomly showed three systems at different volume bins. Again quite similar behavior is observed for the energy and heat capacity, verifying the implementation of the proposed methodology.

94

(a)   (b)
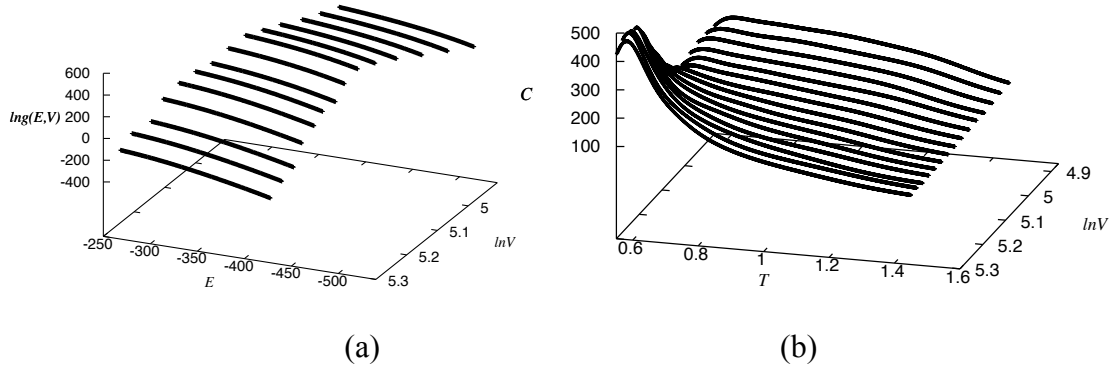
Figure 7.6 (a) Density of states $\ln g(E,V)$ and (b) heat capacity in the *NPT* ensemble
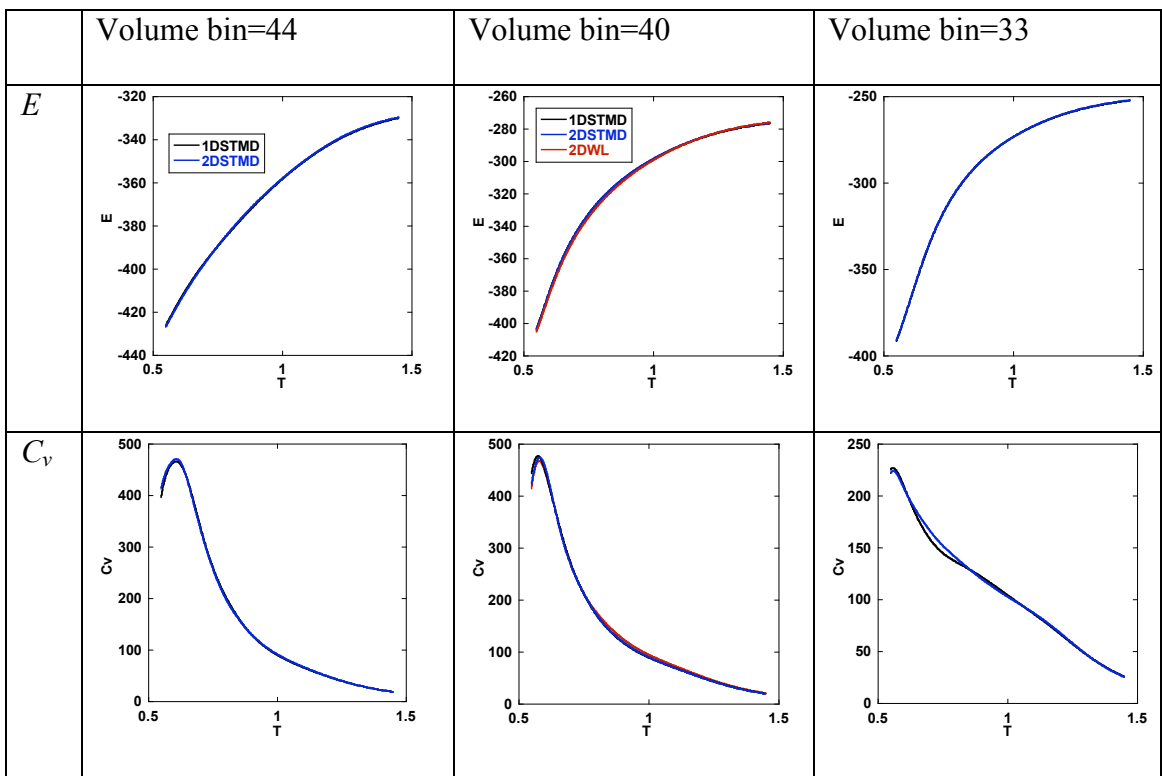
Figure 7.7 Comparisons of the results from different simulations. Note heat capacity of 1D STMD is an average of systems with low and high volume boundaries of a volume bin

CHAPTER 8   CONCLUSIONS AND FUTURE WORK

8.1   Conclusion

Through the application of advanced, high-fidelity computational tools and methods, significant new insight into the structural, and thermodynamic properties (especially free energy and heat capacity) and phase behavior of complex fluids has been provided in this dissertation. These findings are obtained based on efficient and valid density-of-states simulation methods. Extended algorithms for joint density-of-states calculations have also been proposed and validated, enabling study of the entire phase behavior of various ensembles, and thus provide insight that is not generally available by experiment.

Specially, in Chapter 4, the recently developed Wang-Landau algorithm was applied for the first time to study the process of self-assembly using a 3D lattice-based lipid model. The lipid model studied consists of 3 or 5 coarse-grained segments with the empty lattice sites representing water. The molecular interactions between hydrophobic and hydrophilic segments and between hydrophobic segments and water are purely repulsive. A bilayer structure is found to form at low temperatures, with phase transitions to clusters as temperature increases. With the complete density of states obtained, continuous curves of the average energy and heat capacity could be calculated and thus a complete view of the phase behavior provided. For 3-segment chains, varying lattice sizes were studied, with the observation that as the system size increases, the phase transition temperature shifts. Specifically, the ratio of chain number to lattice area (i.e., area per lipid) affects the phase transition temperature. At small ratios, only one phase transition occurs between the bilayer and cluster phases, while at high ratio when the

number of chains is enough to form an additional bilayer, the phase transition occurs in two steps with stable intermediate state. An additional simple CBMC move was implemented and applied to simulate longer 5-segment chains on a lattice of size $L = 20$ and similar phase transition behavior was observed as a function of temperature; however due to the increase in the required number of energy bins as the system size increases, application of the WL method to much larger system would be somewhat impractical.

The STMC method, an extension of the WL algorithm that allows simulations with larger energy bin sizes, was also implemented to study the self-assembly process. The results obtained are found to be in good agreement with the original WL method and largely independent of the energy bin size used. The possibility of running simulations with larger bin sizes by STMC with fewer MC steps is potentially an improvement over the original WL method; however, the longer time taken to perform an individual STMC step soon outweighs any additional benefits.

Thus, to examine complex or continuous systems, parallel simulation and extended to MD simulations are necessary. Two important systems were selected for the phase behavior study. Specifically, in Chapter 5, we examined the structural and phase behavior of self-assembled lipid bilayers modeled with off-lattice amphiphilic molecules in solutions. Simulations were performed using STMD and parallel replica-exchange WL MC methods, both of which allow the direct calculation of the density of states. We have demonstrated that these two very different methods provide essentially identical results when considering the density of states and key transitions in the heat capacity for the model lipid system studied, thereby validating both the methods and the results obtained,

and showing the applicability of these methods to determine the free energy of self-assembling lipid systems.

While both methods provide closely matching results for the systems studied herein, we note that there are considerations to take into account when selecting one method over the other. A particularly advantageous feature of STMD is that it does not require a reduction in the energy bin size as system size is increased, since the statistical temperature can be locally interpolated well by a linear function, making the results from STMD only weakly dependent on the energy bin size. However, to the best of our knowledge, no such interpolation scheme exists for the WL method. Therefore, REWL simulations use a much finer energy binning compared to STMD, making it more costly for all histogram bins to be filled for appropriate convergence. This can make STMD better suited for the examination of larger systems. However, as we observe in our calculations, the error bars associated with STMD tend to be larger than those obtained from REWL, which may be associated with the fact that STMD follows the "natural" dynamics dictated by the equations of motion which are absent in the MC moves underlying WL; as such, hysteresis effects around phase transitions may be more pronounced in STMD, influencing the ability to accurately resolve certain behaviors /transitions. However, both STMD and REWL are important tools, enabling the accurate calculation of the density of states over a wide range of systems and state points.

In general, the nature of the system and state points under consideration will play a key role in determining which method is most suited to a given application. Generally, applicability is largely determined by the underlying algorithm of each method, i.e., MD versus MC, and the way in which parallelism is implemented in that method. For

example, STMD relies on MD-based force calculations that, in this work, are implemented on massively-parallel GPUs; this makes STMD well suited for the efficient examination of large systems, since the MD algorithm is known to scale well on massively multi-threaded GPUs, distributed CPU computing machines, and hybrid CPU-GPU systems.[166, 167] STMD may also lend itself more easily to the study of complex molecules, e.g., biopolymers, again as a result of the parallel nature of the underlying MD calculations but also since it avoids having to develop nonphysical effective MC trial moves to overcome steric hindrance.[30, 31, 170] However, although the single-replica STMD used in this work is highly parallel with regards to the force computations, it may still experience problems associated with getting trapped in local energy minima at low temperature, making it prohibitively expensive to appropriately sample this regime. For example, even for the relatively small, coarse-grained system studied here, ~2 months of computation time was needed to study the lowest temperature window compared to ~5 days for the highest temperature window; however, we note that sufficient convergence was found over the entire range studied in this work. The ability to incorporate specialized MC moves into the WL method can be used to mitigate such problems, increasing sampling and avoid trapping (e.g., cluster moves and identity swaps). The REWL method is also less prone to such trapping issues at low temperatures as a result of the parallel sampling associated with the replica exchange MC moves; prior work has shown this approach to be superior over the original WL method without such exchanges[112] and has additionally been demonstrated for statistical temperature MC[165] (note, statistical temperature MC is closely related to STMD). We note that the replica-exchange technique has been implemented for MD-based computations, specifically in

the context of the STMD method, demonstrating increased performance as compared to a single-replica STMD.[162] However, in this work we chose to concentrate on the aforementioned parallelization direction and the efficient implementation on GPUs.

In Chapter 6, another important system, nano-confined fluids was also simulated. We studied the nature of the fluid-solid phase transition in nano-confined regions by performing STMD simulations. Heat capacity and free energy are calculated to indicate the phase behaivor and better understand the phase transition. For the ideal pore speparation systems, the T_{ODT} decreases to lower values when either the pore separation is increased or the wall-fluid strength ε_{wm} is reduced. Both of these agrees with the idea that a reduced interaction between mobile particles and walls requires a lower temperature to crystallize, and the behavior becomes more bulk like. For the systems with non-ideal spacing seperation, quite different results are obtained with a much lower T_{ODT}, compared to the ideal-pore spacings. The calculation of the free energy between different states at various pore separations shows strong evidence that the ordered structures we obtained are the states with low free energies.

To further the ability to use STMD to calculate the 2D DOS, Chapter 7 reported a hybrid WL/STMD method, making use of both the advantages of Wang-Landau and especially the STMD algorithms. To demonstrate and validate the proposed method, the 2D density of states was calculated in the semi-grand canonical ensemble and *NPT* ensemble with a binary mixture of LJ fluids. Nearly identical results are obtained compared to the known methods (1D/2D WL and standard STMD), when considering the calculated thermodynamic properties for the system studied, thereby validating both the methods and the results obtained.

The hybrid method is not without its own limitations. First, overlapping energy regions are required for performing the movement in the WL direction; thus if extending to wider changes in larger systems, simulation with small energy bin size and multi-windows may be inevitable. In addition, the many sets of simulations varying in composition or volume are actually independent from each other, but they need to wait for each other (to be converged to the current modification factor) so that the whole simulation can proceed to next iteration. The synchronization of this process could be time-consuming. However, for obtaining the 2D DOS, this is unavoidable. In principle, additional parallelization (i.e., multiple walkers) could be implemented to speed convergence. As discussed earlier, a particularly advantage of STMD is that it is only weakly dependent on the energy bin size. In addition, STMD scales better to larger systems and may lend itself more easily to the study of complex molecules, e.g., biopolymers, as a result of the parallel nature of the underlying MD calculations but also since it avoids having to develop nonphysical effective MC trial moves to overcome steric hindrance. All of these advantages led to the idea of hybrid method practical for simulations of larger systems. In sum, we have verified the feasibility of obtaining independent DOSs generated *via* efficient STMD, for calculating the continuous 2D DOS.

8.2   Future work

Despite the progress made, there are still many aspects that require further study and development. Below we discuss recommendations for future studies, both in the aspects of method development and applied systems.

8.2.1 Extended applications

The study of the self-assembly lipids and nano-confined systems, modeled with simple coarse-grained lipids and Lennard-Jones particles, respectively, demonstrated the great potential of the WL/STMD methods for studying the phase behavior of complex systems. Thus, an important next step is to apply these methods to the study of more realistic systems/models; several related systems of immediate interest are described below.

Skin lipids - self-assembly of ceramides

The stratum corneum is the outermost layer of the skin and is responsible for the barrier function in mammalian physiology. Ceramide lipids (CER), the main constituent of this layer, are a class of double-tailed lipids composed of hydrophobic sphingosine and acyl tails, connected together with a hydrophilic head group. From experimental work, a diverse mixture of CERs has been observed in the stratum corneum (12 different CERs), along with cholesterol and free fatty acid.[43, 44, 171] Although the lipid profile of the stratum corneum is well known, experiments are unable to discern the exact molecular level structure, and how the local interactions between these constituents influence barrier function. In previous work from our group,[44] MD simulations have been used to examine the behavior of ceramide bilayers, focusing on nonhydroxy sphingosine (NS) and nonhydroxy phytosphingosine (NP) ceramides. This work proposed a modified version of the CHARMM force field for ceramide simulations, showing better ability to match the results from the experiment (shown in **Figure 8.1**), when compared to the commonly used GROMOS force field. While this method could be directly examined using tools such as STMD, the simulations would likely be too computationally expensive for any practical exploration.

(a)   (b)   (c)
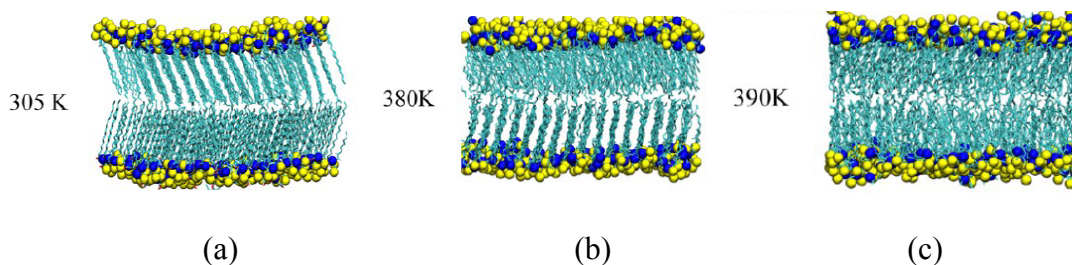
Figure 8.1 Snapshots of the observed phases for CER NS at different temperatures, and from reference;[44] (a) tilted gel bilayer phase (b) untilted gel bilayer phase (c) liquid-like bilayer phase.

Coarse-grained models have been developed and derived from the modified CHARMM force field, enabling the efficient and accurate examination of the phase behavior of CERS as a function of temperature by the STMD method. Such examinations would focus on the heat capacity for detecting all the major transitions as well as subtle intermediate order-order transitions known to occur; our prior examinations with the simplified model outlined in Chapter 5 make us confident in the sensitivity of the heat capacity to determine such transitions if they exist. The nature of the flat histogram sampling in STMD may also prove useful for the formation of equilibrium, rather than metastable phases, as the method allows equal sampling of all phases, eliminating the need to worry about kinetic traps in the phase landscape. Preliminary work has already been undertaken, examining the coarse-grained CER3 system, which consists of 200 lipids and 5000 water molecules. The CER3 molecule is coarse-grained to three head group beads and seven tail group beads show in **Figure 8.2** . At a fixed volume with V, C_v is calculated in **Figure 8.2** , and a simulation snapshot is shown in **Figure 8.3**, demonstrating the formation of an ordered gel bilayer structure. While these preliminary results are promising, it is important to note that the corresponding traditional MD simulations are performed in the *NPT* ensemble, while

STMD is limited to *NVT*. The use of *NVT* may give rise to stressed or defective bilayer structures, since the system cannot adjust the bilayer cross section, which will influence the area per lipid. As such, continued studies will require the use of the hybrid WL/STMD method, previously discussed in Chapter 7.
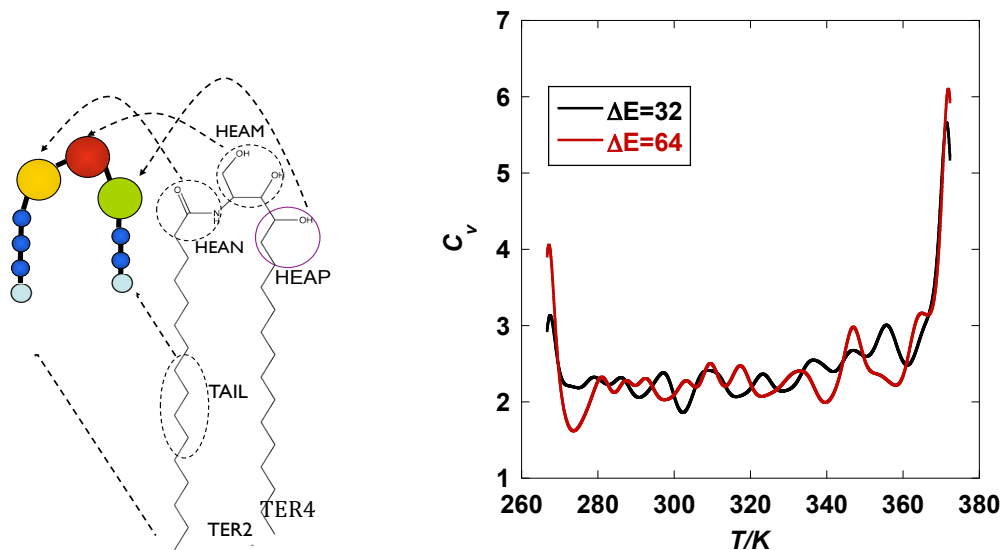
Figure 8.2 (a) Schematic depication of the coarse-grained CER3, (b) heat capacity from STMD with different energy bin sizes.
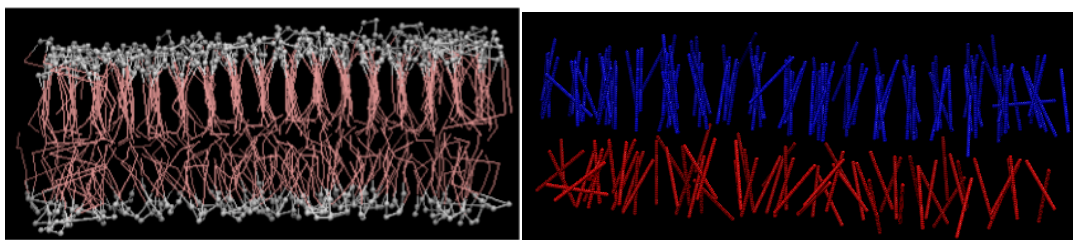
Figure 8.3 Snapshot for the structure of lipid bilayer at *T*=270K. The water particles are not shown, and the right one only shows the tails for better view

Nano-confined thin films of non-polar fluids

As discussed earlier in the thesis, when a liquid is nano-confined in a narrow gap (such as in a porous medium, in a lubricated contact between solids, or near a cell membrane),[76] the dynamic behavior is changed. The phenomenon has been experimentally detected for

various systems such as octamethylcyclotetrasiloxane, cyclohexane, toluene, and alkanes.[71] In the work reported previously from our research group,[68, 71, 72, 92, 93] the dynamical properties for an alkane (dodecane) ultra-thin film confined between solid surfaces were extensively studied. Dodecane ($C_{12}H_{26}$) with a thickness of the chain as ~4 Å, has already been extensively studied by experimental work of Granick *et al.*[76, 81] In the work of Cui *et al.*,[92] where dodecane is described by a united atom model, and the solid walls composed of simple LJ atoms, solid-like structure was found to form when the film thickness reduced to 6 molecular layers (shown in **Figure 8.4**).

(a)   (b)
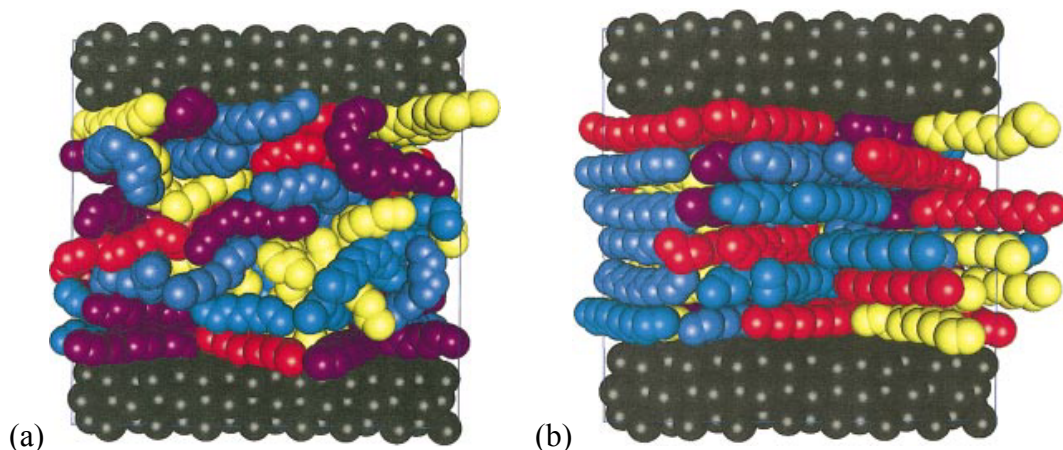
Figure 8.4 The equilibrium configuration of the confined dodecane with a film thickness of (a) 8 molecular layers, (b) 6 molecular layers. The solid wall atoms are shown in black. Individual fluid molecules are colored differently for easy distinction. All are from reference.[92]

As of yet, only traditional MD simulation has been used to study this phase transition. The STMD method is uniquely suited to study this transitions, in much the same way as described in Chapter 6 for simple nanoconfined LJ spheres, i.e., GCMD STMD simulations. Here it is proposed to study the nature of these phase behaviors of the alkanes films by calculating free energy and heat capacity as a function of temperature, with various mica wall separations. Furthermore, other complex systems, so as the

aforementioned cyclohexane, can be examined to provide a better link to experimental studies.

8.2.2   2D DOS method

Efficiency

When examining lower temperatures and more complex molecule system, the computational time cost associated with calculating the 2D DOS will be increased greatly. As previously discussed in the context of WL MC, replica-exchange can be used to greatly speed up convergence.[135] It has also been demonstrated that replica-exchange can be used with 1D STMD.[162] As such, it is an important next step to include replica exchange as part of the hybrid WL/STMD code developed to calculate the 2D DOS. Also, since the WL part is likely to converge more slowly than the STMD part in most cases, the algorithm can be modified slightly so as to increase WL sampling. For example, similar to the concept of having multiple walkers, multiple configurations from the STMD trajectory can be used as starting points to N unique WL walkers.

Application to diblock copolymer

Block copolymer consists of two or more homopolymer subunits linked by covalent bonds, where systems with two distinct blocks (A and B) are called diblock copolymers. Diblock copolymers are attractive because they can microphase separate to form various nanostrutrues such as sheets and cylinders, and will be particularly useful in the self-assembly of nanobuilding blocks (e.g. creating devices for use in computer memory, nanoscale-templating and nanoscale separations).[172] Similar to the ammphiphilic lipids studied in this thesis, the incompatibility between the blocks drives the block copolymers

themselves self-assemble into complex morphologies as they attempt to minimize their free energy by aggregating with species of their own kind; since the blocks are covalently linked to each other, they cannot form macroscopic domains separately, instead forming many complex phases.[15, 173]

The differences in relative lengths of each block (asymmetry) leads to different morphologies. In prior studies[173], blocks with similar length tend to form lamellar phases, while larger difference in block lengths lead to "hexagonally packed cylinder" geometry, shown in **Figure 8.5**. Accurately locating phase transitions between different spatially structured phases formed by the self-assembling of these systems is a computational challenge. Most simulation studies consider a fixed block fraction, calculating phase transitions as a function of temperature. The transition as a function of block fraction is typically interpolated with little else done to quantify the transition as a function of block fraction. With the 2D DOS applied in the semi-grand canonical ensemble (i.e., the relative fraction of A and B can be changed, however the total system is fixed), the phase behavior can be calculated as a both a function of temperature and block fraction. Further study of related systems, such as melts of amphiphilic polymer grafted nanoparticles, are also of interest and could be examined using the same approach.
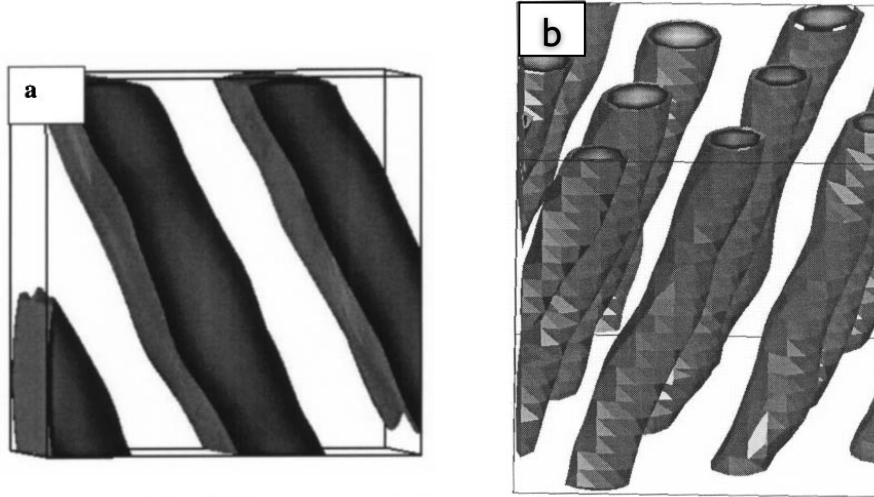
Figure 8.5 Different structures of block copolymer composed of A and B; (a) lamellar phase obtained with $\chi N=52.7$, $f_A=0.5$, (b) hexagonal cylinder at $\chi N=57.9$, $f_A=0.2$. And f_A is the relative fraction of A block.

REFERENCES

D. L. Beveridge and F. M. DiCapua, Annual Review of Biophysics and Biophysical Chemistry **18** (1), 431-492 (1989).

D. P. Landau and K. Binder, *A Guide to Monte Carlo Methods in Statistical Physics*. (Cambridge University Press, Cambridge, 2000).

B. A. Patel, P. G. Debenedetti and F. H. Stillinger, The Journal of Physical Chemistry A **111** (49), 12651-12658 (2007).

F. Wang and D. P. Landau, Physical Review E **64** (5), 056101 (2001).

D. A. Kofke, Fluid Phase Equilib. **228-229**, 41 (2005).

P. Kollman, Chem. Rev. **93**, 2395 (1993).

C. A. Reynolds, P. M. King and W. G. Richards, Molecular Physics **76** (2), 251-275 (1992).

L. D. Gelb, K. E. Gubbins, R. Radhakrishnan and M. Sliwinska-Bartkowiak, Reports on Progress in Physics **62** (12), 1573 (1999).

M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquid*. (Oxford, Clarendon, 1987).

D. Frenkel, *Computer Modeling of Fluids, Polymers and Solids, NATO Advanced Study Institute Series*. (Deventer, Kluwer, 1989).

P. Singh, S. K. Sarkar and P. Bandyopadhyay, Chemical Physics Letters **514** (4), 357-361 (2011).

S. Singh, M. Chopra and J. J. de Pablo, Annual Review of Chemical and Biomolecular Engineering **3** (1), 369-394 (2012).

T. P. Straatsma and J. A. McCammon, The Journal of Chemical Physics **95** (2), 1175-1188 (1991).

J. G. Kirkwood, The Journal of Chemical Physics **3** (5), 300-313 (1935).

M. Muller and K. C. Daoulas, J. Chem. Phys. **128**, 024903 (2008).

F. M. Ytreberg, R. H. Swendsen and D. M. Zuckerman, J. Chem. Phys. **125**, 184114 (2006).

R. W. Zwanzig, The Journal of Chemical Physics **22** (8), 1420-1426 (1954).

S. Park and F. Khalili-Araghi, J. Chem. Phys. **119**, 3559 (2003).

G. E. Crooks, Physical Review E **61** (3), 2361-2366 (2000).

D. J. Evans, Molecular Physics **101** (10), 1551-1554 (2003).

D. A. Kofke and P. T. Cummings, Mol. Phys **92**, 973 (1997).

Y. Okamoto, Journal of Molecular Graphics and Modelling **22** (5), 425-439 (2004).

A. M. Ferrenberg and R. H. Swendsen, Physical Review Letters **63** (12), 1195-1198 (1989).

B. A. Berg and T. Celik, Physical Review Letters **69** (15), 2292-2295 (1992).

J. Lee, Physical Review Letters **71** (2), 211-214 (1993).

K. Hukushima and K. Nemoto, Journal of the Physical Society of Japan **65** (6), 1604-1608 (1996).

A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov and P. N. Vorontsov, The Journal of Chemical Physics **96** (3), 1776-1783 (1992).

E. Marinari and G. Parisi, Europhys. Lett. **19** (6) (1992).

F. Wang and D. P. Landau, Physical Review Letters **86** (10), 2050-2053 (2001).

J. Kim, J. E. Straub and T. Keyes, Physical Review Letters **97** (5), 050601 (2006).

J. Kim, J. E. Straub and T. Keyes, The Journal of Chemical Physics **126** (13), 135101 (2007).

A. Drefahl, M. Wahab, P. Schiller and H.-j. Mogel, Thin Solid Films **327**, 846 (1998).

D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*. (Academic Press, San Diego, 2002).

H. Noguchi and M. Takasu, Physical Review E **64**, 041913 (2001).

M. Kenward and M. D. Whitmore, The Journal of Chemical Physics **116** (8), 3455-3470 (2002).

C. M. Care, J. Phys. C: Solid State Phys **20**, 1 (1987).

D. Brindle and C. M. Care, Mol. Simul. **5** (1990).

A. T. Bernardes, Langmuir **12** (24), 5763-5767 (1996).

A. T. Bernardes, Langmuir **12**, 5763 (1996).

M. G. Wolf, J. A. Jongejan, J. D. Laman and S. W. de Leeuw, The Journal of Physical Chemistry B **112** (43), 13493-13498 (2008).

S. Marrink, E. Lindahl, O. Edholm and A. Mark, Journal of the American Chemical Society **123** (35), 8638-8639 (2001).

G. Brannigan and F. L. H. Brown, *Coarse-Graining of Condensed Phase and Biomolecular Systems*. (CRC Press, Boca Raton, 2009).

D. T. Downing, M. E. Stewart, P. W. Wertz, S. W. Colton Vi and J. S. Strauss, Comparative Biochemistry and Physiology Part B: Comparative Biochemistry **76** (4), 673-678 (1983).

S. Guo, T. C. Moore, C. R. Iacovella, L. A. Strickland and C. McCabe, Journal of Chemical Theory and Computation **9** (11), 5116-5126 (2013).

K. A. Riske, L. Q. Amaral, H. G. Döbereiner and M. T. Lamy, Biophysical Journal **86** (6), 3722-3733 (2004).

V. Sundararajan, *Computational Modeling of Membrane Bilayers*. (Academic Press, San Diego, 2008).

R. G. Larson, J. Chem. Phys. **83**, 2411 (1985).

R. G. Larson, J. Chem. Phys. **89** (3), 1642 (1988).

R. G. Larson, J. Chem. Phys. **96**, 7904 (1992).

A. T. Bernardes, V. B. Henriques and P. M. Bisch, J. Med. Phys. **101**, 645 (1994).

T. B. Liverpool and A. T. Bernardes, J.Phys.II France **5**, 1457 (1995).

C. M. Care, Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases **83** (9), 2905-2912 (1987).

D. Brindle and C. M. Care, J. Chem. Soc., Faraday Trans **88**, 2163 (1992).

C. Tanford, *Hydrophobic Effect: Formation of Micelles and Biological Membranes*. (Wiley-Interscience, New York, 1980).

S. Fujiwara, D. Funaoka, T. Itoh and M. Hashimoto, Comput. Phys. Commun. **182**, 192 (2011).

S. Fujiwara, T. Itoh, M. Hashimoto and R. Horiuchi, J. Chem. Phys. **130**, 144901 (2009).

G. Mohan and D. I. Kopelevich, J. Chem. Phys. **128**, 044905 (2008).

A. Milchev, A. Bhattacharya and K. Binder, Macromolecules **34** (6), 1881-1893 (2001).

J. M. Rodgers, J. Sorensen, F. J. M. de Meyer, B. Schiott and B. Smit, The Journal of Physical Chemistry B **116** (5), 1551-1569 (2012).

T. Nagai, R. Ueoka and Y. Okamoto, Journal of the Physical Society of Japan **81**, 024002 (2012).

S. J. Marrink, J. Risselada and A. E. Mark, Chemistry and Physics of Lipids **135** (2), 223-244 (2005).

S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. de Vries, The Journal of Physical Chemistry B **111** (27), 7812-7824 (2007).

V. Knecht, A. E. Mark and S.-J. Marrink, Journal of the American Chemical Society **128** (6), 2030-2034 (2006).

R. Goetz and R. Lipowsky, Journal of Chemical Physics **108** (17), 7397-7409 (1998).

S. Fujiwara, T. Itoh, M. Hashimoto and R. Horiuchi, Journal of Chemical Physics **130** (14), 144901-144908 (2009).

S. J. Marrink, A. H. de Vries and A. E. Mark, The Journal of Physical Chemistry B **108** (2), 750-760 (2003).

H. J. Risselada and S. J. Marrink, Proceedings of the National Academy of Sciences **105** (45), 17367-17372 (2008).

P. T. Cummings, H. Docherty, C. R. Iacovella and J. K. Singh, AIChE Journal **56** (4), 842-848 (2010).

U. Jonsson and B. Bhushan, Journal of Applied Physics **78** (5), 3107-3114 (1995).

C. R. Iacovella, P. T. Cummings, H. Dochertyc, J. K. Singh, L. Gai and L. Wan, Soft Matter, submitted (2014).

S. T. Cui, P. T. Cummings and H. D. Cochran, The Journal of Chemical Physics **111** (3), 1273-1280 (1999).

H. Docherty and P. T. Cummings, Soft Matter **6** (8), 1640-1643 (2010).

J. Mittal, T. M. Truskett, J. R. Errington and G. Hummer, Physical Review Letters **100** (14), 145901 (2008).

A. L. Demirel and S. Granick, Physical Review Letters **77** (11), 2261-2264 (1996).

M. L. Gee, P. M. McGuiggan, J. N. Israelachvili and A. M. Homola, The Journal of Chemical Physics **93** (3), 1895-1906 (1990).

S. Granick, Science **253** (5026), 1374-1379 (1991).

J. Israelachvili, Y. Min, M. Akbulut, A. Alig, G. Carver, W. Greene, K. Kristiansen, E. Meyer, N. Pesika, K. Rosenberg and H. Zeng, Reports on Progress in Physics **73** (3), 036601 (2010).

J. Klein and E. Kumacheva, The Journal of Chemical Physics **108** (16), 6996-7009 (1998).

J. Klein and E. Kumacheva, Science **269** (5225), 816-819 (1995).

E. Kumacheva and J. Klein, The Journal of Chemical Physics **108** (16), 7010-7022 (1998).

H. W. Hu, G. A. Carson and S. Granick, Physical Review Letters **66** (21), 2758-2761 (1991).

J. Van Alsten and S. Granick, Langmuir **6** (4), 876-880 (1990).

B. Coasne, J. Czwartos, M. Sliwinska-Bartkowiak and K. E. Gubbins, The Journal of Chemical Physics **133** (8), 084701 (2010).

R. Radhakrishnan, K. E. Gubbins, A. Watanabe and K. Kaneko, The Journal of Chemical Physics **111** (19), 9058-9067 (1999).

M. Sliwinska-Bartkowiak, J. Gras, R. Sikorski, R. Radhakrishnan, L. Gelb and K. E. Gubbins, Langmuir **15** (18), 6060-6069 (1999).

R. Radhakrishnan, K. E. Gubbins and M. Sliwinska-Bartkowiak, The Journal of Chemical Physics **112** (24), 11048-11057 (2000).

P. A. Thompson and M. O. Robbins, Science **250** (4982), 792-794 (1990).

A. Jabbarzadeh, J. D. Atkinson and R. I. Tanner, Macromolecules **36** (13), 5020-5031 (2003).

R. E. Belardinelli and V. D. Pereyra, Journal of Chemical Physics **127** (18), 184105-184107 (2007).

Y. Wang, K. Hill and J. G. Harris, Langmuir **9** (8), 1983-1985 (1993).

J. Klein and E. Kumacheva, Physica A: Statistical Mechanics and its Applications **249** (1-4), 206-215 (1998).

S. T. Cui, P. T. Cummings and H. D. Cochran, The Journal of Chemical Physics **114** (16), 7189-7195 (2001).

S. T. Cui, P. T. Cummings and H. D. Cochran, The Journal of Chemical Physics **114** (14), 6464-6471 (2001).

L. Gai, K. A. Maerzke, P. T. Cummings and C. McCabe, The Journal of Chemical Physics **137** (14), 144901 (2012).

Y. W. Li, T. Wüst and D. P. Landau, Physical Review E **87** (1), 012706 (2013).

T. Wüst, Y. Li and D. Landau, J. Stat. Phys **144** (3), 638-651 (2011).

D. F. Parsons and D. R. M. Williams, Physical Review E **74** (4), 041804 (2006).

D. T. Seaton, S. J. Mitchell and D. P. Landau, Brazilian Journal of Physics **38**, 48-53 (2008).

D. T. Seaton, T. Wüst and D. P. Landau, Physical Review E **81** (1), 011802 (2010).

Y. Sliozberg and C. F. Abrams, Macromolecules **38** (12), 5321-5329 (2005).

M. P. Taylor, W. Paul and K. Binder, The Journal of Chemical Physics **131** (11), 114907 (2009).

D. T. Seaton, S. J. Mitchell and D. P. Landau, Braz. J. Phys **38** (1), 48-53 (2007).

C. Gervais, T. Wüst, D. P. Landau and Y. Xu, The Journal of Chemical Physics **130** (21), 215106 (2009).

J. S. Yang and W. Kwak, Computer Physics Communications **181** (1), 99-104 (2010).

C. Zhou, T. C. Schulthess, S. Torbrügge and D. P. Landau, Physical Review Letters **96** (12), 120201 (2006).

Q. Yan, R. Faller and J. J. de Pablo, The Journal of Chemical Physics **116** (20), 8745-8749 (2002).

M. S. Shell, P. G. Debenedetti and A. Z. Panagiotopoulos, Physical Review E **66** (5), 056703 (2002).

P. Poulain, F. Calvo, R. Antoine, M. Broyer and P. Dugourd, Physical Review E **73** (5), 056704 (2006).

A. E. Ferdinand and M. E. Fisher, Phys. Rev. E **185**, 832-846 (1969).

A. G. Cunha-Netto, A. A. Caparica, S.-H. Tsai, R. Dickman and D. P. Landau, Phys. Rev. E **78**, 055701 (2008).

B. J. Schulz, K. Binder, M. Muller and D. P. Landau, Physical Review E **67** (6), 067102 (2003).

T. Vogel, Y. W. Li, T. Wüst and D. P. Landau, Physical Review Letters **110** (21), 210603 (2013).

J. Yin and D. P. Landau, Computer Physics Communications **183** (8), 1568-1573 (2012).

L. Zhan, Computer Physics Communications **179** (5), 339-344 (2008).

M. O. Khan, G. Kennedy and D. Y. C. Chan, Journal of Computational Chemistry **26** (1), 72-77 (2005).

D. P. Landau, S.-H. Tsai and M. Exler, American Journal of Physics **72** (10), 1294-1302 (2004).

D. A. McQuarrie, *Statistical mechanics*. (University Science Books, Sausalito, 2000).

A. Mitsutake, Y. Sugita and Y. Okamoto, Peptide Science **60** (2), 96-123 (2001).

C. Chipot, M. S. Shell and P. Andrew, *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. (Springer, New York, 2007).

M. R. Wilson, Journal of Molecular Liquids **68** (1), 23-31 (1996).

A. S. Keys, C. R. Iacovella and S. C. Glotzer, Journal of Computational Physics **230** (17), 6438-6463 (2011).

A. S. Keys, C. R. Iacovella and S. C. Glotzer, Annual Review of Condensed Matter Physics **2** (1), 263-285 (2011).

J. A. Anderson, C. D. Lorenz and A. Travesset, Journal of Computational Physics **227** (10), 5342-5359 (2008).

http://codeblue.umich.edu/hoomd-blue.

M. Kenward and M. D. Whitmore, Journal of Chemical Physics **116** (8), 3455-3470 (2002).

S. J. Marrink, E. Lindahl, O. Edholm and A. E. Mark, J. Am. Chem. Soc. **123**, 8638 (2001).

M. G. Wolf, J. A. Jongejan, J. D. Laman and S. W. d. Leeuw, J. Phys. Chem. B **112**, 13493 (2008).

F. Wang and D. P. Landau, Phys. Rev. E. **64** (2001).

K. F. Lau and K. A. Dill, Macromolecules **22** (10), 3986-3997 (1989).

S. A. Jonsson, S. Mohanty and A. Irback, J. Chem. Phys. **135** (12), 125102-125107 (2011).

T. S. Jain and J. J. d. Pabloa, J. Chem. Phys. **116**, 7238 (2002).

J. l. Siepmann and D. Frenkel, Mol. Phys **75**, 59 (1992).

A. D. Swetnam and M. P. Allen, Journal of Computational Chemistry **32** (5), 816-821 (2011).

C. Zhou and J. Su, Physical Review E **78** (4), 046705 (2008).

L. Gai, T. Vogel, K. A. Maerzke, C. R. Iacovella, D. P. Landau, P. T. Cummings and C. McCabe, The Journal of Chemical Physics **139** (5), 054505 (2013).

J. Kim, J. E. Straub and T. Keyes, Physical Review Letters **97** (5), 050601 (2006).

E. Bittner, A. Nussbaumer and W. Janke, Physical Review Letters **101** (13), 130603 (2008).

D. Chandler and H. C. Andersen, Journal of Chemical Physics **54** (1), 26-33 (1971).

H. R. Warner, Industrial & Engineering Chemistry Fundamentals **11** (3), 379-387 (1972).

C. Zhou and R. N. Bhatt, Physical Review E **72** (2), 025701 (2005).

E. J. Newman and G. T. Barkema, *Monte Carlo methods in statistical physics*. (Oxford University Press New York, 1999).

S. Schnabel, D. T. Seaton, D. P. Landau and M. Bachmann, Physical Review E **84** (1), 011127 (2011).

D. H. E. Gross, *Microcanonical Thermodynamics* (World Scientific, Singapore, 2001).

J. Gross, T. Neuhaus, T. Vogel and M. Bachmann, Journal of Chemical Physics **138** (7), 074905-074908 (2013).

F. Gulminelli and P. Chomaz, Physical Review E **66** (4), 046108 (2002).

H. Dominguez, M. P. Allen and R. Evans, Molecular Physics **96** (2), 209-229 (1999).

L. Wan, C. R. Iacovella, T. D. Nguyen, H. Docherty and P. T. Cummings, Physical Review B **86** (21), 214105 (2012).

L. Gai, L. Wan, C. R. Iacovella, P. T. Cummings and C. McCabe, J. Chem. Phys. **in preparation** (2014).

G. M. Torrie and J. P. Valleau, Chemical Physics Letters **28** (4), 578-581 (1974).

D. Frenkel and A. J. C. Ladd, The Journal of Chemical Physics **81** (7), 3188-3193 (1984).

T. Kaneko, T. Mima and K. Yasuoka, Chemical Physics Letters **490** (4-6), 165-171 (2010).

C. R. Iacovella, A. S. Keys and S. C. Glotzer, Proceedings of the National Academy of

L. Gai, C. R. Iacovella, P. T. Cummings and C. McCabe, J. Chem. Phys. **in preparation** (2014).

B. A. Berg and T. Neuhaus, Physical Review Letters **68** (1), 9-12 (1992).

M. Fitzgerald, R. R. Picard and R. N. Silver, Journal of Statistical Physics **98** (1-2), 321-345 (2000).

G. R. Smith and A. D. Bruce, Journal of Physics A: Mathematical and General **28** (23), 6623 (1995).

J. S. Wang and R. Swendsen, Journal of Statistical Physics **106** (1-2), 245-285 (2002).

R. Faller and J. J. de Pablo, The Journal of Chemical Physics **119** (8), 4405-4408 (2003).

P. N. Vorontsov-Velyaminov, N. A. Volkov and A. A. Yurchenko, Journal of Physics A: Mathematical and General **37** (5), 1573 (2004).

G. Ganzenmüller and P. J. Camp, The Journal of Chemical Physics **127** (15), 154504 (2007).

J. Yin and D. P. Landau, The Journal of Chemical Physics **134** (7), 074501 (2011).

M. S. Shell, P. G. Debenedetti and A. Z. Panagiotopoulos, The Journal of Physical Chemistry B **108** (51), 19748-19755 (2004).

D. Jayasri, V. S. S. Sastry and K. P. N. Murthy, Physical Review E **72** (3), 036702 (2005).

K. Mukhopadhyay, N. Ghoshal and S. K. Roy, Physics Letters A **372** (19), 3369-3374 (2008).

D. A. McQuarrie, *Statistical Mechanics*. (University Science Books, Sausalito, 2000).

W. M. Brown, A. Kohlmeyer, S. J. Plimpton and A. N. Tharrington, Computer Physics Communications **183** (3), 449-459 (2012).

W. M. Brown, P. Wang, S. J. Plimpton and A. N. Tharrington, Computer Physics Communications **182** (4), 898-911 (2011).

J. S. Lintuvuori and M. R. Wilson, The Journal of Chemical Physics **132** (22), 224902 (2010).

J. Kim, J. E. Straub and T. Keyes, The Journal of Physical Chemistry B **116** (29), 8646-8653 (2012).

J. Kim, J. E. Straub and T. Keyes, Physical Review E **76** (1), 011913 (2007).

G. S. K. Pilgram, D. C. J. Vissers, H. van der Meulen, S. Pavel, S. P. M. Lavrijsen, J. A. Bouwstra and H. K. Koerten, **117** (3), 710-717 (2001).

O. Gazit, R. Khalfin, Y. Cohen and R. Tannenbaum, The Journal of Physical Chemistry C **113** (2), 576-583 (2008).

M. A. Horsch, Z. Zhang, C. R. Iacovella and S. C. Glotzer, The Journal of Chemical Physics **121** (22), 11455-11462 (2004).