©Copyright 2019
Nihit Pokhrel

Understanding lipid membranes' interactions with small molecules and cholesterol using molecular dynamics computer simulations.

Nihit Pokhrel

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Lutz Maibaum, Chair

Anne B. McCoy

Bruce H. Robinson

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Understanding lipid membranes' interactions with small molecules and cholesterol using molecular dynamics computer simulations.

Nihit Pokhrel

Chair of the Supervisory Committee:
Assistant Prof. Lutz Maibaum
Department of Chemistry

Cellular membranes are made up of phospholipids, proteins and cholesterol. The packing of components within the membrane gives rise to important biological phenomena. Understanding these interactions is the most critical step towards predicting membrane behavior. In this dissertation, I focus on lipid bilayer interactions with various molecules. At biologically relevant scale, we investigate how phospholipids interact with small molecules like amino acid residues and cholesterol. At field scale, we study the structure and dynamics of lipid molecules as they bind with clay minerals. I utilize molecular dynamics simulation, both equilibrium and biased, at both coarse-grained and atomistic level to study how these interactions impact membrane thermodynamic properties. I first study membrane permeation by evaluating the efficiency of different computational methods to sample the configurational space of various amino acid and lipid membranes. I show that replica exchange umbrella sampling out-performs others because of the algorithm's ability to sample transition state. Using this method, I calculate cholesterol chemical potential in seven binary lipid membranes. These results demonstrate the sensitivity of the cholesterol chemical potential to the lipid tail unsaturation only if the difference in tail saturation is big, and show that cholesterol has the greatest affinity to saturated PC lipids. These studies provide practical guidance for studying membrane permeation and yield important insight into the

thermodynamic properties of lipid bilayer systems. This work will also contribute towards the understanding of cholesterol's effect of membranes, which is a vast and still ongoing field of research. Finally, this research also sheds lights on soil water repellency at the molecular scale. I investigate structural properties of organic matter that are found in soil and find that zwitterionic lipid aggregates bind to the surface of the clay. In this relatively new field, such molecular simulations will have significant scientific impact.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My graduate study would not have been possible without the guidance, support, help, and love of my family. My parents, Mr. Chiranjibi Prasad Upadhyaya and Mrs. Bimala Sharma have always been my source of inspiration. They have taught me the value of persistence and hard work. My sister, Ankit Pokhrel, has been the most prominent source of positivity.

I would like to thank Professor Lutz Maibaum for the opportunity to work on this research and for his input and guidance throughout my time in his research group. His high standards and boundless energy has made my PhD experience amazing! All the members of the Maibaum group, particularly Dr. Shushan He and Dr. Kayla Sapp, for their support and friendship. I have also had an immense support from the members of the Amphiphiliphiles group. I am grateful to Brenda Kessenich and Professor Jim De Yoreo for driving me in the field of soil science.

Finally, my friends have been the biggest support system during graduate school. I would like to thank Soonee Tan, Julia Nguyen, and Sujata Chakroborty who went through the entire PhD experience with me and have motivated me to stay focused throughout. Arushi Sonkhya deserves a special shout-out for always being there to listen to me.

DEDICATION

to my father, Chiranjibi Prasad Upadhyaya

Chapter 1

INTRODUCTION

1.1 Small Molecule Interactions With Lipid Bilayers

The cellular membrane is the barrier around the cell that separates its internal environment from the surroundings. The tight packing of phospholipids, sterols, and other membrane components together with the hydrophobic interior of the membrane mediate many processes that are vital for cellular organisms. The complex nature of the membrane system makes it a challenging as well as a very widely studied topic.[1–3] The enormous complexity of membrane structure is reduced to a few components by building model lipid systems with controlled lipid compositions, both in experimental and simulation studies. These systems will allow us to better understand the complex processes related to the cellular membrane.

One of the principal roles of biological membranes is to prevent uncontrolled flux of material across the lipid bilayer. This allows, for example, the interior of a cell to form a chemical environment that is distinct from that in the extracellular space. However, not all molecules are hindered from crossing the membrane to the same extent; especially small and uncharged solutes exhibit appreciable permeation rates.[4–6] Being able to predict whether a molecule can enter a cell by passive diffusion across the membrane is critical.[7–10] Small peptides that are positively charged are known to permeate membranes. The exact mechanism of this permeation is unknown and is heavily researched[11–13] in the last decade with experimental and computational techniques. Previous atomistic molecular dynamics studies have shown that membrane permeation is an extremely challenging computational problem .[14–16]

Currently, there is a growing interest in understanding the mechanism of membrane translocation, and many computational methods to predict a peptide's ability to translocate across a membrane have been developed.[17–25] In order to determine if any of the exist-

ing methods provides the best estimate of thermodynamic properties associated with these processes, we compare various computational methods available by computing free energy profiles of several solutes permeating a model bilayer.

Cholesterol is a major component of the cellular membrane, and it plays a vital role in maintaining the fluidity and rigidity of cell membrane . The presence of cholesterol in lipid membranes can induce order, increase packing and decrease membrane permeability.[?,26–28] It is known that cholesterol is enriched in the liquid-ordered phase, which consists of rigid and saturated lipids compared to liquid disordered phase, which generally consists of more flexible unsaturated lipids.[29–33] We quantify this preference of cholesterol by computing the chemical potential in different lipid environments that vary in cholesterol concentration, lipid headgroup type and the degree of unsaturation in the fatty acid lipid tail.

1.2 Structure and Aggregation of Organic Matter on Clay Surface

Soil water repellency is a process where soil has reduced affinity to water. Water-repellent soils resist water infiltration, and they have increased erosion and decreased ability to support crops.[34,35] The large carbon pool in soil, which could either come from an organic coating on soil particles or decomposing organic matter, is known to impact repellency.[36,37] Field studies have shown that both the amount of clay in the soil and the seasonal cycle affects soil water repellency, but there is a lack of mechanistic understanding of soil water repellency.[38] It has been hypothesized that the reorientation of amphiphilic compounds controls soil water repellency. In a collaborative effort with the De Yoreo group at Pacific Northwest National Lab to understand soil water repellency at the molecular level, we study the impact of organic matter, namely lipids, on the surface of clay montmorillonite. Montmorillonite is a phyllosilicate clay mineral, which is a layered mineral, where each layer is a "sandwich" of silica/alumina/silica. The spacing between each layer is about 1 nm and the interlayer is filled with water and ions. In order to fully address the mechanistic details of montmorillonite interaction with organic matter, we use molecular dynamics simulation to investigate lipids binding to the montmorillonite surfaces. The use of molecular dynamics to gain insight into

soil water repellence is relatively new. This is particularly important in this field because previous studies have left the molecular interactions to speculation.

1.3 Specific Aims

The work in this thesis is divided into three main components. In the first part, we focus on membrane permeation. Using atomistic molecular dynamics (MD) simulations with enhanced sampling techniques, we study the mechanism of permeation of model membrane. We compare the efficiency of various enhanced sampling methods in order to establish which method is superior and can most accurately compute the free energy profile. By studying the physical interaction between the permeant and the lipids, we also identify possible sampling problems and show how some enhanced sampling methods fail to overcome these problems due to the nature of the algorithm. In the second chapter, we quantify the cholesterol chemical potential in lipid membranes using coarse grained molecular model for both phospholipids and cholesterol. Using the results from the first chapter, we use the method that can best compute free energy profiles to determine transfer free energies of a single cholesterol from lipid membranes with various cholesterol concentrations to bulk water. We further calculate the free energy gained or required to transfer a cholesterol molecule from one lipid type to another without using water as a reference state. The third chapter discusses the interaction of phospholipids with clay montmorillonite at the atomistic level. Based on the density distribution of lipid atoms, we identify different structural aggregates that are formed when certain phospholipids bind to the clay surface.

1.3.1 Membrane Permeation: Mechanism and Best Practices for Simulation Studies

In this part of the work, we start by investigating the mechanism of membrane permeation. We show that membrane permeation is a very difficult process to study using conventional MD. We demonstrate that the permeation by small solutes like amino acids is a rare event and that it involves a very high barrier which can only be captured using enhanced molecular dynamics techniques. Most importantly, we show that even after using enhanced sampling

techniques, quantifying the thermodynamics of membrane permeation is not trivial. We also deduce that translocation simulations can often face hysteresis and that the results can be very dependent on the simulation setup. These kinds of simulations fail to equilibrate within standard simulation times. The equilibration problems are exhibited not only for positively charged permeants but also for permeants with large net dipole moments. Using the replica exchange umbrella sampling (REUS) method, we demonstrate that the difficulty in relaxation of electrostatic interactions between the solute and lipid headgroups gives rise to hysteresis-like behavior, which only decays after extensive equilibration. In contrast, for neutral solutes the simulations do not suffer from hysteresis due to the absence of strong interactions between solute and lipid headgroups. We also demonstrate the usage of committor distribution functions to check if the transition state ensemble (TSE) is accurately sampled in any of our simulations. Furthermore, we show how standard methods to estimate errors in free energy profile can heavily underestimate the associated uncertainty. Based on our results, we develop helpful diagnostic procudures to evaluate the accuracy of calculated free energy profiles of membrane permeation.

1.3.2 Determination of Cholesterol Chemical Potential in Lipid Bilayer

The primary goal of this project is to accurately compute the chemical potential of cholesterol molecules in different lipid environments. It is known that cholesterol prefers saturated phospholipids, but the exact degree of preferential partitioning is difficult to quantify both experimentally and using computer simulations. Using replica exchange umbrella sampling, we show that the free energy difference associated with cholesterol transferring from a binary lipid bilayer to water depends on various factors including lipid head group chemistry, lipid tail saturation and the cholesterol content in the bilayer. The cholesterol chemical potential in a bilayer increases by about 10 kJ/mol when the cholesterol concentration increases from 0% to 60% in the bilayer. The chemical potential, however, does not show much sensitivity to lipid tails and headgroups. Within the error bars, cholesterol has much greater affinity for doubly unsaturated lipid, while it does not show a significant preference between the singly

unsaturated and fully saturated lipids. We also use a new method called binary bilayer system (BBS) to compute the cholesterol chemical potential without having to calculate transfer free energies; this method makes the computation much more efficient. In the BBS, we develop a lipid bilayer with two types of lipids and cholesterol. The BBS behaves like a semi-permeable membrane, where only cholesterol molecules can laterally diffuse. This allows us to compute the chemical potential more easily and accurately. This method also confirms our previous results from replica exchange umbrella sampling simulations that cholesterol only shows sensitivity if the degree of unsaturation between the lipid tails is noticeably different.

1.3.3 Phospholipid Structure and Thermodynamics on a Montmorillonite Clay Surface

In this collaborative project, we investigated soil water repellency (SWR) at the molecular level. SWR is a season-dependent phenomenon and is hypothesized to be caused by reorientation of amphiphilic molecules during drying and wetting cycle.[39] Amphiphilic molecules are suspected to form surface-bound monolayers. In order to test the idea that amphiphilic molecules reorient themselves and expose the hydrophobic tails which decreases the affinity of water to the soil, we simulated systems with clay montmorollinite and different types of lipids. Our simulation results show that this is not the case, and indeed these amphiphilic molecules form micellar or bilayer-like structures in order to decrease the tail exposure to water. In all our simulations with different phospholipids, the aggregates bind to the surface via the lipid headgroup; the surface-bound aggregates are stable and do not undergo any surface reorientation. We have also illustrated that binding becomes more favorable with increasing lipid concentrations.

Chapter 2

MEMBRANE PERMEATION: MECHANISM AND METHODOLOGY

2.1 Introduction

Cell membranes are impermeable to many ions and hydrophilic compounds. The polar headgroups form a different physiochemical environment than the hydrophobic tails that constitute the interior region of the membrane. The intricate interactions between phospholipids, sterols, and other components make membranes selectively permeable. Some molecules are hindered from crossing the membrane more than others; especially small and uncharged solutes exhibit appreciable permeation rates. Understanding how different classes of molecules move across biological membrane is a prerequisite to predicting a solute's permeation rate, which is a critical factor in the fields of drug design and pharmacology.[2]

Using the language of chemical reactions, one can think of membrane translocation as a process that starts with the solute on one side of the membrane (the reactant, or initial, state) and ends with the solute on the other side (the product, or final, state). The path in-between those states is parametrized by a reaction coordinate, the precise nature of which is typically unknown. The rate at which a reaction occurs is often dominated by high free energies of intermediates, which act as a barrier separating the reactant and product states.

Quantifying the magnitude of such free energy barriers is the first step to predicting the permeation rates of small solutes. Membrane permeation is difficult to measure experimentally because of the complexity of lipid bilayer systems.[1] Computer simulations offer a promising route to calculate the free energy profile, also called the potential of mean force (PMF), that governs the translocation process. However, they also face a significant challenge: spontaneous translocation occurs only rarely, especially if the energetic barriers

involved are large. Even though advances in computing hardware and new algorithms allow conventional MD simulations at atomistic resolution to extend to multiple microseconds, this is by far insufficient to directly observe a statistically significant number (if any!) of translocation events. This is a common problem when performing computer simulations, and a large number of methods have been developed over time to address this issue. In one way or another, these methods bias the system of interest to sample configurations unlikely to be seen in an unbiased simulation, and do so in a way that one can obtain the underlying free energy profile of the unbiased system. Examples of such methods are umbrella sampling,[21] metadynamics (and its variations),[22] replica exchange (and its variations),[23] adaptive biasing force,[24] and thermodynamic integration.[25] These methods improve the statistical sampling and reduce the amount of computational resources required to calculate free energy profiles.

The wealth of available methods is marvelous, but it also raises the question which of them one should choose for a given application, such as calculating the free energy profile of membrane translocation. Only a small number of studies have compared the relative performance of these methods by applying them to the same system. For example, Bochicchio and coworkers found that metadynamics and umbrella sampling performed similarly well when computing the PMF for polypropylene and polyethylene oligomers moving across a phospholipid bilayer.[40] In another study, Lee and colleagues computed the free energy profiles of urea, benzoic acid and codeine in their neutral forms to compare umbrella sampling, replica exchange umbrella sampling, adaptive biasing force, and multiple-walker adaptive biasing force, and found no clear advantage for any of these methods.[41] For the charged solute n-propylguanidinium, the side chain analog of arginine, Neale and coworkers found that the virtual replica exchange method converges faster than umbrella sampling.[42] Even fewer studies have focused on explaining why one method out-performs the others and identifying the sources of sampling errors; the difficulty lies in evaluating the convergence of results and estimating the associated uncertainty within these simulation techniques.

Considering the lack of available information but growing interest in quantifying uncertainty in free energy calculations, we here use the well-tempered metadynamics (WT-

metaD),[43] umbrella sampling (US),[21] and replica exchange umbrella sampling (REUS)[44] methods to study the translocation of a variety of small molecules through a symmetric dioleoylphosphatidylcholine (DOPC) bilayer. Particularly, we use these three methods to compute the free energy of water, various forms of alanine, and arginine as they move across a DOPC bilayer. These solute molecules were chosen because they are well studied in the literature [6,45–50] and represent a class of compounds that commonly appear in translocation problems and span a range of sizes, shapes and hydrophobicities .[51] An extensive list of translocation research can be found in recent review papers .[19,20] It should be noted that the translocation mechanism across a pure phospholipid bilayer might be different from that across a biological membrane, where channel proteins can facilitate the process.[52] While most previous works focus on the mechanistic details of the permeation phenomenon itself, we concentrate on identifying and diagnosing generic convergence issues. We find that while all three methods perform similarly well for solutes that are charge-neutral or that have only a small dipole moment, there are significant convergence problems in the case of charged or highly polar solutes, and that those are most easily detected and overcome by REUS simulations.[53,54] These results show that the strengths of electrostatic interactions are likely to affect the relative efficiency among various methods. This by itself gives valuable guidance as to which method one should choose for a given system. It also creates an opportunity to gather mechanistic information about the basic physical properties from the difference in performance of these methods.

In this chapter we will explore these topics in detail. We begin with a brief overview of the WT-metaD, US, and REUS methods of free energy calculations. We will then showcase how these methods can be used to compute the PMF for various solutes. We demonstrate that observables unrelated to the free energy, in particular the so-called committor of a configuration, can be used to assess the quality of the sampling obtained in simulations. We will then conclude with some general advice for performing free energy calculations of membrane translocation.

2.2 Methods

2.2.1 Simulation Details

We performed all simulations using Gromacs 4.6.7[55] with the Plumed 2.1 plugin[56] under periodic boundary conditions. Temperature and pressure were maintained at 320 K and 1 atm using the Nose-Hoover thermostat and Berendsen barostat, respectively. Long range electrostatic interactions were computed using the fourth order PME method[57] with a Fourier spacing of 0.12 nm. The real space coulombic interaction was calculated up to 1.0 nm. Van der Waals interactions were calculated using a cutoff of 1.0 nm. Bond lengths within the solutes and lipids were constrained using the LINCS algorithm.[58]

The DOPC bilayer was constructed using the united-atom berger forcefield[59] such that each monolayer consisted of 64 lipids. Water molecules were treated using the rigid simple point charge (SPC) model.[60] Each bilayer-water system was equilibrated for 10 ns before adding any permeant to the system. After adding the permeant, simulations were run for different amounts of time based on the convergence of free energy profiles. The permeants were modeled using the all-atom OPLS-AA forcefield.[61] Three different forms of alanine were constructed: the first by truncating the side chain at the β-carbon with the α-carbon replaced by a hydrogen; we call this form the side chain analog, where the alanine residue essentially becomes a methane molecule. This method of truncating amino acids has been used in the past to study amino acid interactions with model bilayer systems.[62,63] The second form of alanine was constructed with neutral termini (NH_2 and COOH). We made the third form with charged termini (NH_3^+ and COO^-), a charge neutral but zwitterionic molecule. Cationic arginine with charged termini (NH_3^+ and COO^-) was constructed using the `pdb2gmx` tool of Gromacs. Thus we have studied the following permeants:

i. Water
ii. Side chain analog of alanine
iii. Alanine with neutral termini

iv. Zwitterionic alanine

v. Arginine

For the simulations of positively charged solutes a single chloride ion was added to achieve overall charge neutrality. Each permeant/water system was equilibrated for 10 ns. The equilibrated water/bilayer and water/permeant systems were then combined to form the final water/bilayer/permeant system, which was again equilibrated for 10 ns before any production run under NPT conditions. The Visual Molecular Dynamics (VMD) software was used to monitor and visually inspect all simulation trajectories.[64]

2.2.2 Well-Tempered Metadynamics

Metadynamics is a biasing technique that overcomes sampling problems by adding a history-dependent bias potential to the system potential energy.[65] In regular intervals a Gaussian-shaped potential is added to the system energy, centered at the current value of the collective variable (CV) z. This potential drives z away from its current value, which facilitates the exploration of the entire range of the CV. Over time, these Gaussians add up to the biasing potential

$$V_{\mathrm{G}}(z,t) = \sum_{t'<t} \omega_0 \exp\left(-\frac{(z - z(t'))^2}{2\sigma^2}\right), \tag{2.1}$$

where ω_0 is the height of the Gaussian potential, σ is its width, and $z(t')$ is the value of the collective variable at time t'. Each hill is deposited at a predefined rate and centered at a previously explored configurations, biasing the system towards configurations that have not yet been explored. In the well-tempered variant of metadynamics, the height of the added Gaussian decreases exponentially with the already deposited bias potential at the current value of the collective variable. In other words, the constant ω_0 in (2.1) is replaced with

$$\omega = \omega_0 \exp\left(-\frac{V_{\mathrm{G}}(z(t'),t')}{k_{\mathrm{B}}T\Delta T}\right), \tag{2.2}$$

where ΔT is an input parameter that effectively increases the temperature at which the CV is sampled.[43] This simulation scheme has been shown to yield the sought-after free energy profile in the limit of long simulation times through[66]

$$F(z) = -\frac{T + \Delta T}{\Delta T} \lim_{t \to \infty} V_{\mathrm{G}}(z, t) \tag{2.3}$$

up to an irrelevant additive constant.

In practice, the choice of metadynamics parameters affects the rate of convergence,[40] and typically requires some experimentation as there is no single, universally accepted set of parameters even for the specific problem of membrane translocation. The parameters we use for this work can be found in Table 2.1. Our chosen values for the initial height of the added Gaussians is larger than what has been reported in translocation simulations of neutral solutes. We find that increasing the hill height is necessary to observe multiple translocation events over the course of our simulations, especially for charged and dipolar solutes. For converged free energy profiles, uncertainties were estimated using the method proposed by Tiwary and coworkers.[67]

Solute	Height(kJ/mol)	Width(nm)	Time of gaussian addition(ps)	Biasfactor $(\frac{T+\Delta T}{T})$
Water	2.2	0.1	3	10.0
Side chain analog of alanine	0.1	0.1	5	2.00
Alanine with neutral termini	10	0.3	3	35.0
Zwitterionic alanine	15	0.3	3	140
Zwitterionic arginine	15	0.3	3	140

Table 2.1: Metadynamics parameters for all solutes studied.

We choose the normal component of the distance vector between the center of mass of the solute and the center of mass of the lipid bilayer as the collective variable z. We used snapshots from a WT-metaD trajectory, with permeants at various positions relative to

the bilayer, as the initial configurations for both US and REUS calculations; the snapshots were taken when the solute first reached the desired distance after a minimum of 100 ns of WT-metaD simulation time.

2.2.3 Umbrella Sampling

Umbrella sampling (US) adds a biasing potential to the system's Hamiltonian to enhance the sampling of configurations that are high in energy .[21] In this case the biasing potential is static. We choose a sequence of N "windows" that span the range of interest of the collective variable z. In the i^{th} window the system is biased to remain close to a predetermined value z_i by using a harmonic umbrella potential

$$V_i(z) = \frac{1}{2}k(z - z_i)^2, \tag{2.4}$$

where k is the stiffness of the potential.

For each of those windows we perform a standard MD simulation, from which we obtain the probability distribution $P_i^{\mathrm{b}}(z)$ of the collective variable in the biased system. We can recover the distribution $P_i(z)$ of the unbiased system over the range of values observed in the i^{th} window via

$$P_i(z) = P_i^{\mathrm{b}}(z)\, \mathrm{e}^{\beta V_i(z)}\, \mathrm{e}^{-\beta C_i}, \tag{2.5}$$

where C_i is an unknown constant to be determined shortly. Having obtained such estimates for the probability distribution for each window, we still need to combine this information to obtain the free energy profile $F(z)$ over the entire range of z. One efficient way to do so is the Weighted Histogram Analysis Method (WHAM),[68,69] which expresses the probability distribution $P(z)$ as a linear superposition of the estimates from each window,

$$P(z) = \sum_{i=1}^{N} w_i(z) P_i(z). \tag{2.6}$$

The weights $w_i(z)$ are chosen to minimize the statistical error of the reconstructed distribution. It can be shown that these weights then satisfy the coupled equations

$$w_i(z) = \frac{S_i \mathrm{e}^{-\beta V_i(z) + \beta C_i}}{\sum_{j=1}^{N} S_j \mathrm{e}^{-\beta V_j(z) + \beta C_j}} \tag{2.7}$$

$$\mathrm{e}^{-\beta C_i} = \int \mathrm{d}z \, \mathrm{e}^{-\beta V_i(z)} P(z) \tag{2.8}$$

which must be solved self-consistently for the unknown C_i. Here, S_i is the number of independent samples obtained from the i^{th} simulation. Once the unbiased distribution (2.6) is obtained, we can compute the free energy profile

$$F(z) = -k_{\mathrm{B}} T \, \ln P(z). \tag{2.9}$$

We use the same reaction coordinate z for the US calculation as for WT-metaD. For each solute we used a series of windows with spacing of 0.1 nm spanning the entire bilayer. A harmonic potential of $k = 1000$ kJ/mol/nm^2 was used. 50 windows were constructed for all three forms of alanine. For the side chain analog and alanine with neutral termini, each window was simulated for 20 ns totaling 1 μs of simulation time. Zwitterionic alanine was simulated for 2.75 μs where each window was run for 55 ns. 40 windows were used for arginine simulation and each window was run for 100 ns, resulting in 4.0 μs of total simulation time. Table 2.2 contains more information about solute specific simulation time.

2.2.4 Replica Exchange Umbrella Sampling

Replica exchange umbrella sampling (REUS) is very similar to umbrella sampling: again we perform N different simulations, also called replicas, each with a different bias potential such as the one shown in (2.4). In REUS, however, these simulations are not independent: at pre-determined time intervals, two replicas exchange their current configuration with a

probability that is determined by the detailed balance condition. The latter guarantees that each replica still samples configurations according to the equilibrium distribution of the biased system. The reconstruction of the free energy profile proceeds in the same way as in US, for example by using the WHAM algorithm.

REUS is different from regular replica exchange simulations, in which the replicas are unbiased in z but are instead held at different temperatures. In that case, simulations at the temperature of interest benefit from the enhanced sampling at higher temperatures.[70] In REUS, trajectories in the i^{th} replica are bound to only explore states close to z_i, and the exchange with other replicas allows the system to overcome free energy barriers in orthogonal directions if those barriers are smaller in other regions of the collective variable. For example, a solute at the center of a bilayer might have very limited rotational freedom. In a regular US simulation, this solute might be trapped there in a specific orientation for a long time. In REUS, on the other hand, the solute is essentially allowed to diffuse to the bilayer surface, rotate, and return to the bilayer center.

To allow for accurate comparisons between the simulation methods we use the exact same parameters for US and REUS. For water, each replica was simulated for 8 ns and 27 ns, respectively. Each replica was run for 55 ns for alanine and 165 ns for arginine. An exchange was attempted every 2 ps for all REUS calculations. Table 2.2 contains information about simulation lengths for each solute using each of the three methods.

Solute	**US**		**REUS**		**WT-metad** (μs)
	length per window (ns)	total (μs)	length/replica (ns)	total * (μs)	
Water	-	-	8	0.32	0.42
Side chain analog of alanine	20	1.00	-	-	1.00
Alanine with neutral termini	20	1.00	-	-	1.00
Zwitterionic alanine	55	2.75	55	2.20	1.40
Zwitterionic arginine	100	4.00	166	6.64	1.00

Table 2.2: Summary of all simulated systems. REUS total simulation time only represent simulations that span the membrane between -2.0 nm$< z \leq 2.0$ nm. We ran additional simulation around $z \leq -2.0$ nm and $z > 2.0$ nm to investigate if solute configuration changes in bulk water phase.

2.2.5 Error Estimation

Every free energy profile obtained from computer simulations should be accompanied by an estimate of the statistical uncertainty. Because all of the above methods are guaranteed to converge to the correct result in the limit of infinite simulation times, a simple and practical way to establish convergence is to show that the estimate of $F(z)$ becomes independent of simulation time. In the case of metadynamics, it is furthermore assumed that the observed dynamics of the collective variable will become diffusive when the added bias potential completely compensates any variation in the system's free energy profile, which can serve as an additional test for convergence.

Quantifying the uncertainty is more difficult. A straight-forward but computationally expensive way to do so is to repeat the entire calculation multiple times, and then compute the uncertainty in $F(z)$ using the standard tools of error analysis. A more convenient and potentially more efficient way to obtain error bars is to divide the data of a single simulation into smaller segments in such a way that they can be considered uncorrelated, and to compute uncertainties based on these segments. Bootstrapping is a method that samples many such partitionings, and can be readily applied to US and REUS simulations. One freely available implementation of this algorithm is the tool `g_wham`, which is part of the Gromacs software package which provides a way to calculate free energy profiles and their uncertainties by

combining the WHAM method with bootstrapping analysis.[71] Therefore, we use `g_wham` to compute the free energy as well as to calculate the uncertainty.

The ever-increasing nature of the bias potential $V_{\mathrm{G}}(z)$ complicates this analysis in the case of metadynamics simulations. However, one can replace (2.3) with a time-independent estimator of the free energy profile $F(z)$, and compute its uncertainties from the variance of such estimates at different simulation times.[67] As long as there are no systematic sampling errors, this methods can estimate the uncertainities for the PMF. For converged free energy profiles, uncertainties were estimated using the method proposed by Tiwary and coworkers. This method could also be applied to asymmetric free energy profiles, but because those simulations are evidently not yet converged the estimated uncertainties would be unreliable.

2.2.6 The Committor and Its Distribution

The free energy profile $F(z)$ is a thermodynamic description of a process that proceeds along the collective variable z. It is the starting point for classical theories of chemical kinetics, which typically consider reactions in which $F(z)$ has two minima that correspond to the reactant state A and the product state B, and that are separated by a peak. This peak forms a barrier whose crossing determines the reaction rate. The location of the peak is sometimes thought of as the transition state, which separates the reactant and the product states.

This picture is accurate only if z is indeed the reaction coordinate, i.e., it describes accurately and completely the progress of the process under consideration. Whether that is the case is usually not clear *a priori* in complex systems such as biological membranes. Kinetic theories of chemical dynamics have developed quantitative metrics to describe the progress of a reaction.[72,73] Rather than relying on a projection of the free energy onto an arbitrary CV, they consider the ensemble of paths that go from A to B. Where a single state lies along a reaction is described by the committor p_{B}, defined as the probability that a trajectory initiated from that state will visit the product state B before the reactant state A. By definition, a state that has $p_{\mathrm{B}} = 1/2$ is a transition state. The set of all transition

states is called the transition state ensemble.

For the translocation process of hydrophilic solutes across a phospholipid bilayer, the stable states A and B correspond to solute positions above and below the membrane, respectively. If the bilayer is symmetric, then one might expect that a solute that lies at the center of the membrane has equal probability of moving to the upper or the lower membrane/water interface when let go. In other words, at least some configurations with $z = 0$ should have a committor value of 1/2.

All three biasing methods discussed above generate configurations with $z = 0$, and each of those has an associated committor value p_{B}. We will see in the next section that the probability distribution $P(p_{\mathrm{B}})$ of these committor values contains valuable information. If the set of configurations with $z = 0$ were the transition state ensemble, then the distribution $P(p_{\mathrm{B}})$ would be non-zero only at $p_{\mathrm{B}} = 1/2$. In general this set will also contain configurations that are not transition states, but at least for a symmetric bilayer the distribution $P(p_{\mathrm{B}})$ must be symmetric to reflect the equivalence of the upper and the lower side of the membrane. In either case one expects a significant peak at $p_{\mathrm{B}} = 1/2$.

We estimate the committor distribution for the ensembles of configurations with $z = 0$ generated by WT-metaD, US, and REUS simulations. For alanine analog and arginine, we select 6 such configurations from each method, and use them together with randomized initial velocities as starting points for four unbiased MD simulations. For zwitterionic alanine, from WT-metaD, we select 3 and from REUS and US, we select 6 such configurations. We then record in how many of these four trajectories the solute reaches the lower ($z < 0$) membrane/water interface before the upper interface. The resulting fraction serves as an estimate for the p_{B}-value of the starting configuration. We then calculate a histogram of these committors to obtain an estimate for the distribution function $P(p_{\mathrm{B}})$ for each sampling method.

It should be noted that both the number of configurations tested and the number of trajectories used to compute a configuration's committor value should be much larger if one wants to calculate an accurate estimate of the committor distribution function.[73] The

significant computational cost of bilayer simulations and the long time scale of solute motion across the membrane limits us to relatively small numbers. However, we are interested only in large, qualitative differences between the distribution functions obtained from the three different methods, for which our approach is sufficient.

2.3 Results

In this section, we summarize our results for the calculation of the translocation of different solutes through a DOPC bilayer. We begin with the transport of a single water molecule; then we consider three different forms of alanine. Finally, we study arginine as an example of a cationic amino acid. For all solutes, we calculate the free energy profiles using WT-metaD and US first. If the result from these methods are similar, we do not do further analysis with REUS. REUS is a modified version of US and is guaranteed to converge if US converges by the nature of the algorithm. However, if the results from WT-metaD and US differ, we use REUS to compute free energy profiles and to diagnose sampling issues.

2.3.1 Water

The movement of water molecules across the cell membrane is of paramount biological importance, because it provides a mechanism that allows the cell to control the concentrations of electrolytes and to maintain an osmotic balance. Water pores and water channels are often studied in the context of peptides and drug translocation as well. While some of this movement is facilitated by membrane proteins, we here focus only on the passive diffusion of individual water molecules through the lipid bilayer.

Following the simulation methodology outlined in the previous section, we calculated the translocation free energy profile of a water molecule (dipole moment 2.27 D as modelled by SPC) through a DOPC bilayer using both the WT-metaD and the REUS methods. Here the water molecule plays the role of a solute. Results are shown in Figure 2.1, which shows that both methods give the same result which also matches the US result available in the literature: as the water molecule enters the bilayer, the free energy increases rapidly at

first, and then plateaus once the molecule reaches the central core of the membrane .[49] The height of this plateau is approximately 27 kJ/mol relative to bilayer/water interface, which is consistent with previous results.[6,74] The flattening of the PMF in the central part of the bilayer coincides with a region of lower lipid tail density and increased disorder.[74,75]

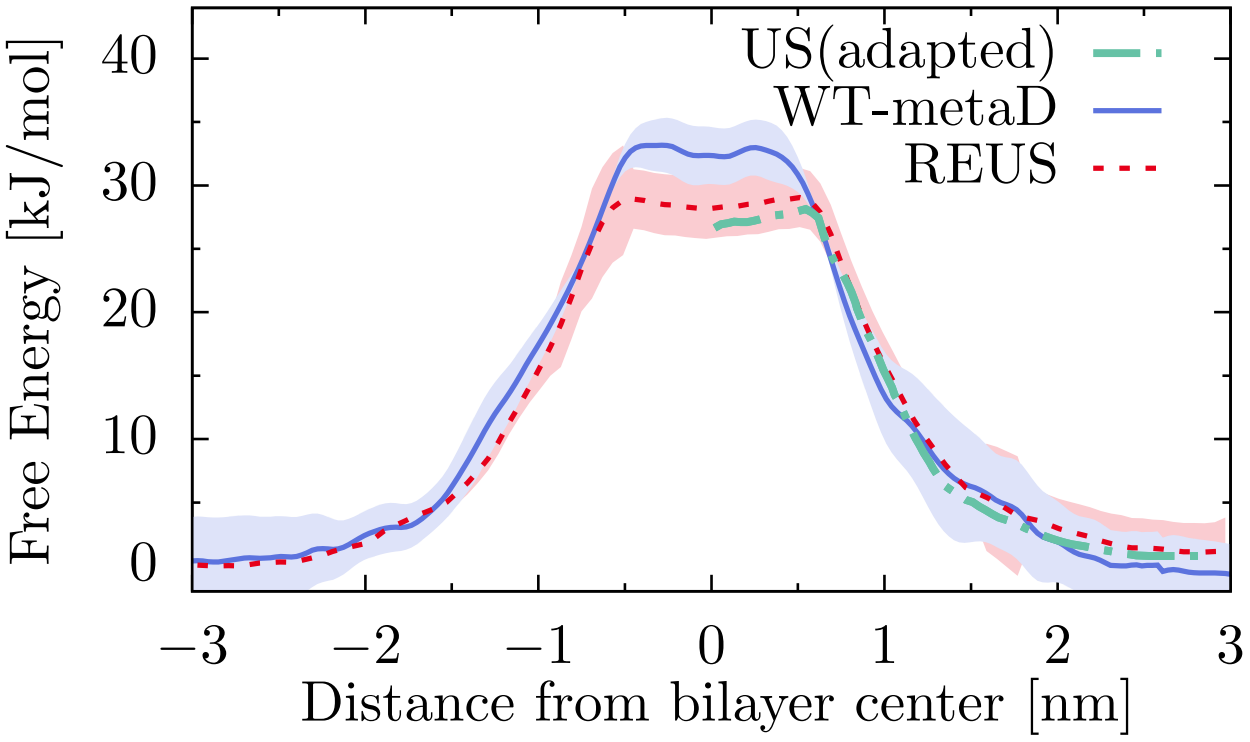


Figure 2.1: Free energy of a single water molecule as a function of distance from the bilayer center using WT-metaD and REUS with shaded area representing uncertainties. The two results are comparable and are similar to previous simulation results, in particular both free energy profiles are symmetric with respect to the bilayer center.

Note that these free energy profiles were calculated over the entire range of the solute position z; the symmetry of the PMF emerged naturally and was not put in manually, for example by averaging over positive and negative z. The fact that the calculated $F(z)$ has

such high symmetry is by itself a good indicator for the convergence of the simulations. These results show that both methods yield the same result with similar uncertainties in a comparable amount of computing times. As such, neither method has a clear advantage over the other.

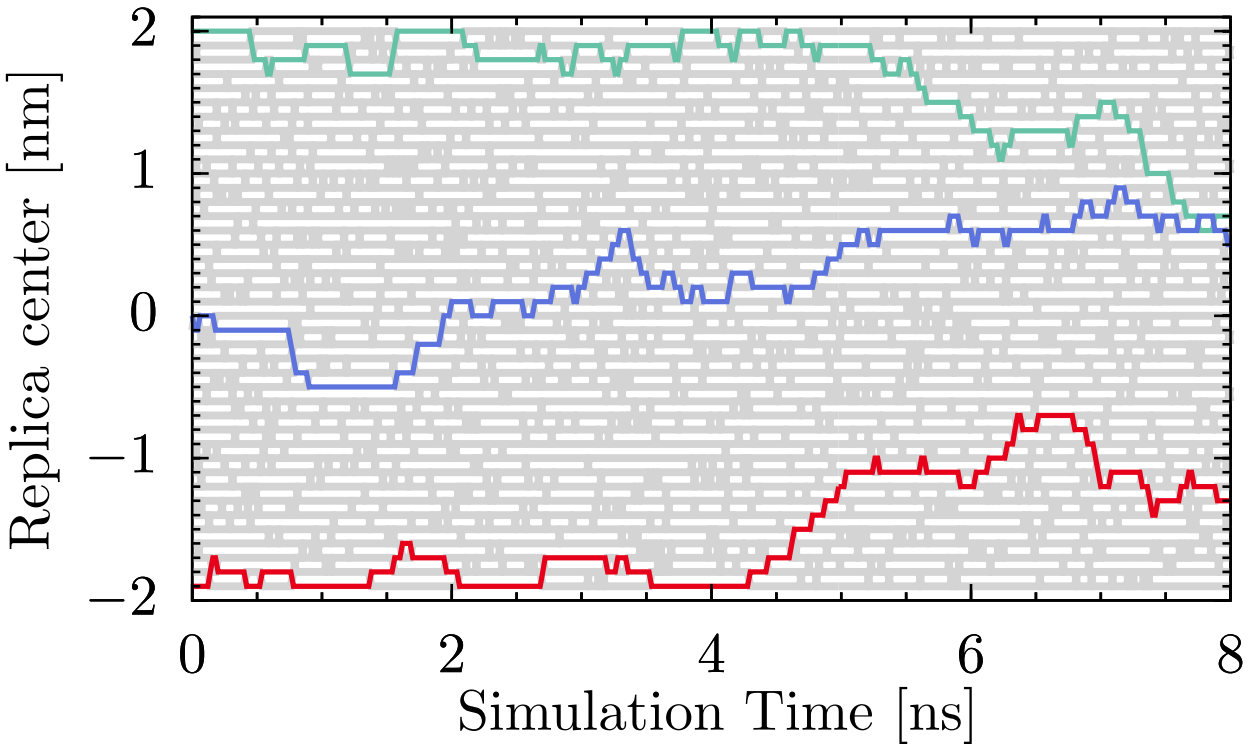


Figure 2.2: REUS results for water. Exchange pattern lines for 40 replicas. Replicas starting at $z = 2.0$ nm, 0 nm and -1.9 nm highlighted. All replica pattern lines are freely diffusing indicating even exchanges between all 40 replicas.

REUS provides additional useful information that is not available from WT-metaD or US. In Fig. 2.2, we look at the pattern of configuration exchanges between neighboring replicas as a function of simulation time. In principle, each trajectory should diffuse through the complete space of replicas due to the ongoing exchange. This is demonstrated in figure 2.2,

which shows the exchange pattern of forty trajectories diffusing through the space of replicas, which span the bilayer from −1.9 nm to 2 nm. We see that exchange occurs frequently between most replicas which further reinforces sufficient sampling. Failure to do so is an important indicator of convergence problems as we will show later in the chapter.

2.3.2 Alanine

The insertion and translocation free energy profiles of peptides are of particular interest for multiple reasons. First, they determine the affinity of proteins to the bilayer, for example through the energetic cost (or gain) of inserting a transmembrane helix. Second, there has been an increased interest in peptide therapeutics, i.e., drugs in the form of peptides that need to cross membranes to become active.[76] Third, so-called cell-penetrating peptides have been found to significantly increase the uptake of attached drug molecules.[77] In all these applications, it is desirable to have quantitative information about the potential of mean force of translocation on a per-residue basis. Here, as an example, we generate three chemical forms of an alanine: side-chain analog, neutral alanine and zwitterionic alanine and computer their PMFs. We will see that these forms of alanine behave very differently, and the cause of these differences lie in the difference of the molecular dipole moments.

Side-Chain Analog of Alanine

One way to model a single residue is by considering only its side chain, and replacing the α-carbon with a single hydrogen atom.[63] In the case of alanine, this so-called side chain analog is simply a methane molecule. Unlike water, methane is hydrophobic, and one would therefore expect a qualitatively different free energy profile. This is indeed the case, as illustrated in Figure 2.3(a): the free energy is minimal at the center of the bilayer, where it is approximately 10 kJ/mol lower than in the surrounding water.

The PMFs shown in this figure were obtained using the well-tempered metadynamics and the umbrella sampling methods. Both methods yield symmetric free energy profiles,

which are the same within the statistical uncertainty. Again, it seems that both methods are equally capable of accelerating the PMF calculation.

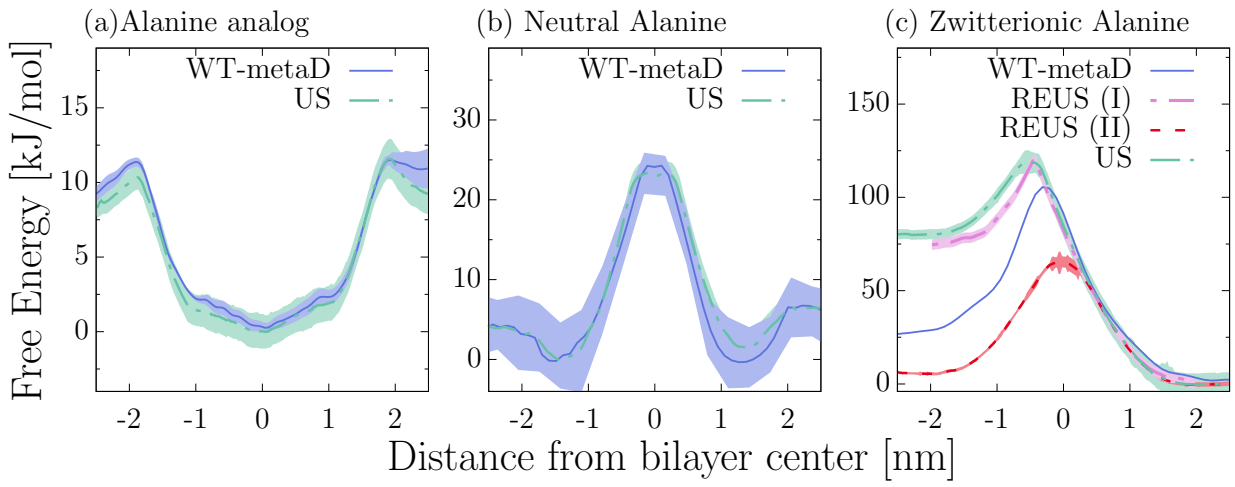


Figure 2.3: Free energy of three forms of alanine translocation as a function of distance from the bilayer center with shaded area representing uncertainties. WT-metaD and US profiles are comparable for side chain analog of alanine (a) and alanine with neutral termini (b). For zwitterionic alanine (c), the four lines differ: profile generated using WT-metad, US, and REUS(I) are asymmetric and the maxima does not lie at $z = 0$ nm. REUS(II) profile after extensive equilibration is symmetric with maximum at $z = 0$ nm.

Neutral Alanine

Considering only the amino acid side chains is insufficient to predict the insertion free energies of membrane proteins, which also depends on the protein backbone.[78] A single amino acid in its un-ionized form contains both a carboxyl (-COOH) and an amino (-NH_2) group. In this form, alanine is charge neutral and has a weak dipole moment.

Figure 2.3(b) shows the free energy profiles of alanine obtained from WT-metaD and US simulations. Both methods yield symmetric profiles that exhibit a substantial maximum in the bilayer core, reflecting the hydrophobic character of this solute. There are shallow minima

in the free energy at the lipid/water interface, which indicate that alanine is weakly attracted to the bilayer surface. As before, both methods perform similarly well when computing such profiles.

In practice one would typically be satisfied when a PMF calculation converged to a point of acceptably small uncertainties and, in case of a symmetric bilayer, yields a symmetric free energy profile. However, to highlight the differences with other, more polar or charged solutes, we showcase how committor analysis can be used to gain further confidence in the correctness of a PMF calculation.

As described in Section 2.2.6, the committor p_{B} of a configuration is the probability that the solute will reach the lower membrane/water interface before the corresponding upper interface in a trajectory that is started from this configuration with random initial velocities. If the bilayer is symmetric, then one would expect that a large number of configurations in which the solute is at the center of the bilayer (i.e., $z = 0$) has a committor value of 1/2.

Both metadynamics and umbrella sampling calculations generate sample configurations with $z = 0$; the former by continually driving the solute across the bilayer, and the latter in a window in which the solute is constrained to remain close to the bilayer center. We can test whether those configurations have the expected committor value by using them as starting points for multiple short trajectories, and by monitoring to which side of the bilayer the solute moves in these trajectories. This calculation is illustrated in Figure 2.4: from the underlying simulation (grey lines in both panel), which is either performed using WT-metaD (a) or US (b), we choose two configurations with $z = 0$, and initiate four unbiased trajectories from each. In all cases, we find that the solute moves towards the upper interface in two out of the four trajectories, and towards the lower interface in the remaining trajectories. Our best estimate for the committor of all four tested configurations is therefore indeed $p_{\mathrm{B}} = 1/2$. It should be noted that this perfect agreement is fortuitous: while the committor of a configuration is a well-defined property, its estimate as obtained from unbiased trajectories follows the binomial distribution. One should not expect an exactly equal number of trajectories to reach the two membrane/water interfaces, and neither should a slight imbalance be consid-

ered proof that a configuration is not a transition state. If in doubt, one can increase the number of trajectories to obtain a more accurate estimate of the committor value.

By extending this process to more than the two configurations shown in Figure 2.4, we can calculate the distribution $P(p_{\mathrm{B}})$ of committor values of the $z = 0$ configurations generated by the two sampling methods. The result for alanine is shown in Figure 2.5, where we used six configurations as the starting points for four trial trajectories each. We see that both WT-metaD and US generate ensembles of states that have a committor distribution sharply peaked at $p_{\mathrm{B}} = 1/2$. This is consistent with the physical intuition that a solute at the center of a symmetric bilayer has no preference for either side of the membrane.

Zwitterionic Alanine

Under physiologically relevant conditions, the carboxyl group of alanine is deprotonated while the amino group is protonated, which yields the zwitterionic form of alanine. While still charge neutral, this form has a significant dipole moment.

This seemingly minor change in solute electrostatics has a dramatic effect on the calculated free energy profiles as shown in Figure 2.3(c) : the PMFs calculated using well-tempered metadynamics, umbrella sampling, and in an initial replica exchange umbrella sampling simulation (denoted REUS (I)) all appear to be different, which is a clear warning sign that these simulations have not converged. Equally disturbing, none of the obtained PMFs are symmetric with respect to the center of the membrane, which we know they should be for a symmetric bilayer. The error bars in the PMFs, calculated using standard error analysis methods, do not reveal any issues with the free energy profile. The error bars are too small, so we cannot trust them to give an accurate measure of the uncertainties.

How then are we to judge the results of a free energy calculation? In practice it is often unfeasible to perform the same calculation using multiple different methods, as we have done

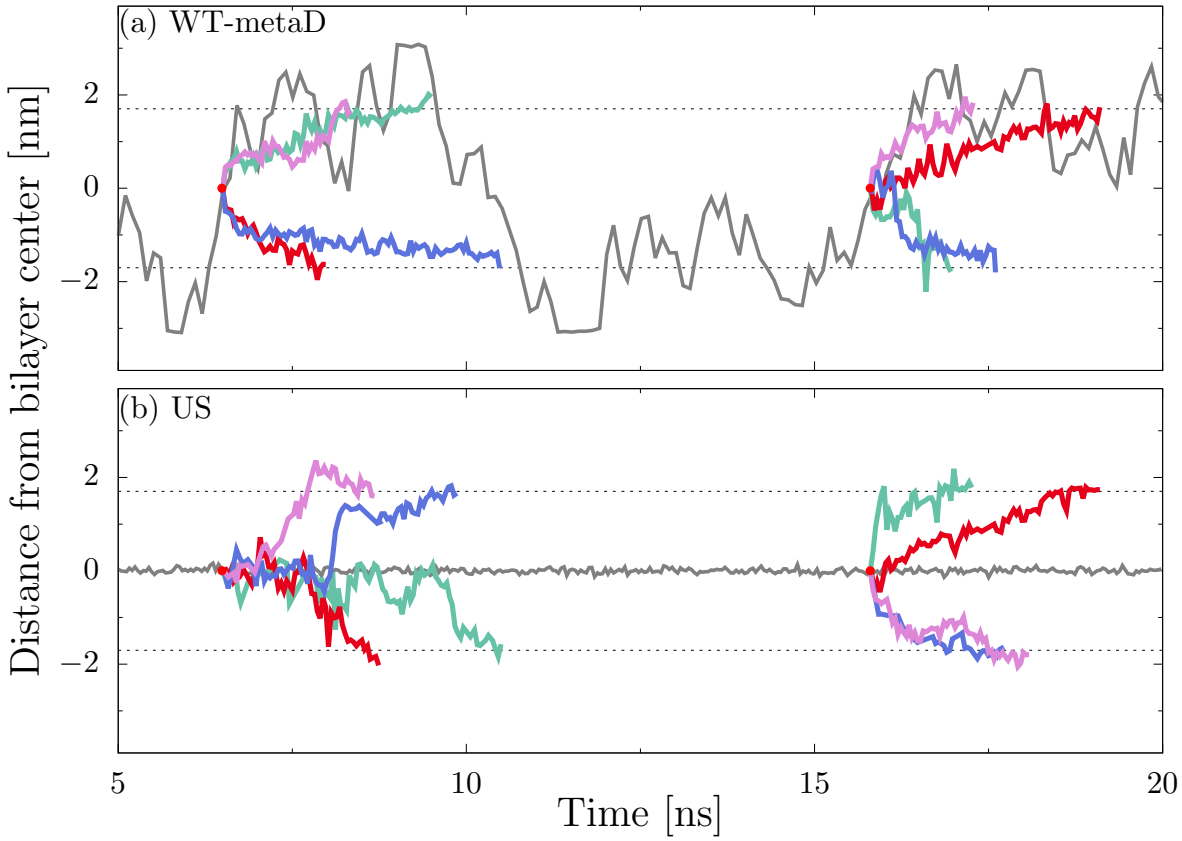


Figure 2.4: Both WT-metaD (a) and US (b) methods sample the transition state of translocation in simulations of neutral alanine. Starting with $z = 0$ configurations generated from a metadynamics trajectory (a) or from the window anchored at the membrane center (b), we initiate four short, unbiased MD simulations. For each of the tested configurations, half of those trajectories evolve towards one side of the bilayer, and the other half to the opposite side. We conclude that these configurations have a committor value $p_B = 1/2$.

here. A cheaper and simpler test is to use the prior knowledge that for a symmetric bilayer, the PMF must also be symmetric. This provides a valuable test for convergence, and is an advantage that should not be given up lightly. For example, one might argue that it should be sufficient to compute the free energy profile only for $z \geq 0$, and fill in the remainder by simply taking the mirror image with respect to the $z = 0$ axis. This would indeed work if the calculation was perfectly converged. However, testing this convergence is most easily done

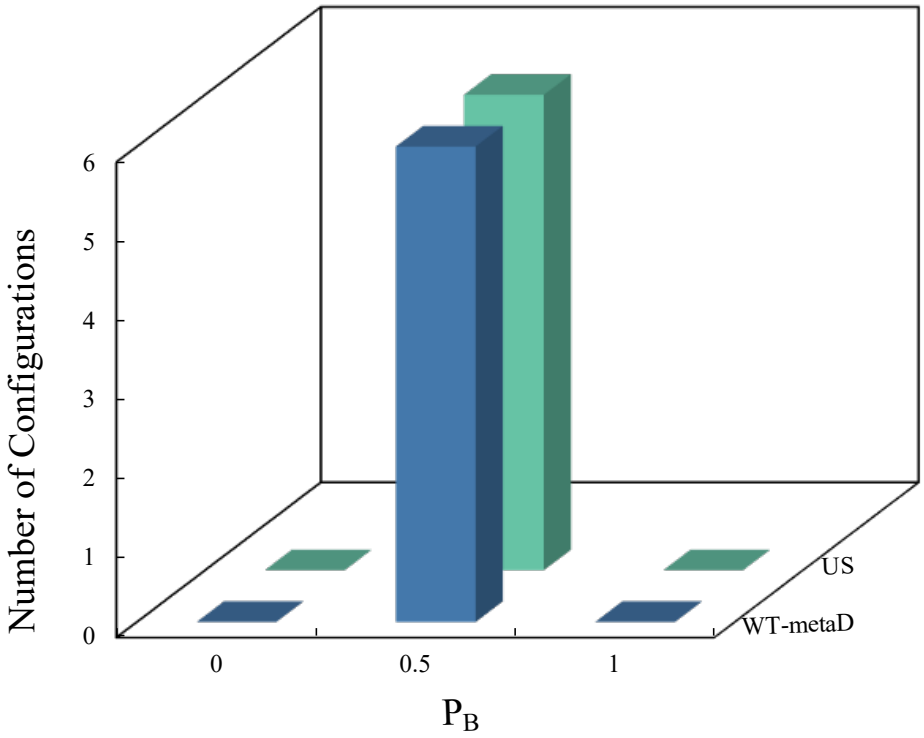


Figure 2.5: Committor distribution for neutral alanine. The committor calculation was performed for 6 WT-metaD and 6 US configurations at $z = 0$. The single peak at $p_B = 1/2$ shows that alanine has no bias towards either side of the membrane in all tested configurations.

by calculating the free energy over the entire range of z. The same warning applies when one calculates the entire PMF, which is then manually symmetrized – the convergence should be checked before further processing the free energy profile.

A first glimpse as to why all three simulation methods might fail to converge over typical simulation time scales can be obtained by visually inspecting the generated trajectories: unlike in the case of un-ionized alanine, the zwitterionic form drags with it both lipids and water molecules as it is driven across the membrane by the biasing algorithm. This is caused

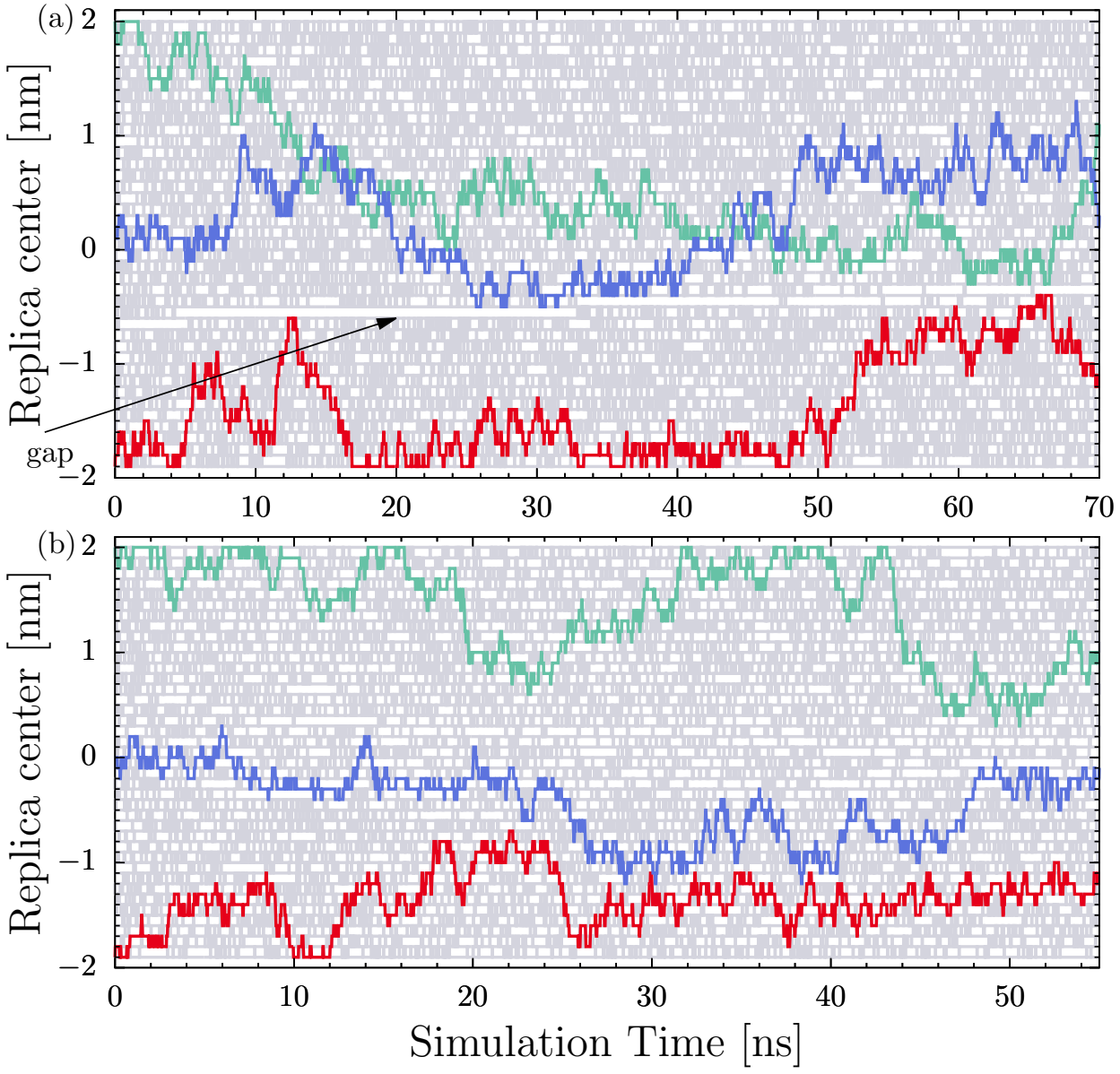


Figure 2.6: REUS results for zwitterionic form of alanine. (a) Exchange pattern lines for 40 replicas. Replicas starting at $z = 2.0$ nm, 0 nm and -1.9 nm highlighted. A gap in the exchange pattern (indicated by the arrow) is clearly visible for REUS(I) in panel (a). Using the final configuration of $z = 0$ replica as the initial condition for the center replicas in a new REUS(II) simulation (b), this gap is no longer present, indicating even exchanges between all 40 replicas.

by strong electrostatic interactions between the solute and the zwitterionic head groups of the phospholipids as well as the polar water molecules. We see that these collective motion creates long-lasting defects in the membrane, and their relaxation exceeds the simulation time scale. In Figure 2.6, we show the exchange pattern of forty trajectories diffusing through the space of replicas, which span the bilayer from -1.9 nm to 2 nm. While exchange occurs frequently between most replicas, there is a suspicious gap, visible as a white line (panel(a)), that indicates that the replicas on either side only rarely switch configurations. Evidently there is a large difference between the configurations in those replicas that significantly decreases the acceptance probability of exchange attempts. The nature of this difference becomes apparent when visualizing the trajectories of replicas on either side of the gap: replicas above the gap contain configurations in which the alanine is in close contact with head groups from the upper leaflet's phospholipids, as illustrated in the upper and left panel of Figure 2.7. Below the gap, on the other hand, are configurations in which the alanine is either surrounded by head groups from either the lower leaflet's lipids (bottom and right panel) or forms a defect in the membrane that affects the upper and the lower leaflet similarly (center panel). This difference is imposed by the initial conditions used in the various replicas.

The effect of this gap on the replica exchange dynamics can be seen in the three highlighted trajectories shown in Figure 2.6(a): tracking a simulation across exchanges, we see a movement that resembles that of a random walk, but that cannot typically cross the gap – it acts almost like a hard wall. The gap thereby separates the replica space into two distinct regions that only rarely exchange configurations with each other. The gap is not, however, entirely static: every so often it seems to move upward, which indicates that a replica from above the gap with a one-sided bilayer deformation has converted to a more symmetric conformation below the gap. This process is extremely slow: over the 70 ns simulation time shown in Figure 2.6(a), the gap has shifted by only three replicas towards the center of the bilayer. Nevertheless, this motion reflects the unswerving progress towards equilibration of the system.

Having realized that it is the slow conversion from one-sided to symmetric defect configu-

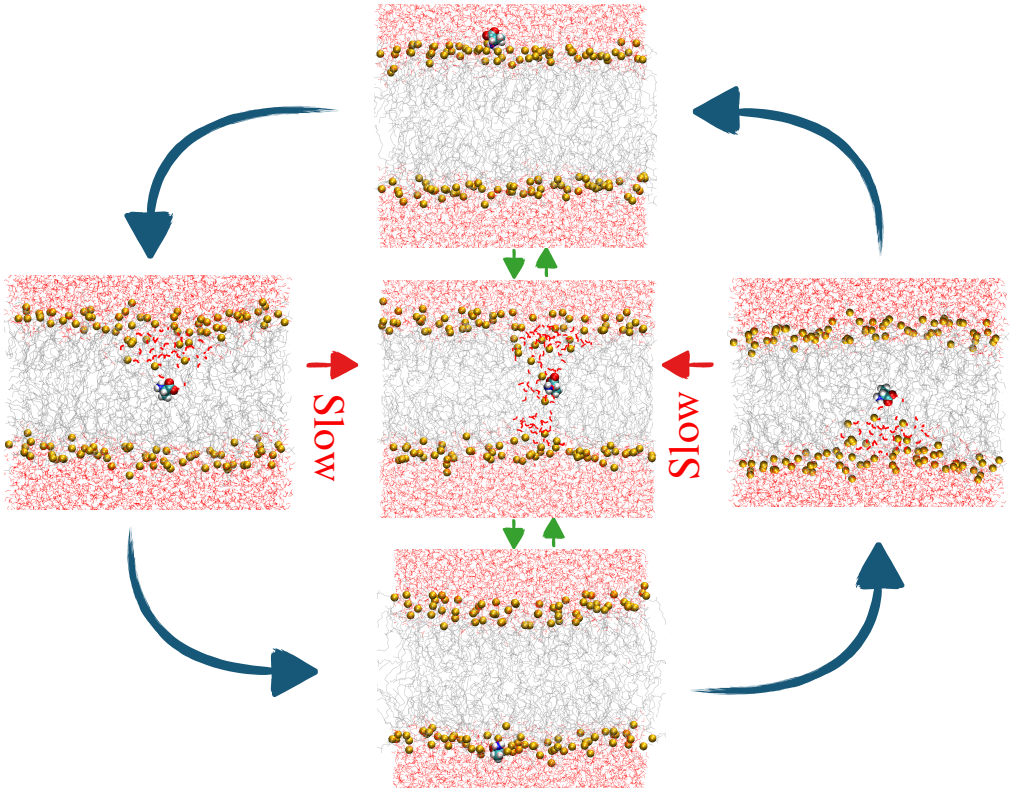


Figure 2.7: Typical configurations encountered in simulations of the translocation of strongly polar and charged solutes. Such solutes tend to localize at one of the membrane/water interfaces (top and bottom panels). When driven inside the bilayer, they create a one-sided deformation of the bilayer (left and right). The transition state, however, is symmetric with respect to the two bilayer leaflets (center). The transition to this symmetric defect is very slow, because it requires crossing of a free energy barrier in a direction other than z, and that is therefore not accelerated by the biasing method.

rations that impedes the convergence of the REUS calculation, it seems possible to side-step this process by using a configuration in which the solute is in close contact with lipids from both leaflets as an initial condition. We therefore performed a new calculation, referred to as REUS (II), that uses the configuration shown in the central panel of Figure 2.7 as the starting point for replicas close to the bilayer center. This simulation's exchange pattern, shown in Figure 2.6(b), no longer shows a long-lasting gap; all replicas appear to exchange equally with their neighbors in a random fashion. Not only does the exchange pattern now

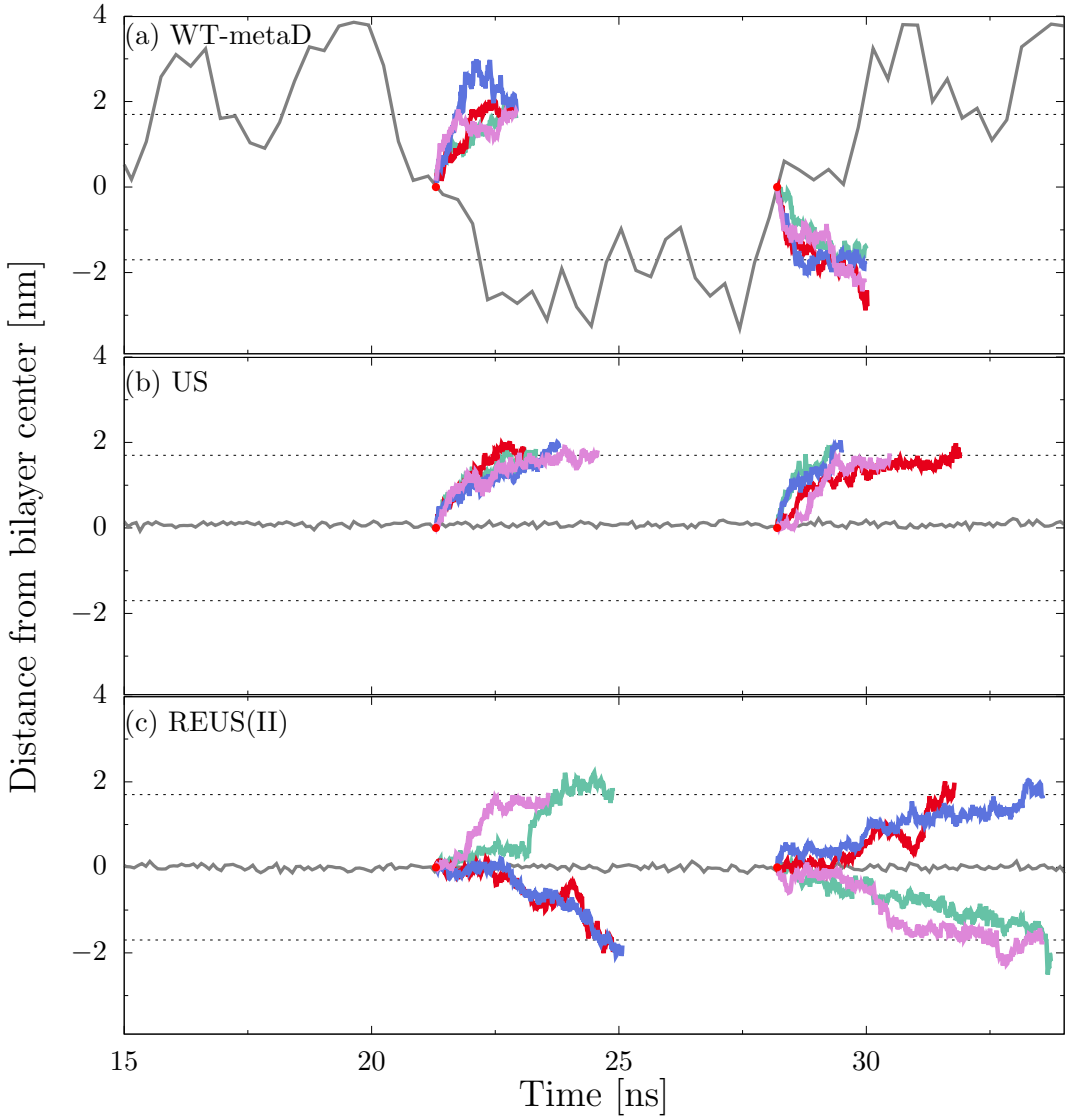


Figure 2.8: For zwitterionic alanine, WT-metaD (a) and US (b) generate configurations with $z = 0$ that are not transition states. Such configurations have a strong tendency to return to the side of the bilayer from which they just came (a) or to the side from which the initial condition in that window originated (b). The second set of REUS simulations (c), on the other hand, generates configurations that show no bias toward either side of the membrane, and that are therefore transition states.

behaves as one would expect for an equilibrated simulation like in Fig. 2.2, the resulting free energy profile at last satisfies the required condition of symmetry: as shown in Figure 2.3,

the PMF obtained from the REUS (II) simulation is an even function of z, and has a smooth maximum at the bilayer center that corresponds to a free energy barrier of 60 kJ/mol.

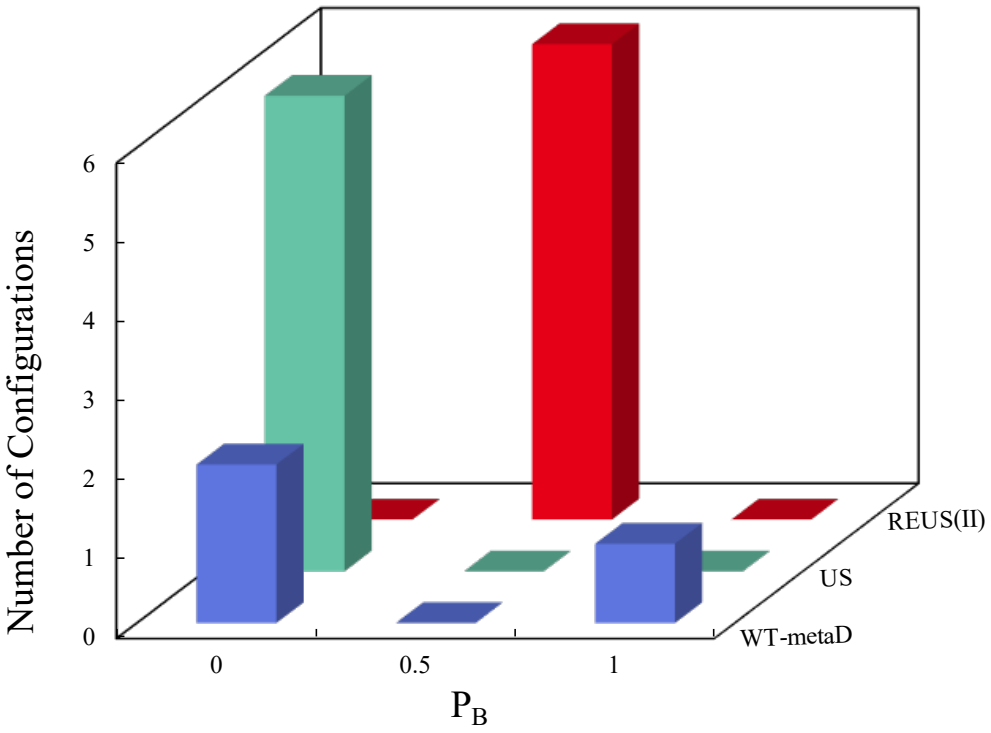


Figure 2.9: Committor distribution for zwitterionic alanine. The committor calculation was performed for 3 WT-metaD, 6 US, and 6 REUS configurations at $z = 0$. For every tested z = 0 configuration obtained from WT-metaD and US simulations, all unbiased trajectories that start from this configuration evolve towards the same side of the membrane. The single peak at $p_B = 1/2$ shows that alanine has no bias towards either side of the membrane in all tested configurations.

So far we have demonstrated how one can use the exchange pattern of REUS simulations

to detect possible convergence problems. This method cannot be used for WT-metaD or US calculations, which is a considerable advantage for the replica exchange method. A generally applicable method to test proper equilibration of solute configurations near the bilayer center can be developed using the committor p_{B}. In Figure 2.4, we showed that for un-ionized alanine, unbiased simulations initiated from $z = 0$ configurations obtained in WT-metaD and US simulations had equal chance of progressing toward the upper and the lower membrane/water interface. Figure 2.8 shows that this is not the case for zwitterionic alanine: for the two shown starting configurations obtained from a metadynamics simulation, alanine return to the side of the membrane from which it just came in four out of four tested trajectories. Similarly, in a US simulation with a harmonic bias potential at the center of the bilayer, and that was started with an initial configuration in which the solute was in close contact with lipids from the upper leaflet, all tested trajectories return to the upper membrane/water interface. Clearly, in both cases the $z = 0$ configurations retain some memory of how the alanine got to the center of the bilayer. This is also the case for the REUS (I) simulation. The better equilibrated REUS (II) simulation, on the other hand, yields configurations in which the solute is near the bilayer center, and that have equal probability of evolving towards the upper and the lower side of the membrane in unbiased simulations, as shown in Figure 2.8(c).

We can estimate the committor distribution function, $P(p_{\mathrm{B}})$, for the ensemble of $z = 0$ configurations generated by the three biasing methods. The results are shown in Figure 2.9; the configurations from WT-metaD and US simulations have a strong bias to either one or both sides of the bilayer, while those obtained in the REUS (II) simulations have an unbiased committor value of $p_{\mathrm{B}} = 1/2$ and are therefore transition states. These important states do not seem to be sampled in either metadynamics or umbrella sampling calculations, which is likely the reason for their failure to converge. We reiterate that neither the number of tested configurations nor the number of unbiased trial simulations are sufficient to obtain a quantitative estimate of the committor distribution function, but they are adequate to test the ability of a simulation method to sample transition states at all.

2.3.3 Arginine

Finally, let us look at arginine, a cationic amino acid. In part due to their prevalence in cell penetrating peptides, the translocation mechanisms of such positively charged amino acids has received much attention.[42,45,46,79] It has been found that motion of arginine through a lipid bilayer is accompanied by the formation of water defect in the hydrophobic core that keeps the amino acid hydrated even when it is near the center of the bilayer. Lipid head groups line this membrane-spanning pore, which cause a major structural change in the bilayer. The calculation of translocation free energy profiles is hindered by the presence of barriers in degrees of freedom other than the position z, in particular due to the coupling of solute orientation and bilayer deformation.[20]

In the previous section we argued that the difficulties encountered in PMF calculations of zwitterionic alanine are due to strong electrostatic interactions between the solute and the lipids or water. If that were the case, then one would expect the same problems, perhaps even in an exacerbated way, to arise for arginine, which contains a positively charged guanidino group in addition to the zwitterionic carboxyl and amino group pair. This is indeed the case, as illustrated in Figure 2.11: WT-metaD, US, and initial REUS simulations yield free energy profiles that are far from symmetric and therefore evidently not converged, even though the calculated uncertainties seem to be rather small. It is worth noting that these PMFs were obtained from long simulations extending over a microsecond or longer, and still they show clear signs of non-convergence.

A look at the replica exchange pattern reveals the same underlying problem as we saw for zwitterionic alanine: there is a persistent gap that separates replica space into two regions with very few exchanges in-between.[53] This gap moves upward even more slowly in the arginine calculations, indicating that the relaxation time scales are even longer for the charged residue. As before, the principal difference between configurations above and below the gap is the symmetry of the defect that the solute creates in the bilayer. It is not surprising that the solution to the problem is therefore also the same: by starting a new set of replica exchange

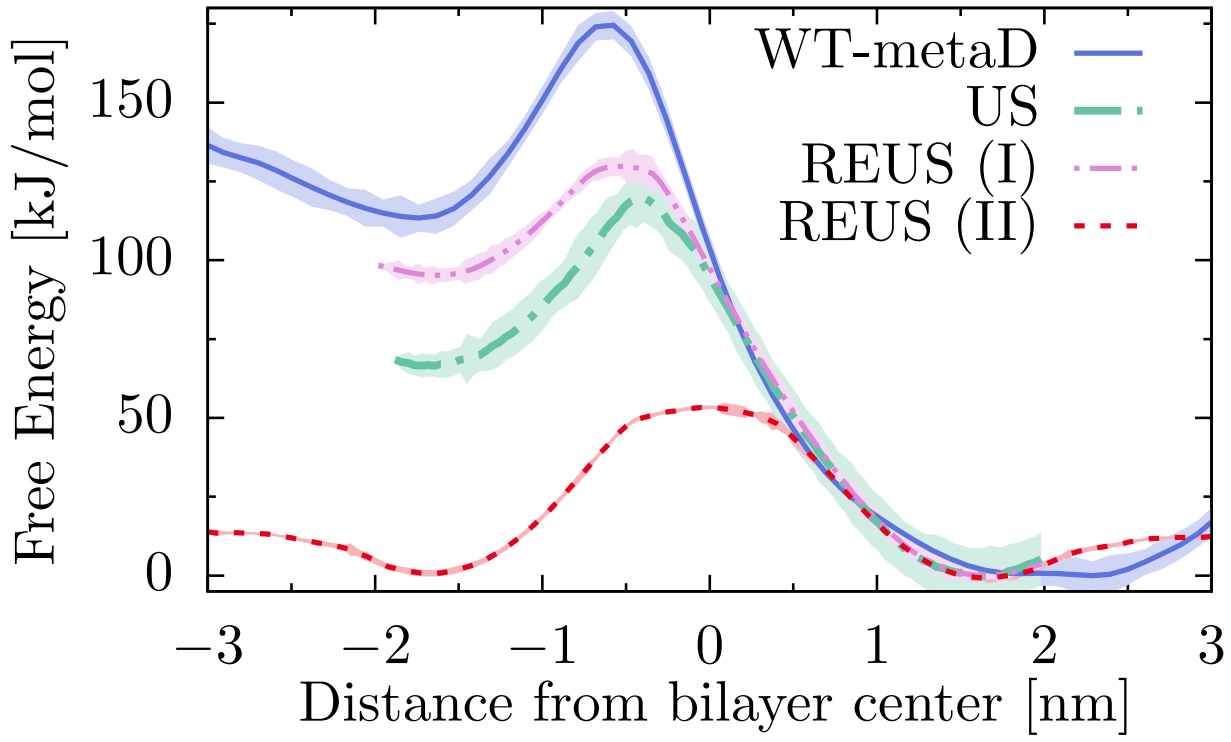


Figure 2.10: Free energy profiles for the translocation of a cationic arginine residue. As illustrated previously in Figure 2.3 for the case of zwitterionic alanine, WT-metaD, US, and initial REUS(I) simulations did not converge, as is evident from their asymmetric PMFs. The REUS (II) profile, obtained after extensive equilibration, is symmetric and has its maximum at $z = 0$.

umbrella sampling simulations with initial conditions that reflect a symmetric defect for replicas near the bilayer center, one obtains the symmetric free energy profile denotes REUS (II) in Figure 2.10. It exhibits shallow local minima for arginine positions close to the lipid head groups ($z \approx \pm$ 1.7 nm), which implies that arginine is attracted to the surface of the membrane. At the hydrophobic center of the bilayer, on the other hand, there is a significant barrier of approximate 50 kJ/mol that hinders translocation.

A committor analysis for the configurations generated by the various biasing methods also paints a familiar picture: the configurations generated in WT-metaD, US, and a first set of REUS simulations that contain arginine at the bilayer center have a strong bias towards

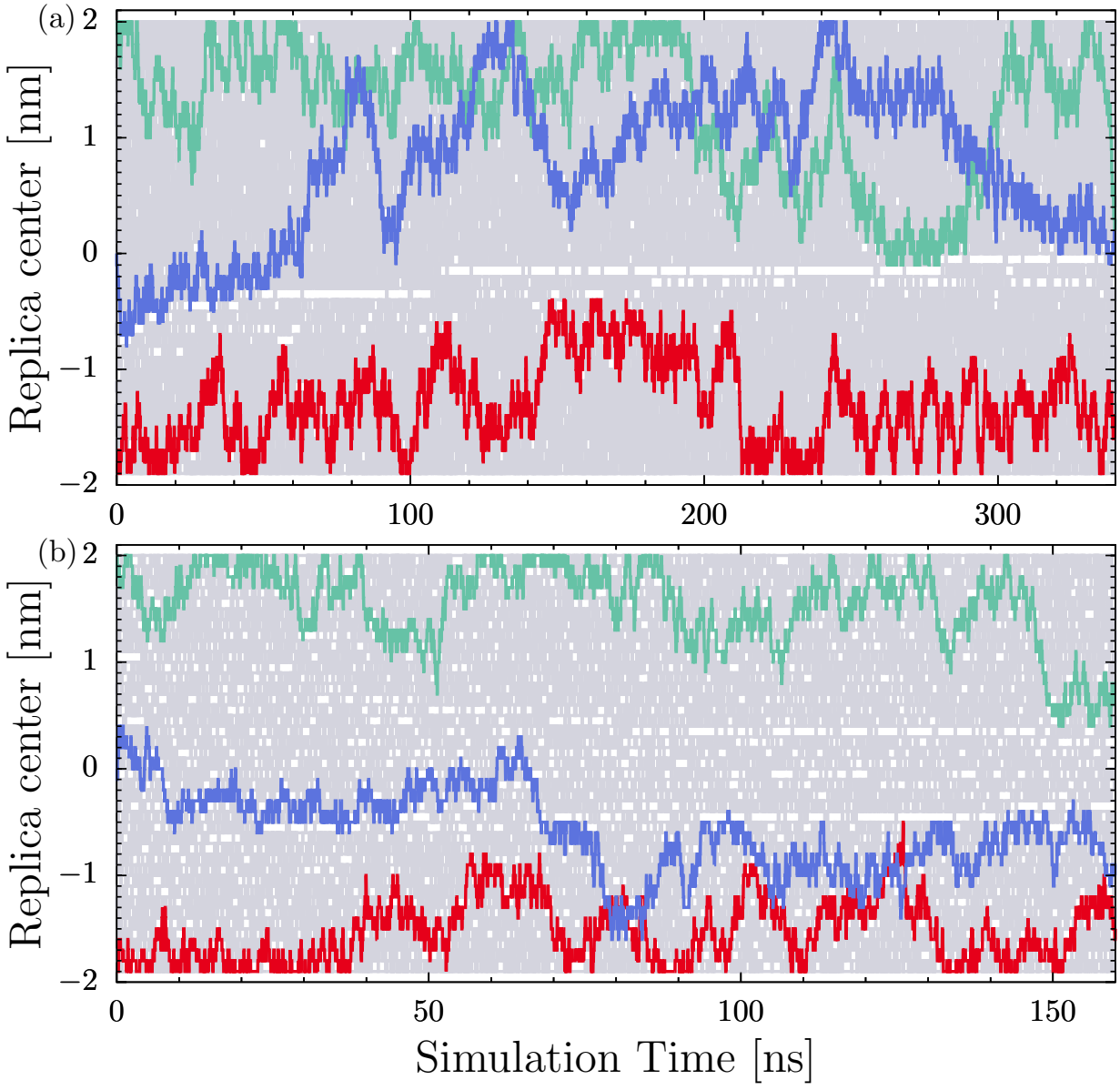


Figure 2.11: REUS results for zwitterionic form of arganine. (a) Exchange pattern lines for 40 replicas. Replicas starting at $z = 2.0$ nm, 0 nm and -1.9 nm highlighted. A gap in the exchange pattern is clearly visible for REUS(I) in panel (a). Using the final configuration of $z = 0$ replica as the initial condition for the center replicas in a new REUS(II) simulation (b), this gap is no longer present, indicating even exchanges between all 40 replicas.

one side of the membrane.[53] In other words, these configurations are not transition states. Repeating this calculation for the second REUS simulation, we find that all tested configurations from the replica centered at $z = 0$ have a committor value close to 1/2. It is this simulation, which includes sampling of the actual transition states, that yields the correct free energy profile as shown in Figure 2.10.

2.4 Discussion

In this chapter, we have seen that all three tested biasing methods perform well for uncharged or weakly polar solutes, but struggle when one wants to compute translocation free energy profiles of strongly polar or charged molecules; interestingly, this is also true for a very small positively charged ion like Na^+.[53] In REUS simulations the underlying cause for this difficulty can be readily seen in the exchange pattern: the slow conversion of configurations in which a solute close to the bilayer center is initially surrounded by lipids from only one leaflet towards a more symmetric state. This process is illustrated in Figure 2.7 by the horizontal arrows, and its progress is reflected by the slow dynamics of the gap in Figure 2.6(a). Once this relaxation has occurred, a second REUS calculation yields reliable PMFs.

While there is no equivalent to the exchange pattern in umbrella sampling calculations, it is likely that the same mechanism is at work here. In both US and REUS simulations, one computes trajectories in which the solute is restrained in the bilayer center by a harmonic bias potential. The simulations will converge as long as the length of these trajectories is longer than the time scale of the required lipid reorganization. Metadynamics simulations, on the other hand, behave differently: because the solute is continually driven across the bilayer, it is unlikely to remain within the hydrophobic core for a sufficiently long time to allow the formation of a symmetric defect or pore. Every time it diffuses towards the bilayer/water interface and then re-enters the lipid tail region, the solute is likely to drag with it lipid head groups and water molecules from the proximal side, thereby creating yet another one-sided defect. An exaggerated illustration of this process is the counterclockwise cycle in Figure 2.7. However, this side-stepping of the actual transition state can likely be avoided by a judicious

choice of the metadynamics parameters.

The behavior depicted in this figure is in fact a common concern in free energy calculations of many complex systems: because we are in interested in $F(z)$, we bias the system along the coordinate z. However, there are additional barriers in other coordinates; for example the lipid reorganization that is required for the horizontal transitions in Figure 2.7. Because we do not bias these "hidden" variables, the calculation of free energy profiles remains challenging.

An obvious solution to this problem is to also bias these collective variables in addition to the position z that is of primary interest. This can be accomplished by introducing multi-dimensional biasing potentials. An intriguing alternative for metadynamics calculations is to simulate multiple trajectories in parallel, each using a different CV. Every so often these trajectories are allowed to exchange configurations similar to REUS calculations. This bias exchange metadynamics (BE-metaD[80]) scheme facilitates the calculation of free energy profiles in one CV by driving the exploration of additional variables.

It is usually not known in advance which variables should be biased to accelerate the convergence of free energy calculations. Our results from the previous section show that breaking and forming solute–lipid and solute–water contacts is a bottleneck that impedes equilibration; therefore biasing the number of such contacts should be beneficial. That this is indeed the case was shown by Ghaemi and coworkers, who studied the translocation of ethanol through a phospholipid bilayer using BE-metaD in multiple CVs, including the coordination numbers that measure the solvation of the solute by lipid heads, tails, and water.[81]

For the small solutes that we have considered here, we have used their center of mass (COM) for the definition of the distance variable z. For more complex solutes this is not necessarily the ideal choice. Hinner and coworkers have shown that the convergence rate of PMF calculations for a voltage-sensitive dye molecule depends on which solute atom is used in the definition of z.[82] PMF calculations for even bigger molecules can benefit from using multiple CVs that measure the distance between different groups of the solute and the membrane.[83] Alternatively, one can bias the orientation of the solute molecule with respect

to the bilayer. If the solute itself has significant conformational flexibility then one should consider including such internal CVs as well in the biasing scheme.

There is a large number of biasing methods and CVs to choose from, which might seem overwhelming at first. One can take comfort in the fact that any combination will at least in principle converge to the correct result if only the simulations are performed over long enough times. As we have seen, however, determining that the calculation is converged can be rather difficult. The standard algorithms to estimate statistical uncertainties fail if there are systematic sampling deficiencies, which seems to occur frequently. How then is one to assess convergence? For the practically important case of symmetric bilayers, we propose the following diagnostic checks:

i. Is the free energy profile symmetric with respect to the bilayer center? This is a strict requirement for symmetric membranes.
ii. Does the free energy profile exhibit a plateau region in the center of the bilayer? Mirrored or symmetrized free energy profiles often exhibit a kink at $z = 0$. This could be a warning sign for insufficient convergence, as such a kink would imply a discontinuity in the mean force, which is unlikely to be physically meaningful.
iii. When performing an REUS calculation with initial conditions taken from simulation trajectories at varying z-positions of the solute, is there an apparent gap in the resulting exchange pattern? The presence of such a gap indicates the existence of barriers in replica space that will impede proper sampling.
iv. For solutes that are most likely found in bulk water or at the membrane-water interface, do the sampled configurations at the center of the bilayer exhibit a bias towards one side of the membrane over the other? For symmetric membranes it is likely that many configurations of the $z = 0$ ensemble are transition states, i.e., they have equal probability of evolving towards states in which the solute is on either side of the membrane. Calculating the committor distribution function is an effective if time-consuming way to test this property: if it is peaked at $p_{\mathrm{B}} = 0.5$ then one indeed

samples the transition state ensemble, which is not the case if it is peaked at $p_{\mathrm{B}} = 0$ and/or $p_{\mathrm{B}} = 1$.

The calculation of membrane translocation and computation of free energies is a very active field of biophysics, and one that has seen dramatic progress over the past decade. New biasing methods, better collective variables, faster computers, more accurate molecular mechanics force fields, and simulation software that has become more user friendly all contribute to this exciting development. It is a sign of success that these calculations are now regularly performed during the development phase of new compounds.

Chapter 3

CHOLESTEROL CHEMICAL POTENTIAL

3.1 Introduction

Cholesterol is an integral and abundant component of mammalian cells. Cholesterol is known to alter the lipid rigidity, permeability, thickness, packing and order.[26–28,84] A number of previous studies, both experimental and computational, done with a variety of techniques, have indicated that cholesterol interacts differently with different types of lipids.[85–91] The general consensus is that cholesterol interacts more favorably with saturated phospholipids than with unsaturated phospholipids in membranes.[29] These favorable interactions reduce the chemical potential of cholesterol and decrease its activity in certain lipid environments. The favorable partitioning of cholesterol into ordered, saturated lipids also alters the phases within lipid bilayers. Lateral domains are formed in biological membranes, and are often discussed in terms of cholesterol-lipid interactions. Cholesterol-lipid interactions have been qualitatively described by three conceptual models. These three models make different predictions for the cholesterol activity within a lipid membrane. In addition, the dependence of cholesterol-lipid interaction on lipid headgroup type versus lipid tail differs on each of these models. The three methods are briefly described below:

Umbrella Model The umbrella model suggests that cholesterol relies on the coverage by the large hydrophilic headgroups from neighboring phospholipids to prevent the exposure of the non-polar part of the cholesterol to water.[92] Because it costs much more energy to cover a cluster than a single cholesterol molecule, cholesterol molecules have a tendency not to form clusters. This model is solely based on cholesterol lipid headgroup interactions and is largely independent of PC lipid tail structure. By this model, regular clusters of lipids are formed at several characteristic compositions at which cholesterol activity changes dramatically.

Superlattice Model The superlattice model suggests that cholesterol forms superlattices in lipid bilayers. At the concentrations where these superlattices are formed, dips in free energy have been predicted.[93] These dips in free energy result in spikes in chemical potential. The repulsive force among cholesterol plays the dominant role in cholesterol-lipid interaction and is independent of lipid tails. In this model, cholesterol accessibility changes at several characteristic compositions.

Condensed-Complex Model The condensed complex model suggests that cholesterol and phospholipids react reversibly to form low energy complexes that occupy smaller lateral areas.[94] Because the condensed complex has a low energy, the model predicts that cholesterol can form condensed complexes with lipids with which it can mix favorably, hence the model is largely based on lipid type. The accessibility increases at a characteristic composition of cholesterol.

The direct measurement of cholesterol's chemical potential in a membrane is experimentally challenging, and as a result activity or accessibility is often calculated as a proxy for chemical potential. Cholesterol oxidase (COD) activity assays have been used to study the accessibility of cholesterol in a lipid bilayer.[95] Other studies have used methyl-β-cyclodextrin assays to study the rate of cholesterol depletion in PC-lipid membranes.[96] From these experiments, different values for the cholesterol concentration at which the activity changes[95,96] have been inferred. In addition, direct cholesterol partitioning coefficients have also been determined between vesicles of different lipid types where methyl-β-cyclodextrin was used as a reference state to resolve the known issues of the low cholesterol exchange rate between vesicles and the difficulties in separating donor vesicles from acceptor vesicles due to vesicle fusion.[30]

In lieu of experimental challenges, molecular dynamics simulations have been used to study cholesterol-lipid interactions.[31,97] These simulations have verified experimental findings like the cholesterol condensing effect and its preference for saturated lipids. In addition, MD simulations have been used to calculate the free energy of cholesterol depletion from selected lipid bilayers to water using both all atom and coarse grained US. These simulations

confirm that the transfer of cholesterol from unsaturated to saturated lipids is exothermic which agrees with calorimetric data.[98] They have also shown that cholesterol has the lowest affinity for polyunsaturated lipid tails. In this chapter, we use MD enhanced by replica exchange umbrella sampling to calculate cholesterol chemical potential in lipid bilayers with various cholesterol concentrations. Equating depletion free energy to chemical potential of cholesterol minus a constant, we show that there is no characteristic concentration where the chemical potential dramatically changes. Our result says that none of the experimental models describe the cholesterol-lipid interaction accurately, and instead the chemical potential increases linearly with cholesterol concentration. Such results have been recently reported in the literature.[99] We also investigate the effect of tail and headgroup chemistry of the phospholipid on cholesterol partitioning into a lipid bilayer and find that the chemical potential is sensitive to both.

Using water as the reference state to calculate the chemical potential can provide great insight into cholesterol behavior within the lipid bilayer, but these calculations are computationally expensive and prone to large uncertainties. An additional goal of the project is to compute equilibrium properties of cholesterol between different lipid environments and to eliminate the usage of water as the reference state. We do so using a method similar to a very recently published one.[100] There, the authors have devised a binary bilayer system (BBS), where two patches of bilayers composed of different lipid types are embedded in bulk water. Cholesterol is distributed evenly within the two patches and is allowed to equilibrate, while the lipids are restrained and not allowed to mix using a soft restraint potential. Our BBS system behaves the same way; we construct a BBS by bringing two lipid patches laterally together and evenly distributing cholesterol between the lipids. We allow for cholesterol to diffuse between either lipid patches;. the lipids are restrained using soft harmonic potentials. The results from BBS are in great agreement with our REUS results, with smaller uncertainty and much faster computational time.

3.2 Methods

3.2.1 Simulation Details

Lipid bilayers with varying cholesterol concentrations were constructed using coarse grained MARTINI 2.2 forcefield.[101] Coarse graining helps speed up bilayer simulations by 3-4 order of magnitude and accelerates the sampling of the phase space by 5-10 fold compared to all atom MD models. The MARTINI force field applies a four-to-one mapping of heavy atoms into a "bead". Explicit solvent representation was included in the force field with a four-to-one mapping of water molecules into a MARTINI water bead. The bonded interactions between beads were implemented as harmonic potentials with respect to stretching and bending degrees of freedom.

We performed all simulation using Gromacs 5.1.4[55] with the Plumed 2.1 and Plumed 2.4 suite.[56] All simulations were performed under periodic boundary conditions and a temperature of 320° K regulated using the Nose-Hoover thermostat. The long range electrostatic interactions were computed using the fourth order PME method [57] with a Fourier spacing of 0.12 nm. Both the van der Waals and Coulombic interactions were set to 1.2 nm. Bond lengths within the solutes and lipids were constrained using the LINCS algorithm.[58]

3.2.2 Bilayer Composition for cholesterol depletion

We systematically constructed 7 lipid bilayer systems for each of the 7 lipid types: i) 1,2-dilinoleoyl-sn-glycero-3-phosphocholine (DLiPC) ii) 1,2-dioleoyl-sn-glycero-3-phosphatidylcholine (DOPC) iii) 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) iv) 1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine (DPPC) v) 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) vi) 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) vii) 1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine (DPPE). 4 out of these 7 lipids, have phosphatidylcholine (PC) headgroups, and the remaining 3 have phosphoethanolamine (PE) headgroups. The PC headgroups are bigger than the PE headgroups. Within the same headgroup, the fatty acid tails also have different degrees of unsaturation. DLiPC (di-18:2-PC) has two double bonds

in each of its tails, DOPC/DOPE (di-18:1-PC/PE) has one double bond in each of its tails, POPC/POPE(16:0-18:1-PC/PE) has one double bond in one of its tails, and DPPC/DPPE (di-16:0-PC/PE)] is fully saturated.

The cholesterol concentration in the lipid bilayer varies from 0% to 60%, covering the biologically relevant range. The total number of lipid components (phospholipids and cholesterol) was always maintained at 200, while individually adjusting the cholesterol and lipid content. From here on, we refer to our lipid system by the concentration of cholesterol in them defined as:

$$\%C = \frac{N_C}{N_L + N_C}, \tag{3.1}$$

where N_C is the number of cholestero moleculel, and N_L is the number of phospholipid molecule in the bilayer respectively.

3.2.3 Depletion Free Energies

In the first chapter, we established that REUS is an efficient biasing method, and will use this technique to calculate depletion free energy, $\Delta G_{depletion}$. It is defined as the free energy gain (or cost) of removing one cholesterol molecule from the lipid bilayer to bulk water. Mathematically,

$$\Delta G_{depletion} = G(N_L, N_C) + \underbrace{G_{solvation}}_{constant} - G(N_L, N_C + 1). \tag{3.2}$$

We calculate the free energy required for this process as a function of the distance between the center of mass of the transferred cholesterol and the lipid bilayer (z). A typical simulation system used to compute the depletion free energy is shown in Fig 3.1. We begin with cholesterol in the membrane as in panel(a) and drag one tagged cholesterol to bulk water as shown in panel(b). Each simulation trajectory is divided into 40 windows; in each window the cholesterol molecule is restrained at a value of z centered 0.1 nm apart. Each window is run for 500 ns at constant pressure of 1 atm using the Parrinello-Rahman baro-

stat, resulting in 20 μs of total simulation time. Like in the first project, the simulation results are stitched together, minimizing the variance using WHAM, and the errors were estimates using bootstrapping as implemented in `g_wham`.[71] We thus calculate free energy for cholesterol depletion for 49 different systems.

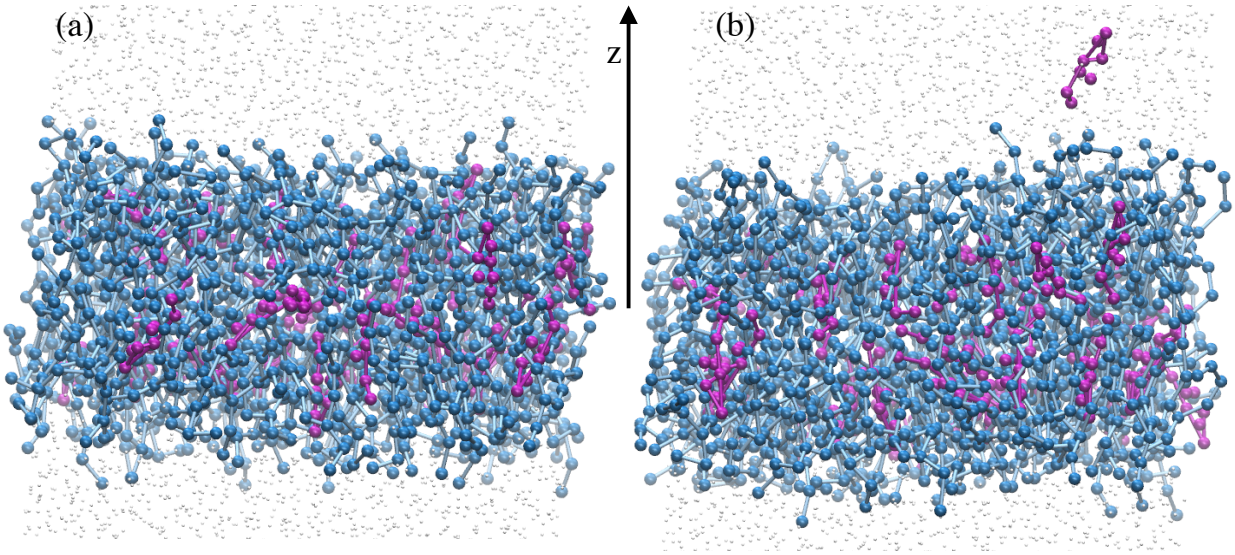


Figure 3.1: Simulation snapshot to compute depletion free energy. We remove one cholesterol molecule from the bilayer with N_L lipids and N_C+1 cholesterol (a) and place it in bulk water (b). The bilayer in (b) has N_L lipids and N_C cholesterol. POPC lipid is represented in blue, cholesterol in purple, and water in faded white.

3.2.4 Chemical Potential

Chemical potential is the molar Gibbs free energy. From the free energy of depletion, we derive the chemical potential of cholesterol (μC) in each lipid bilayer:

$$\mu C = \frac{\delta G}{\delta N_C} = G(N_L, N_C + 1) - G(N_L, N_C)$$

$$\therefore \mu C = G_{solvation} - \Delta G_{depletion}.$$

Our definition of $\Delta G_{depletion}$ calculates the difference between the value of the free energy when then PMF plateaus versus the minimum (at z*). This assumes that cholesterol when in the membrane is restricted to position z* in the bilayer, although cholesterol can diffuse along the lipid. To capture the fact that cholesterol is not restrained at a certain position in a biological membrane, we introduce a correction term ($\Delta G_{correction}$). This correction term accounts for the energy that is released when the positional constraint is removed. The free energy when cholesterol is at z^* is given by

$$\Delta G_{z^*} = -k_{\mathrm{B}}T \ln \frac{e^{-\beta F(z^*)}}{\int_0^{\mathrm{inf}} dz e^{-\beta F(z)}}, \tag{3.3}$$

The free energy when cholesterol is not static and can take any position within the membrane (memb) is

$$\Delta G_{z^{mem}} = -k_{\mathrm{B}}T \ln \frac{\int_0^{\mathrm{mem}} dz e^{-\beta F(z)}}{\int_0^{\mathrm{inf}} dz e^{-\beta F(z)}}. \tag{3.4}$$

The free energy gained when the restraint is removed is

$$\Delta G_{correction} = \Delta G_{z^{mem}} - \Delta G_{z^*} = k_{\mathrm{B}}T \ln \frac{e^{-\beta F(z^*)}}{\int_0^{\mathrm{mem}} dz e^{-\beta F(z)}}. \tag{3.5}$$

So that,

$$\mu C = \underbrace{G_{solvation}}_{\text{constant}} - \Delta G_{depletion} + \Delta G_{correction}, \tag{3.6}$$

where $G_{solvation}$ is a constant offset that is independent of bilayer composition.

3.2.5 Binary Bilayer System

We constructed three BBS with different phospholipids and cholesterol concentration. First, using the MARTINI Maker module of CHARMM GUI [102] we built a pure bilayer with DLiPC lipid and 30%C, then we replaced half of the DLiPC with DPPC resulting in a system that

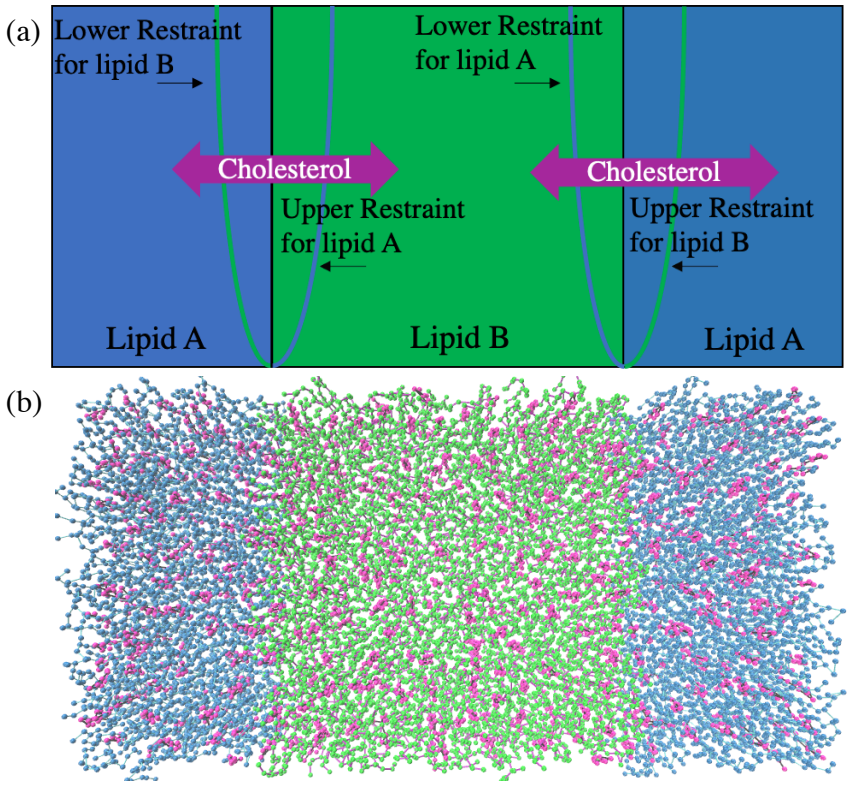


Figure 3.2: Binary bilayer system. We devise a laterally semi-permeable membrane where the cholesterol molecules are free to diffuse between the lipid but the lipid movement is restrained by a restraining potential (a). Simulation snapshot of a BBS containing DPPC in blue, cholesterol in purple and DLiPC in green (b). The cholesterol is initially evenly distributed between the two lipids patches.

contains 35%DLiPC- 30%Cholesterol-35%DPPC; we call this system DLiPC-DPPC. We followed a similar protocol to build two other systems with 40%DOPC- 20%Cholesterol-40%POPC and 40%POPE- 20%Cholesterol-40%POPC and call them DOPC-POPC and POPE-POPC, respectively. 10 replicates of each system were run for 10μs and only the last 8 μs was used for analysis. All BBS systems are anisotropic and the box size does not fluctuate throughout

the simulation. Below, we will see that this required because the restrains that act on lipids are on absolute coordinates. In these systems, cholesterol can laterally diffuse between two lipids. The lipids, however, are restrained using the MultiColvar module in PLUMED 2.4. For each phosphate bead, a restraining potentials of the form

$$k\left(\frac{(x_i - a_i)}{s}\right)^e,$$

is applied. In the equation above, k is the force constant, x_i is the x component of the position of the i^{th} of phosphate atom, a_i is the value of x_i where the potential is imposed, s is the scaling factor, and e is an exponential term. We pre-define k to 100 kJ/mol, s to 1 nm, and e to 12. We also define a_i; at each time step in simulation, a phosphate atom experiences two biasing potential with two values of a_i which determines the upper and the lower limit of the region of the phase space accessible to the atom during the simulation. A schematic of the BBS along with a simulation snapshot is shown in Fig. 3.2.

To validate the simulation approach and to illustrate the accuracy of the model, we compare i) bilayer thickness and ii) tail order parameter of the BBS against two pure membranes each containing equilibrium concentration of the lipid and cholesterol as determined by the BBS.

3.3 Results

3.3.1 Free Energy

We calculate the free energy for removing one cholesterol molecule from the lipid bilayer to the bulk water for all 49 systems in Fig. 3.3 and Fig. 3.4. The depletion free energies are all in the range of 100 kJ/mol - 75 kJ/mol, which matches with previously published simulation result.[31] For all lipids, there are free energy troughs at the equilibrium positions of the cholesterol. Upon addition of cholesterol, the trough moves farther away from the bilayer center, indicating increasing bilayer thickness. Eventually, at higher cholesterol concentration, the position of the minimum either remains constant or slightly moves towards

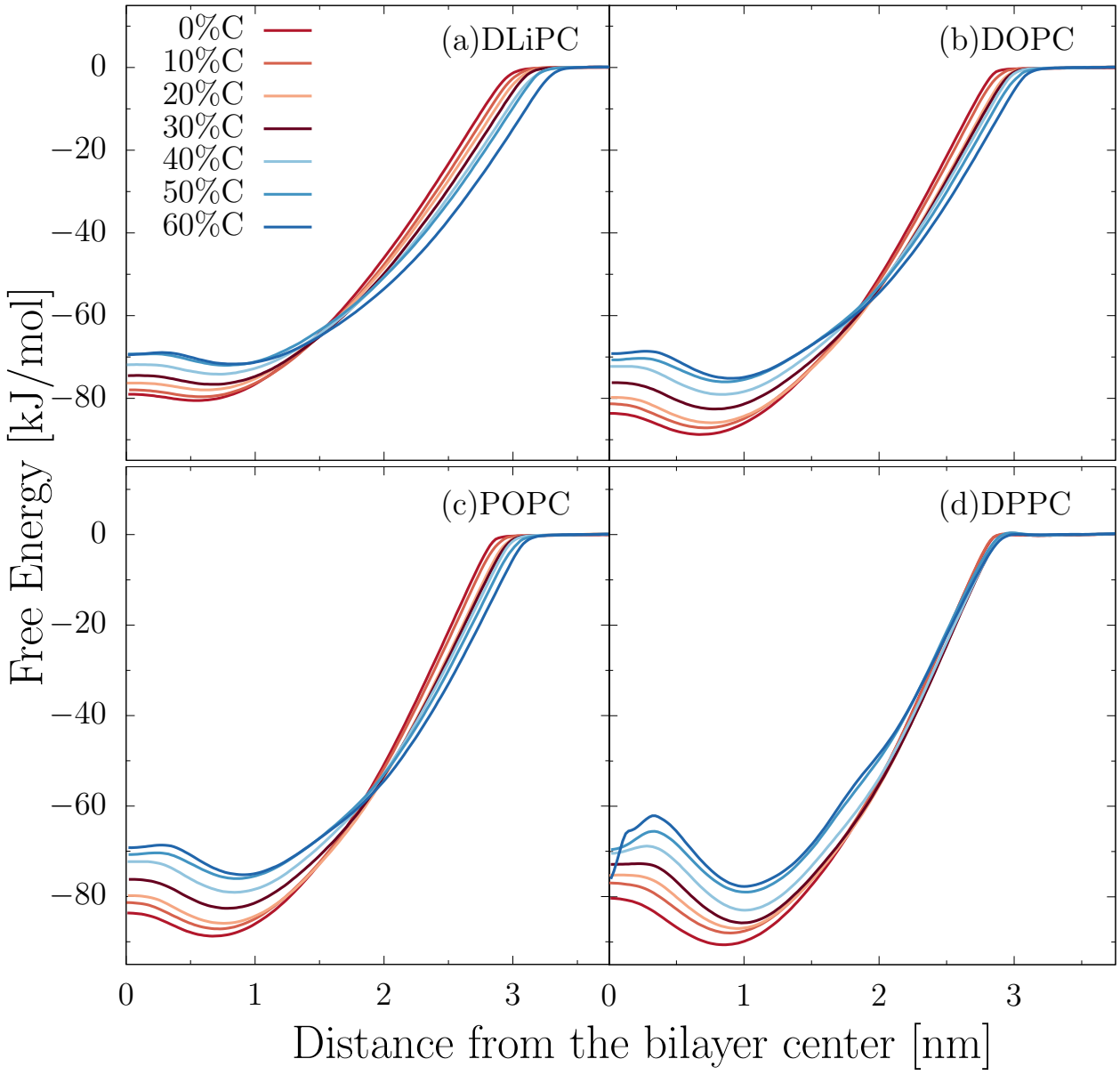


Figure 3.3: Free energy profiles of cholesterol depletion from PC lipids. The free energy of depletion decreases as cholesterol content increases in the membrane, suggesting cholesterol has greatest affinity to bilayer with the lowest cholesterol content.

the bilayer center at high cholesterol in agreement with bilayer thickness results. The free energy profiles smoothly increases as cholesterol depletes from the trough to bulk water,

where the profiles flatten when the cholesterol freely diffuses in water and the cholesterol tail stops interacting with the lipid. The energy required to deplete cholesterol decreases with

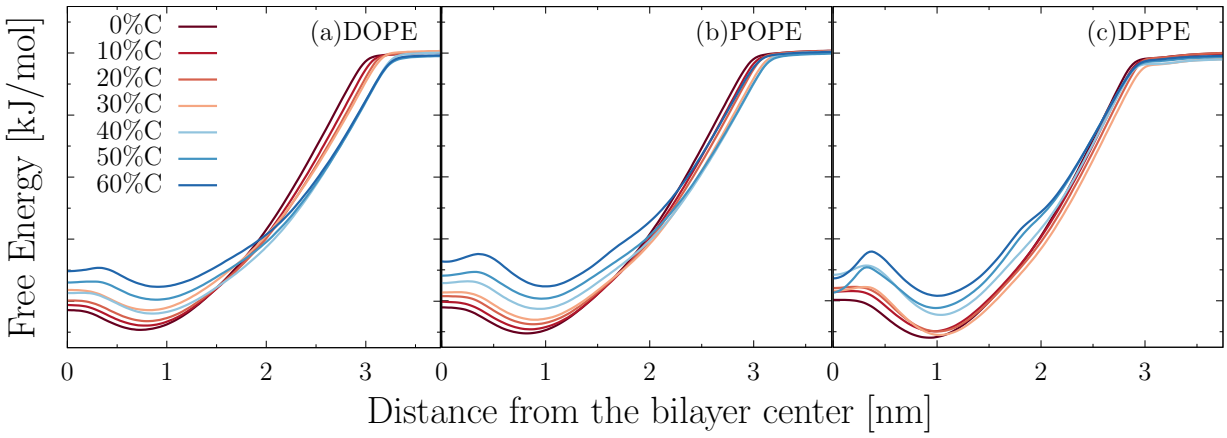


Figure 3.4: Free energy profiles of cholesterol depletion from PE lipids. The free energy of depletion decreases as cholesterol content increases in the membrane, suggesting cholesterol has greatest affinity to bilayer with the lowest cholesterol content.

increasing cholesterol content. One could speculate that with no cholesterol present, there is greater free volume within lipid tails. This free volume allows cholesterol to effectively pack within phospholipid and removal of any cholesterol requires more energy. However, as cholesterol content increases, it induces order and rigidity in the lipid bilayer, as a result of which there is less free volume in the lipid bilayer and at this point, depletion of cholesterol requires less energy.

3.3.2 Chemical Potential

We can approximate the chemical potential of cholesterol from the depletion free energy. Chemical potential is depletion free energy shifted by a constant. Knowing the chemical potential of cholesterol in the lipid bilayer allows to predict the activity of cholesterol in each of the bilayers. In Fig. 3.5, we plot the chemical potential as given by equation (3.2.4) as a function of cholesterol concentration for all 7 bilayers.

The chemical potential line for DLiPC is above all the other lines, suggesting that the interaction between cholesterol and the doubly unsaturated DLiPC is unfavorable. All other lines are indistinguishable within error bars, although two general trends exist: i) the PC lines are higher up than the PE lines and ii) within the same headgroup type, the chemical potential decreases as the degree of unsaturation in the lipid tail increases. The first speculation that cholesterol prefers PE lipids more than PC lipids does not seem to agree with experimental studes.[30,103,104] The second trend, however, reiterates with many experimental results and is due to cholesterol interacting more favorably with saturated lipids than unsaturated lipids. Cholesterol is known to partition into ordered and rigid phospholipids.[30,31,96]

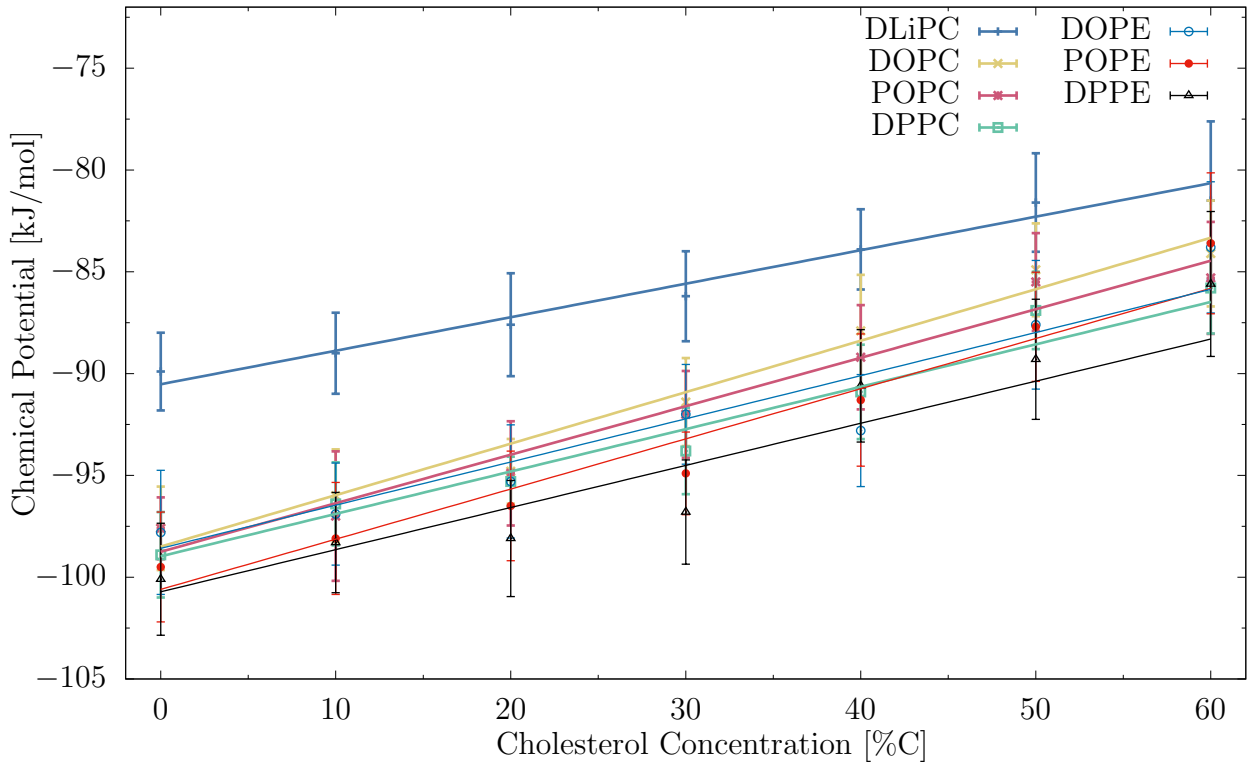


Figure 3.5: Cholesterol chemical potential in all PC and PE bilayers. The chemical potential of the DLiPC bilayer is higher than all other bilayers. The chemical potential curve for PCs are generally higher than those for PEs.

The principal goal of this study is to investigate the change in cholesterol activity within

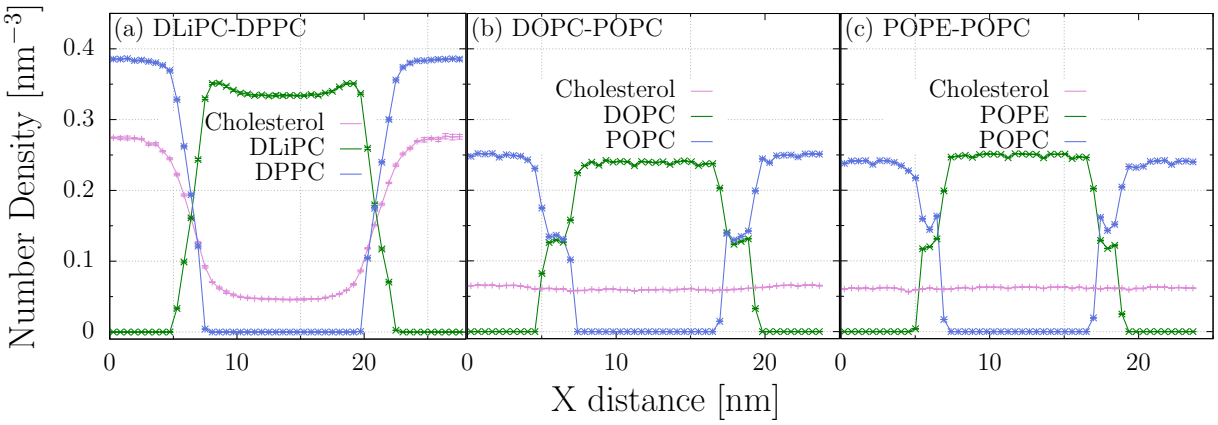


Figure 3.6: Density of cholesterol and two lipid species in the BBS. Cholesterol prefers DLiPC over DPPC (a), but can not differentiate between both DOPC-POPC (b) and POPE-POPC (c), within error bars.

a lipid type as more cholesterol is added in the lipid bilayer in order to determine if any of the models mentioned in the introduction describes the cholesterol chemical potential. As can be seen from Fig. 3.5, the chemical potential increases linearly with increasing cholesterol content. In terms of activity, we can say that activity of cholesterol gradually changes upon more cholesterol being introduced to the bilayer. Although the general increasing trend agrees with all the three proposed models, we find that none of the models capture all of cholesterol-lipid interactions. We do not find characteristic concentration(s) at which the activity changes dramatically, neither do we find dips in chemical potential as suggested by the superlattice model.

Our results do come with some limitations. First, we calculate the chemical potential for lipid bilayers whose cholesterol concentration differs by 10%. Although unlikely, there could be dramatic changes in the chemical potential that we do not capture due to the resolution of the x-axis in Fig. 3.5. Second, the uncertainties associated with chemical potential on the y-axis are relatively big ($\pm$ 3 kJ/mol) as well. The uncertanties and can obscure us from noticing any drastic change.

3.3.3 Binary Bilayer System

Replica exchange umbrella sampling is an excellent method to calculate the depletion free energy. Often, we are interested in the partition free energy which quantifies the amount of energy required for a cholesterol molecule to move from one type of bilayer to another. Many experiments have reported the partition free energies to be in the range of 1 kJ/mol - 5 kJ/mol[30,32] regardless of the lipid type. In principle, the partition free energy can be calculated from depletion energies. The problem arises due to the magnitude of the associated uncertainties, which makes it difficult to quantify the cholesterol affinity between two lipid types. The binary bilayer system works around this by eliminating the usage of water as the reference, and makes the calculation much more efficient.

Cholesterol Partitioning

From the BBS, we can calculate the density of each species as a function of the box length as shown in Fig. 3.6. In panel a), the density of cholesterol is much higher in the region where the density of DPPC is higher than the corresponding DLiPC, suggesting cholesterol interacts much more favorably with the saturated tails of DPPC. From panel (b) and (c), however, we see that the cholesterol density is relatively uniform between both DOPC and POPC, and POPE and POPC within the error bars. This is consistent with the results from REUS, where cholesterol could not differentiate between these lipid types. In panel (b) and (c), we also notice a kink in the interface region where two lipids are allowed to diffuse and mix at $\sim$7 nm and $\sim$17 nm.

We can further look at the cholesterol mole fraction to quantify the cholesterol content in individual lipid bilayer within a BBS. From Fig. 3.7 we can see that the DLiPC-DPPC BBS with 30% global cholesterol concentration, equilibrates into two distinct lipid patches with $\sim$ 42 % local cholesterol concentration in the DPPC rich phase and $\sim$ 13 % in the DLiPC rich phase. The DOPC-POPC BBS with 20% global cholesterol concentration remain as it is; after long equilibration the system forms two distinct phases both with $\sim$ 20 % local

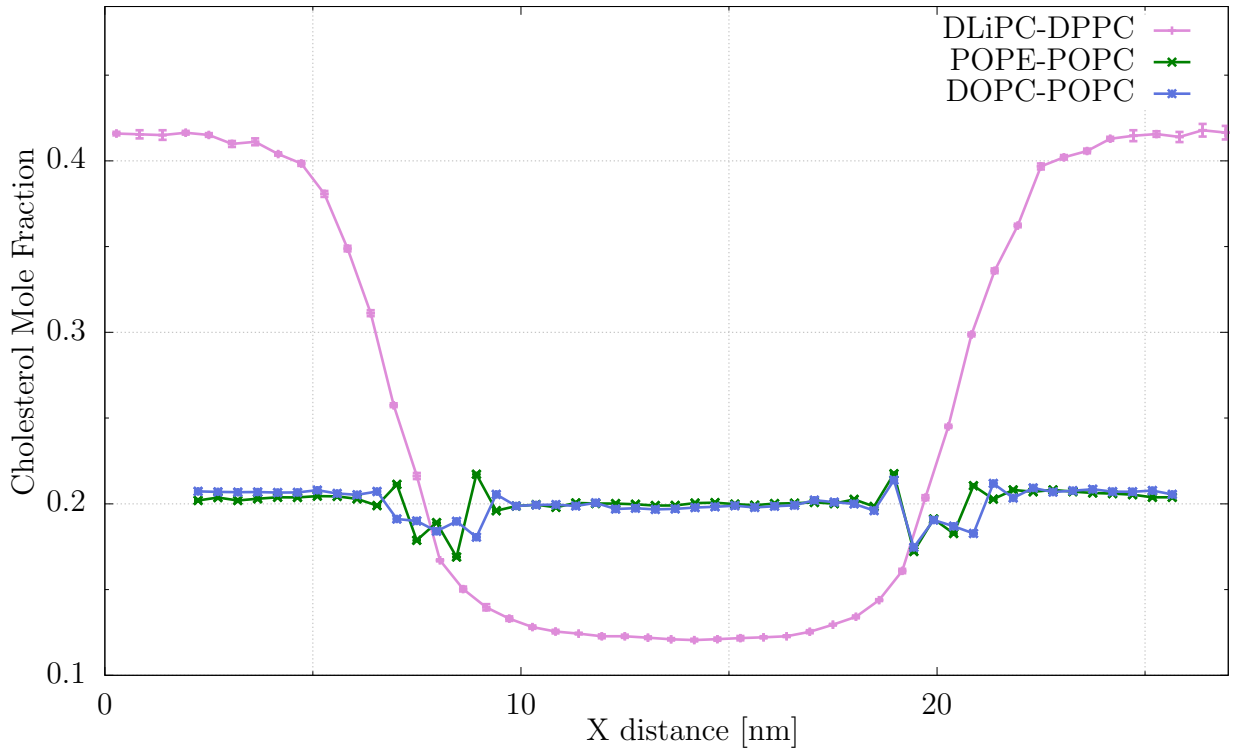


Figure 3.7: Mole fraction of cholesterol in three simulated BBS. The mole fraction of cholesterol is much higher in DPPC rich region than DLiPC rich region. The mole fraction of cholesterol is the same in DOPC and POPC and POPE-POPC within error bars.

cholesterol concentration in DOPC rich and POPC rich phase. The result for POPE-POPC is very similar; the equilibrated system contains two different lipids patches with $\sim$ 20 % local cholesterol concentration.

Equilibrium properties

Before we make thermodynamic predictions from the BBS, it is important to verify that the equilibrated BBS behaves the same as pure bilayers. From here on, we refer to a bilayer as pure if it only contains one type of phospholipid and cholesterol molecule. To verify this, we run 6 new simulations with 42% cholesterol in DPPC, 13% cholesterol in DLiPC, 20% cholesterol in DOPC, 20% cholesterol in POPC, 20% cholesterol in POPE, 20% cholesterol in

	BBS	Pure Lipid
DLiPC	3.55 ± 0.1	3.44 ± 0.1
DPPC	4.33 ± 0.1	4.30 ± 0.1
DOPC	3.80 ± 0.1	3.79 ± 0.1
POPC	3.96 ± 0.1	4.01 ± 0.1
POPE	3.90 ± 0.1	3.95 ± 0.1
POPC	3.94 ± 0.1	4.01 ± 0.1

Table 3.1: Bilayer thickness of BBS and pure lipid bilayer. Within error bars, the lipid thickness stays the same for both BBS and pure lipid.

POPC. The amount of cholesterol in these pure bilayers were determined by mole fractions from the corresponding BBS.

We calculate the bilayer thickness and lipid tail order parameter in Fig 3.8 and Fig. 3.9 to determine if the lipid patches of BBS behave as pure bilayers. In Fig. 3.8, we plot the phosphate number density versus the distance from the bilayer center. Each peak is horizontally shifted such that $z = 0$ corresponds to the center of the bilayer. In panel a), we show the results for DLiPC-DPPC. The distance between the two peaks, which gives the measurement for the bilayer thickness, is the same for both DPPC and DLiPC in the BBS and in the pure bilayer. The DPPC thickness in a BBS is 4.33 nm and the thickness is 4.30 nm in a pure DPPC bilayer containing 42% cholesterol. The DLiPC thickness in a BBS is 3.55 nm and the thickness in a pure DLiPC with 13% cholesterol is 3.44 nm. In panel (b) , a similar analysis for DOPC-POPC shows that the bilayer thickness of POPC as a part of BBS is 3.94 nm and that of a pure POPC with 20% cholesterol is 4.01 nm and the bilayer thickness of DOPC is 3.80 nm and 3.79 nm in the BBS and in pure bilayer system. In panel (c), the bilayer thickness of POPE is 3.90 nm and 3.95 nm, in a BBS and a pure bilayer, respectively. This shows that the patches of binary bilayer in BBS, behave the same way as a pure isolated binary lipid bilayer.

The lipid tail order parameter P_2, defined as $P_2 = \langle \frac{1}{2}(3cos^2\theta - 1) \rangle$, provides structural information about the lipids. In the equation above, θ is the angle between the C−H bond

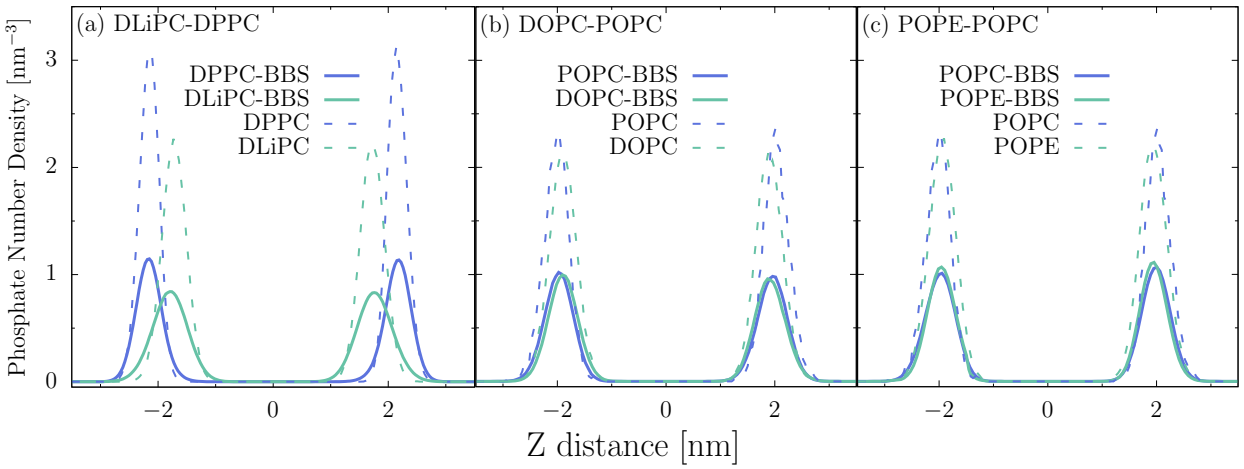


Figure 3.8: Phosphate density in the BBS and in pure binary bilayers. The distance between the peak of the phosphate density stays the same for each lipid in both BBS and in pure binary setup.

vector and bilayer normal averaged over all of the lipids and all of the sampling time in an all atom molecular dynamic simulation[105] . However, in our case the order parameter describes the orientation of bond connecting two CG beads with respect to bilayer normal[106] and is only qualitatively correct. Perfect alignment with the bilayer normal is indicated by $P_2 = 1$, perfect anti-alignment with $P_2 = -0.5$, and a random orientation with $P_2 = 0$. Although this value can not be used to compare our results against NMR order parameters, we can compare different simulation results using the MARTINI order parameter. In figure 3.9, we plot order parameters for all three systems and their corresponding bilayer patches. In the plot, the order parameter for BBS are represented in points and the order parameter for the pure lipids are represented in lines. As we can see, the lines nearly connect all the points, suggesting that all lipids in the BBS behave the same way as in pure bilayers. For all PCs, in all three panels, the first three points are the same because PC lipids differ in their tail beads and not their head beads. In panel (a), we see that that DPPC lipid beads are more aligned than DLiPC beads. In panel (b), we see that the tail beads of POPC are

more ordered than of DOPC. Between POPE and POPC in panel (c), the profiles are very similar and the difference only arises in the head group region. The phosphate-choline bond (bond number 1) is slightly less ordered in case of PE than in PC. Both the results from bilayer thickness and order parameter suggest that lipid patches in BBS after equilibrium indeed behave as pure bilayers.

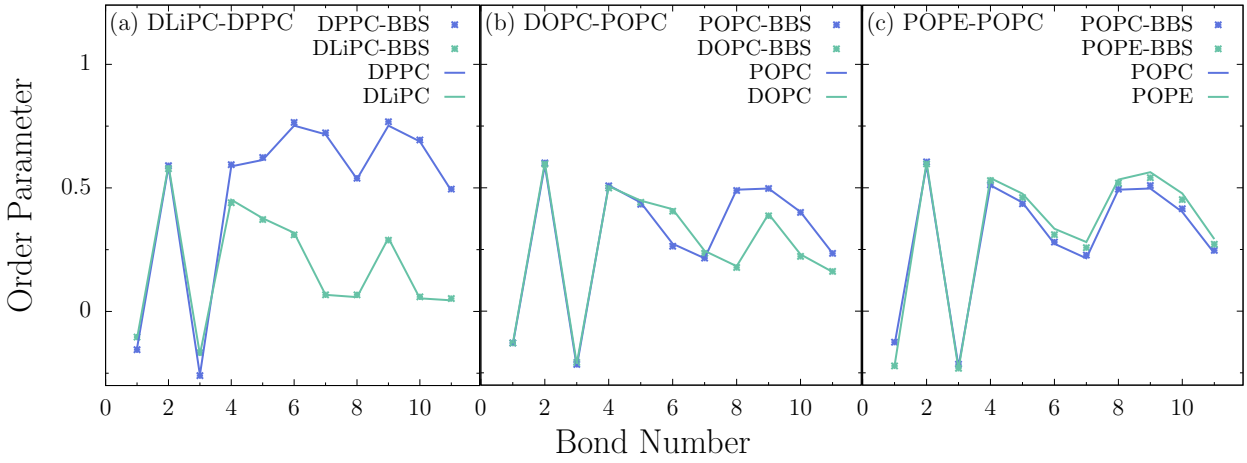


Figure 3.9: Lipid order parameter of consecutive bonds with respect to the bilayer normal in the BBS and in pure binary bilayers. The order parameter for each lipid stays the same for each lipid in both BBS and in individual binary setup.

Partition Free Energy

Thus far we have verified that the BBS is a reliable simulation method and that any thermodynamic properties inferred from BBS will correspond to pure lipid bilayers. The partition coefficient of cholesterol (K), a quantity that is determined in experiments, can be calculated using the density of cholesterol in two patches within the BBS. The partition coefficient (K) is simply the ratio of cholesterol density in two patches, which can be used to calculate the partition free energy ($\Delta G_{partition}$), using $\Delta G_{partition} = -k_B T \ln(K)$, where k_B is the Boltzmann constant, T is the temperature, and K is the cholesterol partition coefficient. The values of

	Mole Fraction	Sim. K	Exp. K	Sim. $\Delta G_{partition}$ (kJ/mol)	Exp. $\Delta G_{partition}$ (kJ/mol)[30]
DLiPC-DPPC	.13 .42	5.4 ± 0.13		-4.39 ± 0.05	
DOPC-POPC	.20 .20	1 ± 0.08	1.4 ± 0.11	0 ± 0.2	-0.5 ± 0.02
POPE-POPC	.20 .20	1 ± 0.02	0.33± 0.04	0 ± 0.05	-0.13 ± 0.004

Table 3.2: Estimated partition coefficient and free energy from simulated BBS. The experimental and simulation values do not match due to the magnitude of the uncertainty.

$\Delta G_{partition}$ are summarized in Table 3.2. From the table, we can see that the results from simulation do not match the experiments mainly because of the magnitude of the uncertainties. The experimental $\Delta G_{partition}$ values are all less than 1 kJ/mol and our simulations can not predict such results accurately. However, our results show that BBS can predict larger $\Delta G_{partition}$, as in the case of DLiPC-DPPC. Unfortunately, DLiPC is not widely used in experiments and one would need more isothermal calorimetry results to validate our finding for DLiPC-DPPC BBS as done in Ref[32,33] .

Although comparison of results between BBS and experiments is not straight forward, we can compare the results from BBS to our REUS results. In table 3.3, we list the chemical potential values for pure bilayer with different cholesterol concentration corresponding to lipid patches in BBS. The results from DLiPC-DPPC would suggest that, a DPPC bilayer with 42% cholesterol will be in equilibrium with 13% cholesterol and the chemical potential of any cholesterol molecule would be the same in these two pure bilayer. In order to verify this, we run two new REUS calculations for pure bilayers with 42% cholesterol concentration in DPPC and 13% cholesterol concentration in DLiPC bilayer. From the table, we see that this is actually the case.

3.4 Discussion

In this work, we have summarized our result from coarse grained molecular dynamics simulations. By using two distinct simulation techniques, we have described the thermodynamics

Lipid concentration	μC (kJ/mol) from REUS
13%Cholesterol-DLiPC	-88.8 ± 2.1
42%Cholesterol-DPPC	-89.6 ± 1.8
20%Cholesterol-DOPC	-94.8 ± 1.6
20%Cholesterol-POPC	-94.9 ± 2.6
20%Cholesterol-POPE	-96.5 ± 2.7

Table 3.3: Cholesterol chemical potential for pure bilayers. The REUS and BBS results are consistent within error bars. BBS is computationally much cheaper than REUS.

of cholesterol in lipid bilayers. We systematically calculated cholesterol depletion energies for 49 lipid bilayer systems with varying cholesterol concentration. From our results, we have shown that cholesterol has greater affinity for lipid bilayers with i) less cholesterol content and ii) less degree of unsaturation. Between PC and PE, cholesterol does not prefertially partition into either; this result is not consistent with previous experimental results.[30,103,104] However, our depletion energy results in the range of 100 kJ/mol - 75 kJ/mol is in allignment with previous MARTINI simulations results.[31] In addition, our estimate of chemical potential is in excellent agreement with previous MARTINI results.[99] Therefore, this discrepancy against experiments should be due to the inaccuracy of the forcefields; further calculations with more accurate forcefields may verify this. Within a binary bilayer with phospholipids and cholesterol only, we have shown that the chemical potential gradually increases as the concentration of cholesterol increases. This result shows that neither of the models discussed in the introduction accurately describes the cholesterol-lipid interactions. The cholesterol-lipid interactions depends on both the headgroup and the tail unsaturation.

Finally, we have developed a BBS method that allows us to eliminate the usage of water as the reference state, which makes the calculation of partition free energy much faster and computationally cheaper. BBS allows us to answer a more biologically relevant question. One can calculate the free energy required for cholesterol molecule to move between two bilayers. Using the DLiPC-DPPC system, we have shown the ability of BBS to easily calculate the cholesterol partition coefficient. Our result show that cholesterol can differentiate between

PC lipids that have very different degrees of unsaturation. These results open the way to a detailed, thermodynamically based description of lipid-cholesterol mixtures, which is much needed to enhance our understanding of bio-membrane related processes.

Chapter 4

PHOSPHOLIPID STRUCTURE AND THERMODYNAMICS ON A MONTMORILLONITE CLAY SURFACE

4.1 Introduction

Soil contains a large, but poorly understood, carbon pool. Soil carbon impacts a range of soil properties, including that of soil water repellency (SWR) .[36,37] SWR is a hydrophobic process where water can not infiltrate the soil. This results in increased erosion, increased water overland flow, and decreased ability to support crops.[34,35] SWR is common in sandy soil and adding clay in sandy soil can alleviate water repellence. Inspite of these environmental and agricultural impacts, there is a lack of mechanistic understanding of SWR, which limits the design of effective remediation protocols.[107] There is an abundance of SWR field studies and investigations of natural organic matter that has not been molecularly characterized.[107,108] Only limited work has been done on model surfaces with well characterized compounds, with most focused on fatty acids .[109–111] Prior work with specific compounds has typically been done on quartz and not on clays, even though clay has been used to remediate SWR and the clay fraction of soils has been found to be a key factor governing the rate of soil organic matter decomposition.[111] The results of these prior studies show that saturated fatty acids induce water repellency in sand, which tends to plateau after a critical amount has been added to the sand.[110,112] As clays are mixed into hydrophobic sandy soils to ameliorate water repellency by an unknown mechanism ,[38] it is desirable to know more about clay-organic matter interaction. Generally, there is a need for more work that both measures water repellency on clay systems and investigates microscale and macroscale characteristics, as prior work frequently does one or the other, but rarely both.

Lipids have been selected for this project, because prior studies have determined that the

lipid fraction of soil organic matter impacts soil hydrophobicity.[113] Lipids are particularly important when considering seasonal cycling of SWR, as lipids have the ability to change orientation, exposing either their hydrophilic or hydrophobic moieties, depending on ambient conditions.[39] While the chemistry of lipids may influence SWR, the surface of the clay itself is also known to influence hydrophobicity.[114,115] Therefore, the goal of this study is to advance a mechanistic understanding of SWR by considering the chemistry of organic matter and a model clay surface. Specifically, we investigate a system of the clay montmorillonite (MMT) and 4 lipids with similar structures. MMT is the most widespread phyllosilicate clay mineral; phyllosilicates have two-dimensional sheets of silica in their structure. MMT is a multi-layered mineral where each layer is a "sandwich" of silica/alumina/silica. Isomorphic substitutions in the the layers give rise to a negative clay layer charge that is compensated by the presence of cations in the interlayer space and on the surface of MMT particles.[116] A MMT unit cell is shown in Fig 4.1.

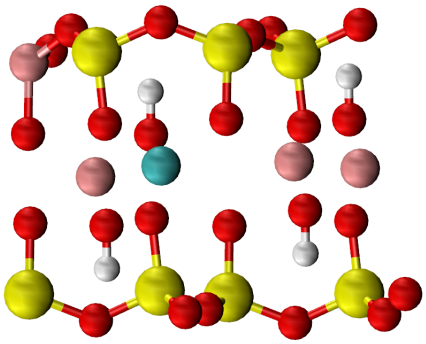


Figure 4.1: The unit cell contains silicon (yellow), oxygen(red), aluminum(pink), magnesium (blue), hydrogen(white) atoms. There are substitutions at both silica and alumina layers.

The application of molecular modeling in geochemistry has significantly increased in the recent years .[117–121] Both ab-initio and classical simulations have been used to study properties of clay pertaining to swelling, cation exchange and interlayer properties.[122–126] Simulations can provide atomic scale descriptions of organic matter interactions with the

surface and also aid in quantifying these interactions. Using biased and unbiased molecular dynamics simulation, we seek to understand the structural properties of surface-bound organic molecules. We find that the lipid aggregates with zwitterionic headgroup form micellular structures of different sizes and bind strongly to the surface of MMT. The results will contribute to our overarching goal of understanding water repellency.

4.2 Methods

4.2.1 Experimental Details

All the experimental work was done by Brenda Kessenich and the procedure will not be discussed here. Particularly, measurements and data analysis related to film topography that is not pertinent to the molecular dynamics study will not be a part of this thesis. However, relevant results from our collaborative study will be discussed. Experimental details can be found in Ref.[127]

4.2.2 Simulation Details

We performed all simulation using Gromacs 5.1.4[55] with the Plumed 2.3[56] plugin under periodic boundary conditions. Temperature and pressure were maintained using the velocity-rescale thermostat and Parrinello-Rahman barostat, respectively. Long range electrostatic interactions were computed using the PME[57] method with a Fourier spacing of 0.12 nm. The real space Coulombic interaction was calculated up to 1.2 nm. Van der Waals interactions were calculated using a cutoff of 1.2 nm. Bond lengths were constrained using the LINCS algorithm.[58] Initial configurations were generated by inserting lipid molecules into previously equilibrated clay/water systems at random positions. These configurations were energy minimized and then equilibrated for 10 ns. After the system was equilibrated, production runs were run for 500 ns.

The clay model is based on the pyrophyllite unit cell structure, and all simulations contain a total of 96 unit cells. Four unit cells within the simulation have a cell formula

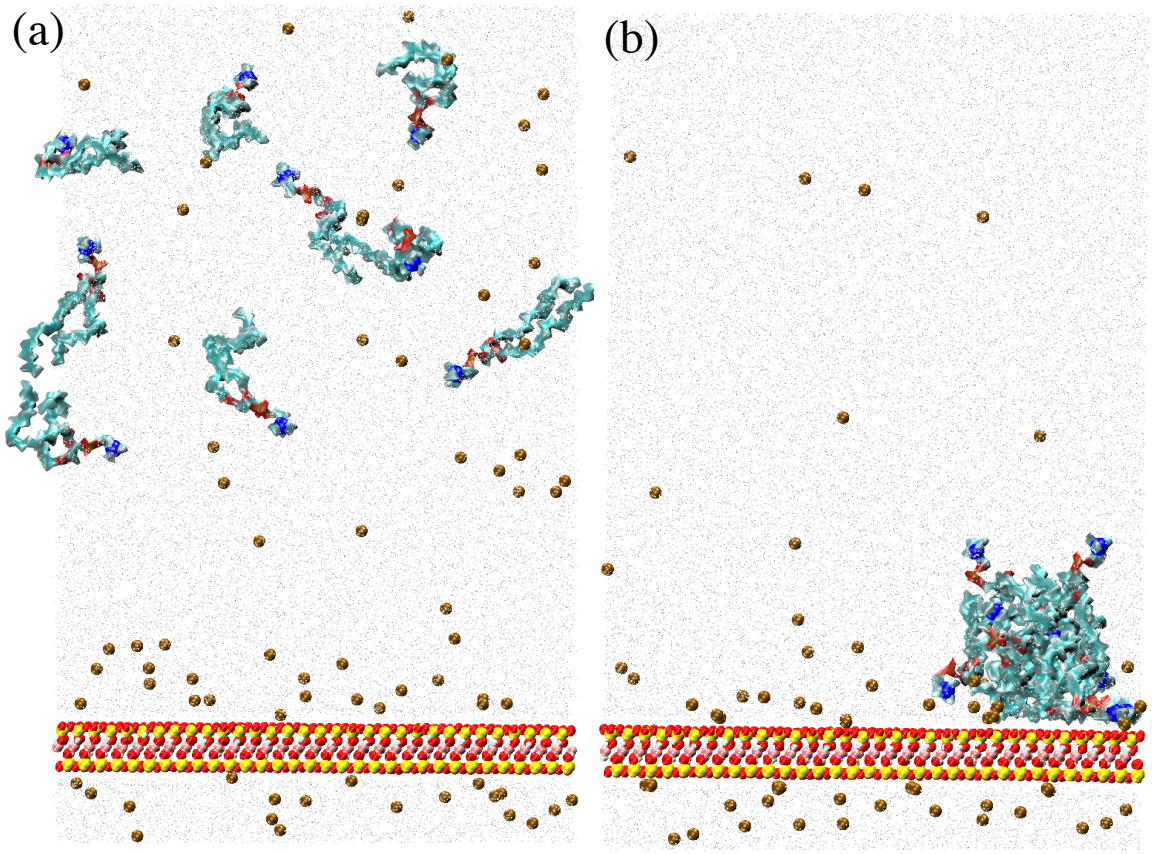


Figure 4.2: Standard simulation snapshot. Shown here is the simulation result of 10% CAM DOPC on MMT surface. Lipids are inserted randomly in water (a). All lipids form aggregates to minimize tail exposure to water. The aggregates bind to the surface on the MMT b) with their headgroups facing the surface. The lipids are represented in cyan, nitrogen in the headgroup is represented in blue, and oxygen in red. The color scheme for MMT is defined in Figure 4.1. Sodium ions are represented in tan and water is represented as faded white dots.

$(Si_{31}Al)(Al_{14}Mg_2)O_{80}(OH)_{16}.nH_2O$, as proposed by Cygan.[128] We employ the CLAYFF force field, which consists of nonbonded (electrostatic and van der Waals) terms and accurately represents the local charge inhomogeneities formed in the clay. The lipids were modeled by the CHARMM36[129] forcefield and water was modeled using the TIP3P[130] forcefield at

	Lipid	Melting Temp. (°K)	Number of Double Bonds	Charge	State at Room Temperature
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine	257	2	±	liquid
DSPE	1,2-distearoyl-sn-glycero-3-phosphoethanolamine	347	0	±	solid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine	256	2	±	liquid
DSPG	1,2-distearoyl-sn-glycero-3-phosphoglycerol	328	0	-	solid

Table 4.1: We study four lipids with different headgroups and fatty acid tails. We simulate four systems for each of the lipids listed at room temperature. Additionally, we simulate four more systems for DSPE at 353°K.

the desired temperatures. In each simulation box, the resulting net charge was balanced by adding interlayer Na^+ ions, forming Na-MMT.

We first generate a slab of the clay. We then construct four systems for each of the four lipids by inserting lipid molecules into already equilibrated clay/water systems at random positions (Fig 4.2(a)). All these configurations were energy minimized and then equilibrated for 10 ns. After the system was equilibrated, unbiased production runs were performed for 1μs and biased runs were performed for 500 ns. The lipids we model are listed in table 4.1. For each lipid we simulated a system with just one lipid and the clay slab to determine the orientation of the lipid when on the surface of MMT. We additionally simulated 10%Coverage as Monolayer (CAM), 25%CAM, and 50%CAM for all lipids at room temperature and also at 353°K for DSPE. DSPE at room temperature is solid while at 353°K, it is liquid; the high temperature simulation will allow us to investigate if the phase of the lipid impacts the structure of surface-bound lipids. The %CAM and hence the number of lipids to simulate was determined by the same protocol as in our experiments:

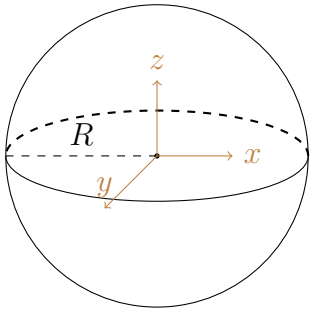
$$\%CAM = \frac{\textit{Total area of lipid based on area per molecule}}{\textit{Surface area of montmorillonite}} \tag{4.1}$$

where the area of each lipid molecule was estimated to be 60 ^{2}.

4.2.3 Determination of aggregate structure

From the simulation, our primary goal is to understand the structure of surface-bound aggregates. In all unbiased simulations for all lipids, freely floating individual lipid molecules come together and form aggregates. These aggregates then either move to the surface of MMT or remain in water depending on the type of the lipid. The shapes of lipid aggregates were inferred from 1-dimensional (1-d) density profiles of the headgroup nitrogen or phosphorous (for DSPG) atom and the last carbon atom in each tail. We used the `gmx_density` tool of Gromacs to compute these profiles. This protocol has been used by a previous simulation study of organic matter on mica surface to study aggregate reorientation.[131]

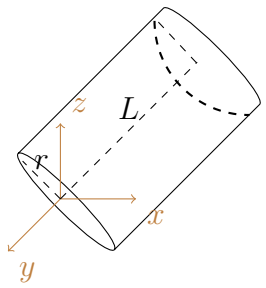
Here, we derive 1-d density equations to distinguish between a spherical and a cylindrical micelle. The aggregates formed in our simulations are not perfect spheres or cylinders; however, equation 4.2 and 4.3 will help us make estimates of the structure of the surface-bound lipid aggregates. For a sphere of radius R, the density profile is constant and does not depend on the axis of projection (z in the following derivation).



The average area density $(\rho) = \frac{N}{A}$, where N is the number of molecules and area (A) = $4\pi R^2$. Then the volume density $(\psi(\vec{r})) = \rho\delta(r - R)$. The volume density projected in 1-d,

$$\begin{aligned} f(z) &= \int dy \int dx\, \psi(\vec{r}) = \int dy \int dx\, \rho\delta(r - R) \\ &= \rho \int dy \int dx\, \delta(\sqrt{x^2 + y^2 + z^2} - R) \\ &= \rho \int dy\, \frac{2R}{\sqrt{R^2 - y^2 - z^2}} = 2\pi R\rho = \frac{N}{2R} \end{aligned} \tag{4.2}$$

The density of cylinder depends on z, and has two sharp peaks. For a cylinder of height L and radius R, the average area density $(\rho) = \frac{N}{A}$, where N is the number of molecules and area (A) $= 2\pi RL$.



Then the volume density $(\psi(\vec{r})) = \rho\delta(r - R)$, where $r = x^2 + z^2$. The volume density projected in z axis,

$$f(z) = \int dx \int_0^L dy\, \psi(\vec{r}) = \int dz \int_0^L dy\, \rho\delta(r - R)$$

$$= \rho \int dx \int_0^L dy\, \delta(\sqrt{x^2 + z^2} - R)$$

$$= L\rho \int dx\, \delta(\sqrt{x^2 + z^2} - R) = L\rho \frac{2R}{\sqrt{R^2 - z^2}} = \frac{N}{\pi R} \frac{1}{\sqrt{1 - (z/R)^2}} \tag{4.3}$$

.

4.2.4 Determination of binding energies

We calculate the binding energy of lipids (both single and aggregates) onto the surface of MMT by using replica exchange umbrella sampling as in the previous chapters. For a single lipid molecule binding to the clay surface, we simulated 35 windows separated by 0.1nm. The collective variable we use is the normal distance between the nitrogen atom (or phosphate for DSPG) and the surface of the MMT. Each window was restrained using a harmonic potential of 500 (kJ/mol)/nm^2, and exchange attempts with neighboring replicas occured every 2 ps. To determine the binding energy of aggregates, we used a slightly different protocol. Starting with aggregates that were formed in unbiased simulation, we bias the

distance between the center of mass of the aggregate and the clay surface using a harmonic potential of 500 (kJ/mol)/nm^2, and exchange configurations with neighboring replicas every 2 ps. We simulated 28 replicas, each for 500 ns. To finally generate the free energy profile, the results from biased replicas, were combined using the weighted histogram analysis method. The statistical uncertainties were estimated using bootstrapping technique implemented in `gmx wham`.[71]

4.3 Results

Our experimental results show that the static contact angle does not change when lipids are added to MMT.[127] This result is in contradiction with previous experimental results that showed significant changes in contact angle after addition of stearic acid (organic matter) on quartz sand .[111] The authors attributed this to the organic matter forming monoloyers on the surface, exposing their hydrocarbon tails outward. The hydrophobic tails then prevented the water droplet from infiltrating the surface which resulted in an increase in contact angle. We investigate if similar phospholipid monolayers are formed on the surface of MMT. We find that lipid aggregates do not form monolayers and instead form hydrophilic aggregates on the MMT and therefore do not increase the hydrophobicity of the clay. All our simulation results are summarized below.

4.3.1 Aggregate Structures

We plot the density of the headgroup nitrogen atom (or phosphorous in the case of DSPG) and the tail carbon atom in both lipid tails of the aggregates as a function of position along the normal axis with respect to the surface of MMT in Fig. 4.3. In all 5 panels, the density profile for 10% CAM is missing because it was not possible to gather much structural information. First, let us look at panels (a)-(d). The density profiles of the nitrogen atoms in the lipid headgroups are asymmetric in all cases. All the profiles have a peak that is close to the surface of MMT. This shows that there are more lipids that are bound to the surface than lipids that are not. For DOPE in panel (a), the nitrogen density at both 25% and 50%

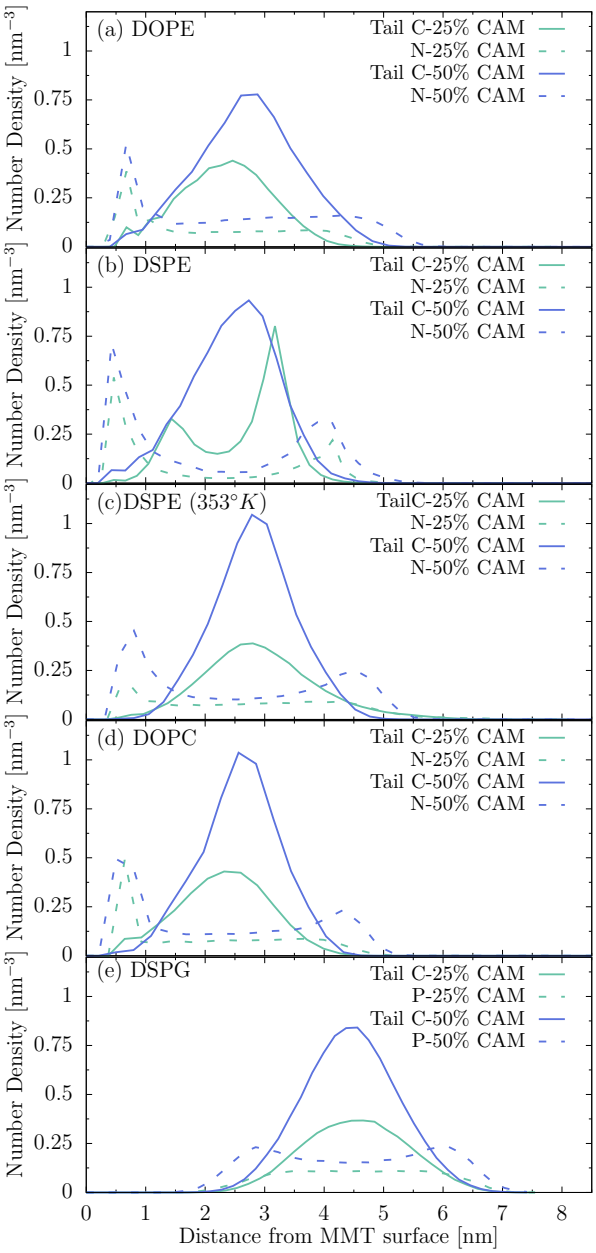


Figure 4.3: Number density of headgroup nitrogen or phosphorous and tail carbon. Other than DSPG, all other aggregates adsorb on the surface of MMT as suggested by the peak in nitrogen density very close to the surface. Results for 10% CAM are not shown here because it was not possible to infer any structural information from the density profiles.

coverage falls drastically with distance, but does not completely vanish, becoming constant beyond 1 nm. The carbon peak for both aggregates is unimodal. From equation 4.2, we can deduce the formation of a spherical micelle in both cases. For DSPE at room temperature (panel (b)), however, both nitrogen and carbon profiles at 25% coverage exhibit bimodal distributions with negligible nitrogen density between the peaks. For 50% CAM DSPE, the nitrogen profile is bimodal, but that of carbon is unimodal. These distinct profiles correspond to interdigitated lipid bilayer arrangements and regular bilayer arrangement for 25% CAM and 50% CAM, respectively. In the heated state (panel (c)), DSPE at 25% CAM behaves similarly to DOPE (both 25% and 50%)and forms a spherical micelle. When the lipid concentration increases to 50% CAM, the nitrogen density profile becomes bimodal with significant density between the peaks. This structure corresponds to a cylindreical micelle from equation 4.3. DOPC (panel (d)) behaves very similarly as heated DSPE (panel (c)). At 25% CAM, spherical micelles are formed and at higher coverage, cylinders are formed. Therefore, in the case where the lipid headgroup is zwitterionic (panel (a)-(d)), the aggregates bind to the surface via the lipid headgroup, and the lipids arrange themselves in such a way that the fatty acid tails are protected from being exposed to water.

DSPG lipids, which have negatively charged headgroups, however, do not bind to the surface. The MMT surface is also negatively charged, so the aggregates are repelled from it, and freely float in water instead. The headgroups orient towards the water, forming perfect spherical micelles at 25%CAM or cylindrical micelles at 50%CAM. We infer this from symmetric nature of all four profiles in Fig. 4.3(e). The tail density profiles for both %CAM is unimodal. The phosphate density for 25% CAM does not have any peak, and that of 50% is bimodal.

The structure of the lipid aggregates does not seem to be dependent on the phase of the lipid at the experimental temperature. At the simulated temperature, DOPE, DSPE(353 °K), and DOPC are liquid; we also simulated DSPE and DSPG that are solid at room temperature. We do not find a correlation between the structure of surface-bound aggregates and the phase of the lipid. DSPE (solid), however, behaves very differently, and forms bilayers

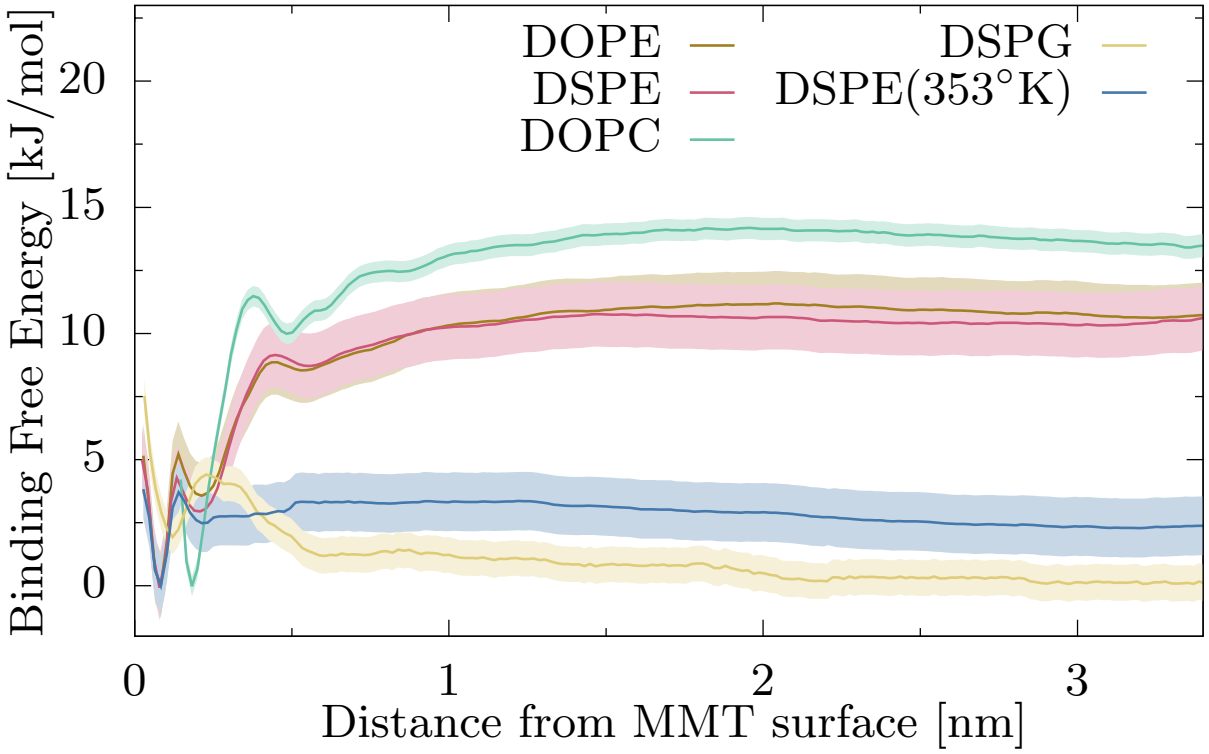


Figure 4.4: Free energy of lipid binding to MMT surface. Lipids with zwitterionic headgroup (PE and PC) bind favorably with MMT. DSPG prefers to freely float in bulk water. The shaded areas represent uncertainty.

while other lipids form micelles. This could be because the solid saturated tails become much more ordered than liquid unsaturated tails when the headgroups bind to the clay surface.

4.3.2 Binding Energies

In Fig. 4.4, we plot binding energies for a single lipid as a function of the nitrogen distance from the surface of the clay (phosphorous in the case of DSPG). All the profiles are shifted vertically such that their minimum lies at 0. From the figure we see that lipids with zwitterionic headgroups (PE and PC) have their minimum on the surface of MMT showing that they bind favorably with the surface. The minimum of all PE profiles lie at the exact same position, but that of DOPC is displaced because of the size of the headgroups; PE is smaller

than PC. The free energy profile of DOPE and DSPE at room temperature is very similar, suggesting the difference in their tail unsaturation does not affect single lipid binding. The energy gain due to binding for both DOPE and DSPE is 10 kJ/mol. For DSPE at elevated temperature, the free energy gain is about 2.5 kJ/mol, which is much less than that at room temperature. At elevated temperature, the entropic contribution (due to fluctuation) to the free energy becomes comparable to the enthalpic contribution (due to binding) resulting in a smaller gain in free energy. Although the shape of the DOPC profile may look similar to that of DOPE, the free energy gain due to binding is higher in the case of DOPC. There is a sharp increase in the free energy profile for DOPC at the surface when the hydrophobic choline headgroup is exposed to water. DSPG on the other hand does not bind to the surface and shows a maximum in the free energy profile at the surface. A single DSPG molecule prefers bulk water by 2 kJ/mol over the MMT surface.

Lipid aggregates may behave differently than single lipid molecules. In order to quantify how efficiently these aggregates bind to the clay, we calculate the binding energy of these aggregates as a function of the normal distance between the center of mass of the aggregates from the surface of MMT. In Fig. 4.5, we show the results for the studied lipids. We do not calculate the binding energy for DSPG because the results from single lipid binding show that they prefer water more than the MMT surface. Our results show that the binding becomes more favorable with increasing lipid concentration for all lipids at all temperatures. The gain in energy for a single DOPE molecule is 10 kJ/mol. This energy increases to 12 kJ/mol when the coverage is 10%. Similarly, the binding energy gain goes up to 20 kJ/mol and 25 kJ/mol when the coverage increases to 25% and 50%, respectively. As the % CAM increases, bigger aggregates, as predicted from the position of the minimum, with more headgroups that bind to the clay surface are formed. The increase in free energy gain, however, is not linear. The free energy gain for DSPE at room temperature goes from 10 kJ/mol (single lipid) to 30 kJ/mol (50%), and for DOPC goes from 12 kJ/mol to 27 kJ/mol The free energy gain for high temperature DSPE at lower % CAM is relatively lower than other aggregates at low %CAM due to increased entropic contribution. However, at 50%CAM the binding energy

increases to 25 kJ/mol like for other lipids. One possible explanation for this observation is an increase in enthalpy due to higher nitrogen contacts with the clay at 50% CAM than at lower concentrations.

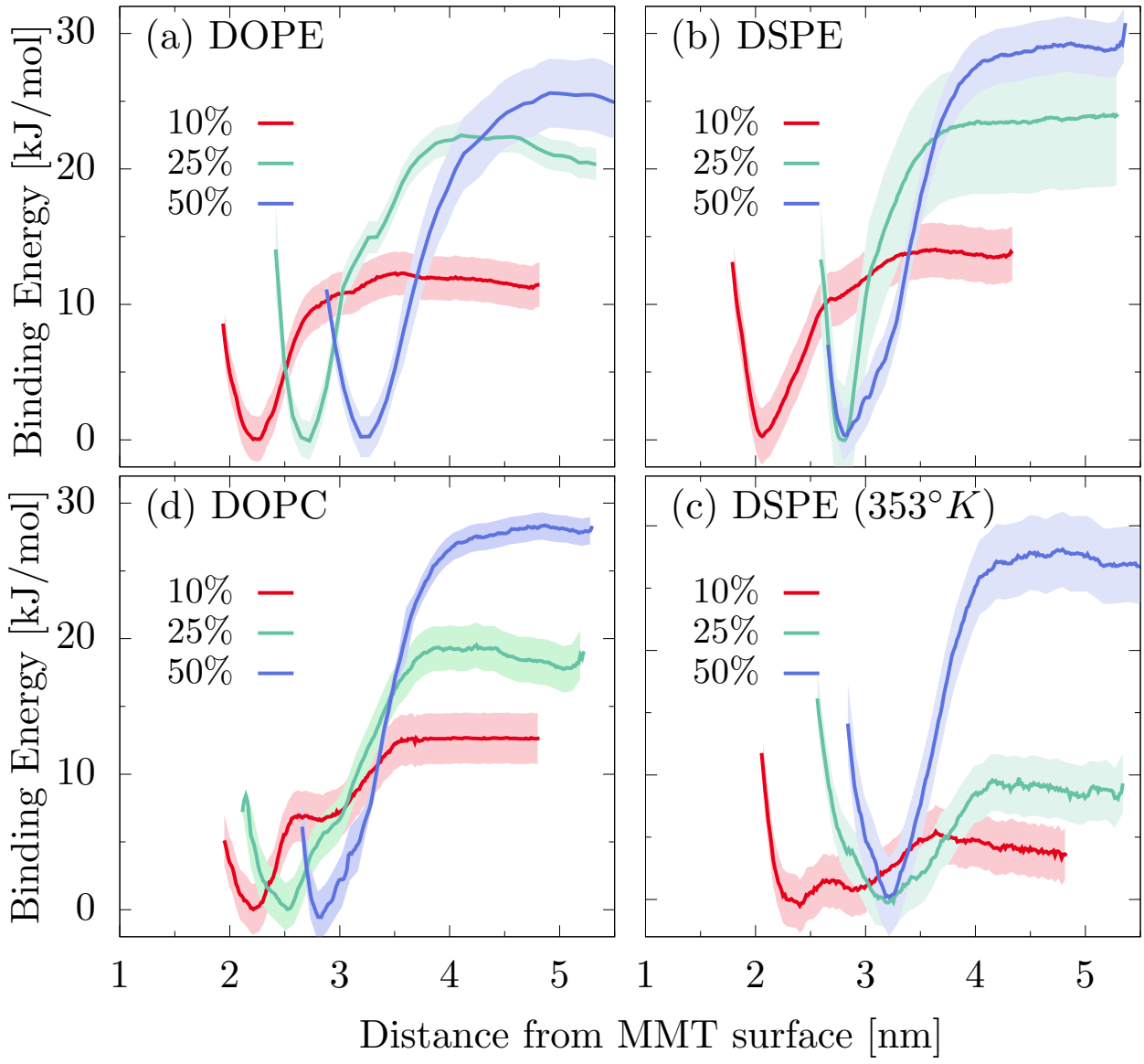


Figure 4.5: Free energy of lipid aggregate binding to MMT surface. Larger aggregates bind more favorably to the surface of MMT in all cases. The shaded areas represent uncertainty.

4.4 Discussion

In this chapter we have studied the cause of soil water repellency at the molecular level. We have looked at the phenomenon with MMT and lipids as the clay and organic molecule, respectively. Using molecular dynamics simulations, we have found that individual lipid molecules come together and forms aggregates to prevent the exposure of their hydrophobic tails. These structures, based on their headgroups, either remain in bulk water or bind to the clay. Our structural analysis suggests that only zwitterionic lipid aggregates bind to the surface. These aggregates are not monolayers, but range from spherical to cylindrical micelles and are not water repellent. In addition, we have also computed the free energy difference between the surface bound state and freely floating state for phospholipid aggregates. The free energy gain of aggregate binding ranges from 5 kJ/mol to 30 kJ/mol. The binding increases as the % CAM increases; this increase is not linear. With more lipids, more headgroups benefit from binding to the negatively charged surface.

Chapter 5

CONCLUSION AND FUTURE WORK

In chapter 2, we investigated the mechanism of cell permeation. The nature of membrane permeation is both fundamental and extremely complicated. Because the cell membrane is highly dynamic, it is very challenging to capture membrane deformations and all possible degrees of freedom, and thus to gather statistically meaningful results. Finding the most relevant degree of freedom for a translocation process has been the major focus of the chapter. We were able to identify that the strong electrostatic interactions between zwitterionic lipid head groups and charged or highly polar solutes are the slowest degree of freedom. The system can exhibit very slow relaxation timescales that pose severe problems to the accuracy of free energy profiles. We propose to resolve this issue by using the REUS simulation technique, which is a modified version of the ubiquitously used US method. With this modified method, we were able to accurately calculate free energy profiles of membrane translocation for several small molecules. We showed that not all free energy algorithms can be used to understand the mechanism of membrane permeation, as some approaches fail to sample the transition state of the translocation process. We also show how standard methods to estimate error can underestimate the error in free energy profiles. With the new method we used, transition states are properly sampled which shed light on the mechanism of membrane permeation. At the end of the chapter, we provide ways to assess the sampling of transition state using the committor distribution function which is a novel way of looking at free energy profiles of translocation.

With the ability to accurately compute free energy profiles, in chapter 3, we studied cholesterol activity in lipid species with different head and tail groups. We quantified cholesterol activity by calculating depletion free energies from 49 systematically constructed sys-

tem using REUS. We have also attempted to resolve disagreements about which model of cholesterol-lipid interactions best describe the chemical potential of cholesterol in bilayer systems. Within the resolution of the systems modeled using the MARTINI forcefield, we have shown that none of the three models can capture all the physics of cholesterol-lipid interactions. Simulating many more systems with all atom forcefields will validate our hypothesis; experimentally, more thermodynamic studies for unsaturated lipids will be required. We also developed the BBS, and showed its efficacy in calculating cholesterol partitioning between two bilayers without using water as a reference state. This simulation procedure immensely reduces computational cost and reproduces the result comparable to the more expensive REUS method. This method can be expanded to more complex systems with many more lipid components and possibly other biomolecules.

In chapter 4, we reported our results of an ongoing collaboration with the De Yoreo group at Pacific Northwest National Lab. In the project, we investigated the structure and energetics of lipid aggregates formed on a MMT surface, which is hypothesized to be an important factor that controls the ability of soil to retain water. A speculation in the field was that lipids may form monolayers with their tails exposed to water, which would repel water and decrease soil's affinity to water. We have shown that this speculation is inaccurate, and we have found that lipids easily aggregate on the surface of the clay forming structures that can range form micelles to bilayers but not monolayers. These aggregates bind to clay surfaces and are unlikely to unbind when dried. With our collaborators, we have investigated soil water repellency at various length ranging from the molecular scale to macro-scale and explained why clay is useful in remediating water repellent soil.

In summary, this research has not only provided important insights into the fundamental structure and function of biomolecules, but also resulted in many dexterous simulation protocols. By using statistical mechanics to model cell membrane and topological transitions in the membranes, we have contributed the most to the field of computational biophysics by providing helpful guidance to effectively conduct molecular simulation research. In the ever growing field of molecular simulations, these results can tremendously aid in effective

sampling. This work provides crucial insight into the behavior of lipid molecules when they interact with various solutes (amino acids, cholesterol) and surfaces (clay). Our findings are expected to have impact across many fields and processes, including drug discovery, drug delivery, material design and other related biomolecular applications. We shed some light and pose many intriguing challenges into many long-standing fundamental questions in the field.

BIBLIOGRAPHY

[1] R. J. Clarke. The dipole potential of phospholipid membranes and methods for its detection. *Adv. Colloid Interface Sci.*, 89:263, 2001.

[2] R. V. Swift and R. E. Amaro. Back to the Future: Can Physical Models of Passive Membrane Permeability Help Reduce Drug Candidate Attrition and Move Us Beyond QSPR? *Chem. Biol. Drug Dse.*, 81(1):61, 2013.

[3] A. E. Cardenas, R. Shrestha, L. J. Webb, and R. Elber. Membrane Permeation of a Peptide: It Is Better to be Positive. *J. Phys. Chem. B*, 119(21):6412, 2015.

[4] S. Futaki. Arginine-rich peptides: potential for intracellular delivery of macromolecules and the mystery of the translocation mechanisms. *Int. J. Pharm.*, 245(1-2):1, 2002.

[5] M. Zorko and Ü. Langel. Cell-penetrating peptides: Mechanism and kinetics of cargo delivery. *Advanced Drug Delivery Reviews*, 57(4 SPEC.ISS.):529, 2005.

[6] M. Orsi and J.W. Essex. Passive permeation across lipid bilayers : a literature review. In Mark S P Sansom and Philip C. Biggin, editors, *Molecular Simulations and Biomembranes From Biophysics to Function*, page 76. Royal Society of Chemistry, Cambridge, UK, 2010.

[7] H. D. Herce and A. E. Garcia. Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. USA*, 104(52):20805, 2007.

[8] Y. Hu and S. Patel. Thermodynamics of cell-penetrating HIV1 TAT peptide insertion into PC/PS/CHOL model bilayers through transmembrane pores: the roles of cholesterol and anionic lipids. *Soft Matter*, 12:6716, 2016.

[9] M Lindgren, M. Hällbrink, a Prochiantz, and Ü. Langel. Cell-penetrating peptides. *Trends in Pharmacol. Sci.*, 21(3):99, 2000.

[10] S. Piantavigna, M. E. Abdelhamid, C. Zhao, X. Qu, G. A. McCubbin, B. Graham, L. Spiccia, A. P. O'Mullane, and L. L. Martin. Mechanistic details of the membrane perforation and passive translocation of TAT peptides. *Chempluschem*, 80(1):83, 2014.

[11] S. Deshayes, M. C. Morris, G. Divita, and F. Heitz. Cell-penetrating peptides: tools for intracellular delivery of therapeutics. *Cell. Mol. Life Sci.*, 62(16):1839, 2005.

[12] C. Bechara and S. Sagan. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.*, 587(12):1693, 2013.

[13] F. Milletti. Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discovery Today*, 17(15-16):850, 2012.

[14] S. Yesylevskyy, S. J. Marrink, and A. E. Mark. Alternative mechanisms for the interaction of the cell-penetrating peptides penetratin and the TAT peptide with lipid bilayers. *Biophys. J.*, 97(1):40, 2009.

[15] C. Neale, J. C. Y. Hsu, C. M. Yip, and R. Pomés. Indolicidin binding induces thinning of a lipid bilayer. *Biophys. J.*, 106(8):29, 2014.

[16] W. F. D. Bennett, C. K. Hong, Y. Wang, and D. P. Tieleman. Antimicrobial Peptide Simulations and the Influence of Force Field on the Free Energy for Pore Formation in Lipid Bilayers. *J. Chem. Theory Comput.*, 12:4524, 2016.

[17] G. J. Schütz, M. Sonnleitner, P. Hinterdorfer, and H. Schindler. Single molecule microscopy of biomembranes (review). *Mol. Membr. Biol.*, 17(1):17, 2000.

[18] C. Ciobanasu, J. Peter Siebrasse, and U. Kubitscheck. Cell-penetrating HIV1 TAT peptides can generate pores in model membranes. *Biophys. J.*, 99(1):153, 2010.

[19] E. Awoonor-Williams and C. N. Rowley. Molecular simulation of nonfacilitated membrane permeation. *Biochim. Biophys. Acta, Biomembr.*, 1858(7):1672, 2016.

[20] C. Neale and R. Pomès. Sampling errors in free energy simulations of small molecules in lipid bilayers. *Biochim. Biophys. Acta, Biomembr.*, 1858(10):2539, 2016.

[21] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187, 1977.

[22] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99(20):12562–12566, 2002.

[23] Y. Sugita and Y. Okamoto. Replica exchange molecular dynamics method for protein folding simulation. *Chem. Phys. Lett.*, 314(1):141, 1999.

[24] E. Darve and A. Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115(20):9169, 2001.

[25] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(1935):300, 1935.

[26] T. L. Steck, J. Ye, and Y. Lange. Probing red cell membrane cholesterol movement with cyclodextrin. *Biophys. J.*, 83(4):2118–2125, 2002.

[27] T.P.W. McMullen, R. N.A.H. Lewis, and R. N. McElhaney. Cholesterol-phospholipid interactions, the liquid-ordered phase and lipid rafts in model and biological membranes. *Curr. Opin. Colloid Interface Sci.*, 8(6):459–468, 2004.

[28] J. Pan, T. T. Mills, S. Tristram-Nagle, and J. F. Nagle. Cholesterol perturbs lipid bilayers nonuniversally. *Phys. Rev. Lett.*, 100(19):1–4, 2008.

[29] J. R. Silvius. Role of cholesterol in lipid raft formation: Lessons from lipid model systems. *Biochim. Biophys. Acta - Biomembr.*, 1610(2):174–183, 2003.

[30] S. Niu and B. J. Litman. Determination of Membrane Cholesterol Partition Coefficient Using a Lipid Vesicle–Cyclodextrin Binary System: Effect of Phospholipid Acyl Chain Unsaturation and Headgroup Composition. *Biophys. J.*, 83(6):3408–3415, 2002.

[31] W. F. D. Bennett, J. L. Maccallum, M. J. Hinner, S. J. Marrink, and D. P. Tieleman. Molecular View of Cholesterol Flip-Flop and Chemical Potential in Different Membrane Environments. (10):12714–12720, 2009.

[32] A. Tsamaloukas, H. Szadkowska, and H. Heerklotz. Thermodynamic Comparison of the Interactions of Cholesterol with Unsaturated Phospholipid and Sphingomyelins. *Biophys. J.*, 90(12):10089–10092, 2006.

[33] A. Tsamaloukas, Halina S., P. J. Slotte, and H. Heerklotz. Interactions of cholesterol with lipid membranes and cyclodextrin characterized by calorimetry. *Biophys. J.*, 89(2):1109–1119, 2005.

[34] J. P. Terry and R. A. Shakesby. Soil hydrophobicity effects on rainsplash: Simulated rainfall and photographic evidence. *Earth Surf. Process. Landforms*, 18(6):519–525, 1993.

[35] J.V. Witter, B. P.D. Jungerius, and M.J ten Harkel. Modelling water erosion and the impact of water repellency. *CATENA*, 18, 1991.

[36] J. Lehmann and M. Kleber. The contentious nature of soil organic matter. *Nature*, 528(7580):60–68, 2015.

[37] P. M. King. Comparison of methods for measuring severity of water repellence of sandy soils and assessment of some factors that affect its measurement. *Aust. J. Soil Res.*, 19(3):275–285, 1981.

[38] I. McKissock, E. L. Walker, R. J. Gilkes, and D. J. Carter. The influence of clay type on reduction of water repellency by applied clays: A review of some West Australian work. *J. Hydrol.*, 231-232:323–332, 2000.

[39] S. H. Doerr, R. A. Shakesby, and R. P.D. Walsh. Soil water repellency: Its causes, characteristics and hydro-geomorphological significance. *Earth Sci. Rev.*, 51(1-4):33–65, 2000.

[40] D. Bochicchio, E. Panizon, R. Ferrando, L. Monticelli, and G. Rossi. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *J. Chem. Phys.*, 143(14):144108, 2015.

[41] C. T. Lee, J. Comer, C. Herndon, N. Leung, A. Pavlova, R. V Swift, C. Tung, C. N. Rowley, R. E. Amaro, C. Chipot, Y. Wang, and J. C. Gumbart. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *J. Chem. Inf. Mod.*, 314(4):721, 2016.

[42] C. Neale, C. Madill, S. Rauscher, and R. Pomés. Accelerating convergence in molecular dynamics simulations of solutes in lipid membranes by conducting a random walk along the bilayer normal. *J. Chem. Theory Comput.*, 9(8):3686, 2013.

[43] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2):20603, 2008.

[44] Y. Sugita, A. Kitao, and Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113(15):6042, 2000.

[45] S. Dorairaj and T. W. Allen. On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc. Natl. Acad. Sci. USA*, 104(12):4943, 2007.

[46] J. L. MacCallum, W. F. D. Bennett, and D. P. Tieleman. Transfer of arginine into lipid bilayers is nonadditive. *Biophys. J.*, 101(1):110, 2011.

[47] J. Gumbart, C. Chipot, and K. Schulten. Free-energy cost for translocon-assisted insertion of membrane proteins. *Proc. Natl. Acad. Sci. USA*, 108(9):3596, 2011.

[48] M. A. Wilson and A. Pohorille. Mechanism of Unassisted Ion Transport across Membrane Bilayers. *J. Am. Chem. Soc.*, 118(14):6580, 1996.

[49] N. Sapay, W. F. D. Bennett, and D. P. Tieleman. Thermodynamics of flip-flop and desorption for a systematic series of phosphatidylcholine lipids. *Soft Matter*, 5(5):3295, 2009.

[50] J. Comer, K. Schulten, and C. Chipot. Calculation of Lipid-Bilayer Permeabilities Using an Average Force. *J. Chem. Theory Comput.*, 10:554, 2014.

[51] W. C. Wimley and S. H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.*, 3(10):842, 1996.

[52] A. C. V. Johansson and E. Lindahl. Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proc. Natl. Acad. Sci. USA*, 106(37):15684, 2009.

[53] N. Pokhrel and L. Maibaum. Free Energy Calculations of Membrane Permeation: Challenges Due to Strong Headgroup-Solute Interactions. *J. Chem. Theory Comput.*, 14(3):1762, 2018.

[54] N. Pokhrel and L. Maibaum. Free energy calculation of membrane translocation: What works when, and why? In M.L. Berkowitz, editor, *Biomembrane Simulations: Computational Studies of Biological Membranes*, page 125. CRC Press, Boca Raton, FL, 2019.

[55] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and J. C. Herman. GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26(16):1701, 2005.

[56] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, Paolo Raiteri, Davide D., Fabrizio M., Fabio P., R. A. B., and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comp. Phys. Commun.*, 180(10):1961, 2009.

[57] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089, 1993.

[58] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463, 1997.

[59] O. Berger, O. Edholm, and F. Jähnig. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.*, 72(5):2002, 1997.

[60] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. *Interaction Models for Water in Relation to Protein Hydration*, volume 14, page 331. Springer, Dordrecht, Netherlands, 1981.

[61] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides . *J. Phys. Chem. B*, 105(28):6474, 2001.

[62] A. C. V. Johansson and E. Lindahl. Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins*, 70(2):1332, 2008.

[63] A. C. V. Johansson and E. Lindahl. The role of lipid composition for insertion and stabilization of amino acids in membranes. *J. Chem. Phys.*, 130(18):185101, 2009.

[64] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graphics*, 14(October 1995):33, 1996.

[65] A. Laio and F. L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, 71(12):126601, 2008.

[66] J. F. Dama, M. Parrinello, and G. A. Voth. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.*, 112:240602, 2014.

[67] P. Tiwary and M. Parrinello. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B*, 119(3):736, 2015.

[68] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.*, 13(8):1011, 1992.

[69] B. Roux. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, 91(1–3):275–282, 1995.

[70] K. Hukushima and K. Nemoto. Exchange monte carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.*, 65(6):1604, 1996.

[71] J. S. Hub, B. L. D. Groot, and D. Van Der Spoel. g_wham–A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.*, 6:3713, 2010.

[72] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler. TRANSITION PATH SAMPLING : Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.*, 53(1):291, 2002.

[73] C. Dellago, P. G. Bolhuis, and P. L. Geissler. *Transition Path Sampling*, volume 123, page 1. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

[74] S. J. Marrink and H. J. C. Berendsen. Simulation of water transport through a lipid membrane. *J. Phys. Chem.*, 98(15):415, 1994.

[75] J. A. Marqusee and K. A. Dill. Solute partitioning into chain molecule interphases: Monolayers, bilayer membranes, and micelles. *The J. Chem. Phys.*, 85(1):434, 1986.

[76] K. Fosgerau and T. Hoffmann. Peptide therapeutics: current status and future directions. *Drug Discovery Today*, 20(1):122–128, 2015.

[77] S. Deshayes, M. C. Morris, G. Divita, and F. Heitz. Cell-penetrating peptides: tools for intracellular delivery of therapeutics. *Cell. Mol. Life Sci.*, 62(16):1839–1849, 2005.

[78] W. C. Wimley and S. H. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.*, 3:842–848, 1996.

[79] J. L. MacCallum, W. F. D. Bennett, and D. P. Tieleman. Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys. J.*, 94(9):3393, 2008.

[80] S. Piana and A. Laio. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111(17):4553–4559, 2007.

[81] Z. Ghaemi, M. Minozzi, P. Carloni, and A. Laio. A novel approach to the investigation of passive molecular permeation through lipid bilayers from atomistic simulations. *J. Phys. Chem. B*, 116(29):8714–8721, 2012.

[82] M. J. Hinner, S. J. Marrink, and A. H. D. Vries. Location, Tilt, and Binding : A Molecular Dynamics Study of Voltage-Sensitive Dyes in Biomembranes. *J. Phys. Chem. B*, 113(48):15807, 2009.

[83] R. Sun, J. F. Dama, J. S. Tan, J. P. Rose, and G. A. Voth. Transition-Tempered Metadynamics Is a Promising Tool for Studying the Permeation of Drug-like Molecules through Membranes. *J. Chem. Theory Comput.*, 12(10):5157, 2016.

[84] C. L. Wennberg, D. Van Der Spoel, and J. S. Hub. Assess the nature of cholesterol-lipid interactions through the chemical potential of cholesterol in phosphatidylcholine bilayers. *Biophys. J.*, 104(13):5372–5377, nov 1999.

[85] R. A. Demel, J. W.C.M. Jansen, P. W.M. van Dijck, and L. L.M. van Deenen. The preferential interactions of cholesterol with different classes of phospholipids. *BBA - Biomembr.*, 465(1):1–10, 1977.

[86] W. D. Ehringer, S. R. Wassall, W. Stillwell, and A. C. Dumaual. Cholesterol Condensation of Alpha-Linolenic and Gamma-Linolenic Acid-Containing Phosphatidylcholine Monolayers and Bilayers. *Biophys. J.*, 66(2):A287–A287, 1994.

[87] L. Fugler, S. Clejan, and R. Bittman. Movement of cholesterol between vesicles prepared with different phospholipids or sizes. *J. Biol. Chem.*, 260(7):4098–4102, 1985.

[88] S. Niu, D. C. Mitchell, and B. J. Litman. Manipulation of Cholesterol Levels in Rod Disk Membranes by Methyl-β-cyclodextrin. *J. Biol. Chem.*, 277(23):20139–20145, 2002.

[89] A. Polozova and B. J. Litman. Cholesterol dependent recruitment of di22:6-PC by a G protein-coupled receptor into lateral domains. *Biophys. J.*, 79(5):2632–2643, 2000.

[90] P. W.M. Van Dijck, B. De Kruijff, L. L.M. Van Deenen, J. De Gier, and R. A. Demel. The preference of cholesterol for phosphatidylcholine in mixed phosphatidylcholine-phosphatidylethanolamine bilayers. *BBA - Biomembr.*, 455(2):576–587, 1976.

[91] K. K. Halling, B. Ramstedt, J. H. Nyström, J. P. Slotte, and T. K.M. Nyholm. Cholesterol interactions with fluid-phase phospholipids: Effect on the lateral organization of the bilayer. *Biophys. J.*, 95(8):3861–3871, oct 2008.

[92] J. Huang and G. W. Feigenson. A Microscopic Interaction Model of Maximum Solubility of Cholesterol in Lipid Bilayers. *Biophys. J.*, 76(4):2142–2157, 1999.

[93] P. Somerharju, J. A. Virtanen, K. H. Cheng, and M. Hermansson. The superlattice model of lateral organization of membranes and its implications on membrane lipid homeostasis. *Biochim. Biophys. Acta - Biomembr.*, 1788(1):12–23, 2009.

[94] A. Radhakrishnan and H. M. McConnell. Chemical activity of cholesterol in membranes. *Biochemistry*, 39(28):8119–8124, 2000.

[95] M. R. Ali, K. H. Cheng, and J. Huang. Assess the nature of cholesterol–lipid interactions through the chemical potential of cholesterol in phosphatidylcholine bilayers. *Proc. Natl. Acad. Sci.*, 104(13):5372 LP – 5377, mar 2007.

[96] J. P. Litz, N. Thakkar, T. Portet, and S. L. Keller. Depletion with Cyclodextrin Reveals Two Populations of Cholesterol in Model Lipid Membranes. *Biophys. J.*, 110(3):635–645, 2016.

[97] Z. Zhang, L. Lu, and M. L. Berkowitz. Energetics of cholesterol transfer between lipid bilayers. *J. Phys. Chem. B*, 112(12):3807–3811, 2008.

[98] S. Jo, H. Rui, J. B. Lim, J. B. Klauda, and W. Im. Cholesterol flip-flop: Insights from free energy simulation studies. *J. Phys. Chem. B*, 114(42):13342–13348, 2010.

[99] C. Díaz-Tejada, I. Ariz-Extreme, N. Awasthi, and J. S. Hub. Quantifying Lateral Inhomogeneity of Cholesterol-Containing Membranes. *J. Phys. Chem. Letters*, 6(23):4799, 2015.

[100] S. Park and W. Im. Quantitative Characterization of Cholesterol Partitioning between. *J. Chem. Theory Comput.*, 14:2829–2833, 2018.

[101] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007.

[102] S. Jo, T. Kim, V. G. Iyer, and W. Im. Charmm-gui: A web-based graphical user interface for charmm. *Journal of Computational Chemistry*, 29(11):1859–1865, 2008.

[103] A. Björkbom, T. Róg, K. Kaszuba, Ma. Kurita, S. Yamaguchi, M. Lönnfors, T. K.M. Nyholm, I. Vattulainen, S. Katsumura, and J. P. Slotte. Effect of sphingomyelin headgroup size on molecular properties and interactions with cholesterol. *Biophys. J.*, 99(10):3300–3308, 2010.

[104] O. Engberg, T. Yasuda, V. Hautala, N. Matsumori, T.K.M. Nyholm, M. Murata, and J. P. Slotte. Lipid Interactions and Organization in Complex Bilayer Membranes. *Biophys. J.*, 110(7):1563–1573, 2016.

[105] T. J. Piggot, J. R. Allison, R. B. Sessions, and J. W. Essex. On the Calculation of Acyl Chain Order Parameters from Lipid Simulations. *J. Chem. Theory Comput.*, 13(11):5683–5696, 2017.

[106] S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B*, 108(2):750–760, 2004.

[107] R. Kaltenbach, D. Diehl, and G. E. Schaumann. Links between nanoscale and macroscale surface properties of natural root mucilage studied by atomic force microscopy and contact angle. *J. Colloid Interface Sci.*, 516:446–455, 2018.

[108] V. K. Truong, E. A. Owuor, P. Murugaraj, Russell J. C., and D. E. Mainwaring. Impact of particle nanotopology on water transport through hydrophobic soils. *J. Colloid Interface Sci.*, 460:61–70, 2015.

[109] D. A. L. Leelamanie, J. Karube, and A. Yoshida. Characterizing water repellency indices: Contact angle and water drop penetration time of hydrophobized sand. *Soil Sci. Plant Nutr.*, 54(2):179–187, 2008.

[110] L. Petridis, H. Ambaye, S. Jagadamma, S. Michael Kilbey, B. S. Lokitz, V. Lauter, and M. A. Mayes. Spatial arrangement of organic compounds on a model mineral surface: Implications for soil organic matter stabilization. *Environ. Sci. Technol.*, 48(1):79–84, 2014.

[111] K. Mainwaring, I. L. Hallin, P. Douglas, S. H. Doerr, and C. P. Morley. The role of naturally occurring organic compounds in causing soil water repellency. *Eur. J. Soil Sci.*, 64(5):667–680, 2013.

[112] M. Ma'Shum and V. C. Farmer. Origin and assessment of water repellency of a sandy south australian soil. *Aust. J. Soil Res.*, 23(4):623–626, 1985.

[113] E. de Blas, M. Rodríguez-Alleres, and G. Almendros. Speciation of lipid and humic fractions in soils under pine and eucalyptus forest in northwest Spain and its effect on water repellency. *Geoderma*, 155(3-4):242–248, 2010.

[114] M. Nosonovsky and B. Bhushan. Roughness optimization for biomimetic superhydrophobic surfaces. *Microsyst. Technol.*, 107(10-11):969–979, 2005.

[115] Y. C. Jung and B. Bhushan. Contact angle, adhesion and friction properties of micro- and nanopatterned polymers for superhydrophobicity. *Nanotechnology*, 17(19):4970–4980, 2006.

[116] M. F. Brigatti, E. Galan, and B. K. G. Theng. Structures and mineralogy of clay minerals. In F. Bergaya, B. K.G. Theng, and G. Lagaly, editors, *Handbook of Clay Science*, page 19. Elsevier.

[117] M. Geramian, M. Osacky, D. G. Ivey, Q. Liu, and T. H. Etsell. Effect of Swelling Clay Minerals (Montmorillonite and Illite-Smectite) on Nonaqueous Bitumen Extraction from Alberta Oil Sands. *Energy and Fuels*, 30(10):8083–8090, 2016.

[118] N. T. Skipper, K. Refson, and J.D.C McConnell. Computer simulation of interlayer water in 2:1 clays. *J. Chem. Phys.*, 94(11):7434–7445, 1991.

[119] N. T. Skipper. Computer simulation of aqueous pore fluids in 2:1 clay minerals. *Mineral. Mag.*, 62(5):657–667, 1998.

[120] E. S. Boek, P. V. Coveney, and N. T. Skipper. Monte Carlo Molecular Modeling Studies of Hydrated Li-, Na-, and K-Smectites: Understanding the Role of Potassium as a Clay Swelling Inhibitor. *J. Am. Chem. Soc.*, 117(50):12608–12617, 1995.

[121] E. S. Boek, P. V. Coveney, and N. T. Skipper. Molecular Modeling of Clay Hydration: A Study of Hysteresis Loops in the Swelling Curves of Sodium Montmorillonites. *Langmuir*, 11(12):4629–4631, 1995.

[122] R. T. Cygan, J. A. Greathouse, H. Heinz, and A. G. Kalinichev. Molecular models and simulations of layered materials. *J. Mater. Chem.*, 19(17):2470–2481, 2009.

[123] E. J.M. Hensen and B. Smit. Why clays swell. *J. Phys. Chem. B*, 106(49):12664–12667, 2002.

[124] A. Delville. Structure of Liquids at a Solid Interface: An Application to the Swelling of Clay by Water. *Langmuir*, 8(7):1796–1805, 1992.

[125] A. Delville. Modeling the Clay-Water Interface. *Langmuir*, 7(3):547–555, 1991.

[126] D. A. Young and D. E. Smith. Simulations of Clay Mineral Swelling and Hydration: Dependence upon Interlayer Ion Size and Charge. *J. Phys. Chem. B*, 2000.

[127] B. L. Kessenich, N. Pokhrel, E. Nakouzi, C. J. Newcomb, M. Flury, L. Maibaum, and J J. De Yoreo. Connecting wettability, topography, and chemistry in a simple lipid-montmorillonite system. *J. Colloid Interface Sci.*, 2019.

[128] R. T. Cygan, J. Liang, and A G Kalinichev. Molecular Models of Hydroxide , Oxyhydroxide , and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B*, pages 1255–1266, 2004.

[129] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles,. *J. Chem. Theory Comput.*, (9):3257–3273.2012.

[130] W. L. Jorgensen, J. Chandrasekhar, and J.D. Madura. Comparison of simple potential functions for simulating liquid water Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 926(1983), 1983.

[131] J. Das, C. Eun, S. Perkin, and M. L. Berkowitz. Restructuring of hydrophobic surfaces created by surfactant adsorption to mica surfaces. *Langmuir*, 27(19):11737–11741, 2011.