ABSTRACT

CURTIS, EMILY MARIE. A Novel, Systematic Multiscale Modeling Method to Calculate Coarse-Grained Parameters for the Simulation of Biomolecules. (Under the direction of Carol K. Hall.)

We developed new intermediate resolution implicit solvent models for lipids, "LIME" and DNA molecules, "DIME," designed for use with discontinuous molecular dynamics (DMD) simulations. A multi-scale modeling approach was used to extract both the LIME and DIME parameters from explicit solvent atomistic simulations. We applied LIME to study the spontaneous formation of lipid bilayers, the behavior of mixed lipid systems at different pH values and the interaction between membranes and nanoparticles. DIME was used to investigate the structural properties of DNA and the process by which two DNA strands hybridize in solution.

In LIME, 14 coarse-grained sites that are classified as 1 of 6 types represent DPPC. DMD simulations performed on a random solution of DPPC lipids resulted in the spontaneous formation of a defect free bilayer in less than 4 hours. The speed at which the formation of the bilayer was observed is close to an order of magnitude faster than the fastest reported speed for a coarse-grained, implicit solvent model. The bilayer formed quantitatively reproduces the main structural properties (e.g. area per lipid, bilayer thickness, bond order parameters) that are observed experimentally. In addition, the bilayer transitions from a liquid-crystalline phase to a tilted gel phase when the temperature is reduced. Transbilayer movement of a lipid from the bottom leaflet to the top leaflet is observed when the temperature is increased.

Our initial LIME model was extended to include the description of the geometry and energetics of DPPC, 1,2-distearoyl-*sn*-glycero-3-phospho-L-serine (DSPS) and 1,2-dihenarachidoyl-sn-glycero-3-phosphocholine (21PC) at both neutral and low pH at 310K. In the model, 14 coarse-grained sites represent DPPC, 17 coarse-grained sites represent DSPS and 18 coarse-grained sites represent 21PC. Each of these coarse-grained sites is classified as 1 of 10 types. LIME/DMD simulations performed on bilayers containing different compositions of DPPC/DSPS and 21PC/DSPS showed similar heterogeneous domain formation at both a neutral and low pH.

We demonstrate how the combination of DMD and LIME can be used to model the interaction between lipid membranes and nanoparticles of different sizes, densities and hydrophobicities. In this work we run "proof of concept " simulations to demonstrate that our model can be evolved to examine more specific nanoparticle-membrane systems. We studied the wrapping process for nanoparticles with diameters from 5Å to 100Å and found that DPPC bilayers do not wrap nanoparticles with a diameter less than 20Å. Instead, we found that these particles become embedded in the bilayer surface where they can easily interact with the hydrophilic head groups of the lipid molecules. We also investigated the interaction between hydrophobic nanoparticles with diameters from 5Å to 40Å. According to our results, the hydrophobic nanoparticles do not undergo the wrapping process; instead they directly penetrate the membrane and embed themselves within the inner hydrophobic core of the bilayers. The density of the hydrophilic and hydrophobic nanoparticles did not appear to affect the way in which they interact with the membranes.

In DIME, three coarse-grained sites are used to represent each nucleotide (one for each sugar, phosphate and base molecule). Each of these coarse-grained sites is classified as 1 of 6 types for sugar, phosphate, cytosine, guanine, adenine and thymine. DMD simulations performed on an initial random configuration of two single-stranded Dickerson-Drew dodecamer chains resulted in the formation of a double-helical structure within approximately 0.17 CPU hours. An alternative procedure for calculating the square-well width for each pair of interaction sites, which involves the second virial coefficient, was also investigated. Simulations run using this second set of parameters did not result in the spontaneous formation of a double helix even though the double helix remained stable at low temperature.

A Novel, Systematic Multiscale Modeling Method to Calculate Coarse-Grained Parameters for the Simulation of Biomolecules

by
Emily Marie Curtis

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2013

APPROVED BY:

_____	_____
Carol K. Hall Committee Chair	Erik Santiso
_____	_____
Michael Flickinger	Gregory Reeves

BIOGRAPHY

Emily Marie Curtis was born in 1982 in Manhasset, NY. She is the daughter of William and Nancy Boehler and has younger sister, Sarah. In 2009 she married Douglas Curtis. She attended The Pennsylvania University in University Park, PA and earned her Bachelor of Science in Chemical Engineering in December 2003. During her undergraduate years she participated in co-ops and had internships at several pharmaceutical companies including Pfizer in Groton, CT, Amgen in Thousand Oaks, CA, Johnson and Johnson Pharmaceutical Research and Development L.L.C. in Raritan, NJ and The R.W. Johnson Pharmaceutical Research Institute in Raritan, NJ. After receiving her B.S. degree Emily worked as a Process Engineer at GlaxoSmithKline (GSK) in Research Triangle Park, NC from 2004 to 2007. While at GSK Emily obtained her Masters of Science in Chemical Engineering from North Carolina State University in Raleigh, NC. Upon completion of her M.S. Emily in 2007 decided to return to school full time to pursue her Ph.D. in Chemical and Biomolecular Engineering with Professor Carol Hall. During the pursuit of her doctoral degree she completed a graduate internship with MedImmune in Gaithersburg, MD.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all of the individuals who made it possible for me to complete this dissertation.

First, I would like to express my sincere gratitude to Professor Carol K. Hall for her wisdom, guidance and encouragement. Her work ethic and dedication to research has been an inspiration to me and has helped motivate me whenever I encountered roadblocks. Professor Hall has been an incredible mentor, providing support and advice, both personal and professional, throughout the pursuit of my doctorate. I greatly appreciate the research environment she established for me in which I felt as though my work had a clearly established focus and goal, and yet, I had the freedom to pursue my own ideas.

I would like to thank all of the Hall research group members that I currently work with and that I have worked with in the past. Their friendship and enthusiasm about our work has made our lab environment an extremely enjoyable place to work. I acknowledge them for all of their illuminating discussions and for their assistance with debugging code and improving algorithms. I would like to especially thank the following group members: Dr. Victoria Wagoner, for teaching me the basics about discontinuous molecular dynamics simulations and writing computer code; Lauren Ridge for her invaluable help debugging EMBLEM, the code that was used to run all of the discontinuous molecular dynamics simulations presented in this work; Dr. Amir Bahrami for his collaboration on the bilayer/nanoparticle simulations; and Dr. Abhishek Singh for his collaboration on the DNA simulations.

I would also like to thank all of the system administrators, Lauren Ridge, David Latshaw, David Rutkowski and Gary Gatling, for the work that they have done to provide the Hall Lab with one of the most up to date and reliable computer clusters.

I would like to thank my parents and my sister for their love and encouragement throughout my life. My parents raised me to love math and science and have cheered me on at all stages of my career.

I would like to thank my husband Doug, for all always being there to support me in whatever I do. I cannot possible express how much his love, kindness and friendship mean to me. I would also like to thank Doug for the technical advice he has given me regarding programming languages. Without his input my code would not be nearly as efficient.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Motivation

The technological advances that have been made to date in the many fields of research aimed at exploiting the unique properties of biomolecules and nanoparticles are astonishing. For example, lipid bilayers and liposomes are now being used to create novel devices for the targeted delivery of proteins, nucleic acids and drugs to treat a wide variety of diseases.[1,2,3] In addition, DNA microarrays are used to measure gene expression levels [4,5,6,7] and the potential of DNA to replace silicon transistors as the next generation of storage technology [8] was recently demonstrated. Drugs that function by targeting DNA molecules have also been developed.[9,10,11] The interaction between nanoparticles and biomolecules has also become a popular area of study due to the increasing applications of nanoparticles in drug delivery and the concerns associated with nanoparticle toxicity. In order to fully realize the potential of the numerous emerging biomolecular technologies, a tool that would provide a better understanding of biomolecules and nanoparticles on a molecular level is needed. One approach that is commonly used to gain insight into the behavior of such biomolecular systems on a molecular level is computer simulation.

Molecular dynamics computer simulations can be divided roughly into two categories: high-resolution models and low-resolution models. High-resolution or atomistic models are based on a realistic representation of membrane geometry and energetics and typically account for the motion of every atom on the membrane and every solvent atom.

One drawback associated with atomistic models is that the detail that makes these models so realistic and appealing also makes them extremely computational intensive and prevents them from examining large conformation changers or long time scales. In contrast to high-resolution models, low-resolution models, which are also known as coarse-grained models, are based on a simplified representation of molecular geometry and energetics. In a coarse-grained model a single interaction site is used to represent the behavior of a group of several atoms. This reduces the total number of sites whose trajectories must be calculated, thereby increasing the speed of the simulation.

This thesis describes our work to develop new intermediate resolution implicit solvent models for lipids, "LIME," and DNA molecules ,"DIME,." designed for use with discontinuous molecular dynamics computer simulations We provide a detailed explanation regarding the multi-scale modeling approach that was used to extract both the LIME and DIME parameters from explicit solvent atomistic simulations. We applied LIME to study the spontaneous formation of lipid bilayers, the behavior of mixed lipid systems at different pH values and the interaction between membranes and nanoparticles. DIME was used to investigate the structural properties of DNA and the process by which two DNA strands hybridize in solution. The long term goal of our work is to develop a systematic approach that can be used to gather the geometric and energetic parameters required to run implicit-solvent, coarse-grained simulations with discontinuous molecular dynamics of any system of biomolecules. In addition, we would like to use these models to provide researchers with molecular level insights in order to facilitate the exploration and design of novel biomolecular structures and devices.

1.2 Overview

In this section, we summarize Chapters 2 – 6 of this thesis. Each chapter contains a literature review and a bibliography.

Chapter 2 describes in detail the development of "LIME," which is a new intermediate resolution implicit model for lipid molecules. LIME was designed for use with discontinuous molecular dynamics (DMD) simulations. The model was developed using a multi-scale modeling approach in which the geometric and energetic parameters are obtained by collecting data from atomistic simulations of a system composed of 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) molecules and explicit water. In the model, 14 coarse-grained sites are used to represent DPPC and each of these sites is classified as 1 of 6 types. DMD/LIME simulations performed on a random solution of DPPC resulted in the formation of a defect free bilayer in less than 4 CPU hours. The bilayer formed quantitatively reproduces the main structural properties (e.g. area per lipid, bilayer thickness, bond order parameters) that are observed experimentally. In addition, the bilayer transitions from a liquid-crystalline phase to a tilted gel phase when the temperature is reduced. Transbilayer movement of a lipid from the bottom leaflet to the top leaflet is observed when the temperature is increased.

In Chapter 3 we discuss the expansion of LIME to describe the geometry and energetics of DPPC, 1,2-distearoyl-*sn*-glycero-3-phospho-L-serine (DSPS) and 1,2-dihenarachidoyl-*sn*-glycero-3-phosphocholine (21PC) at both a neutral pH and at a low pH at 310K. A multi-scale modeling approach was used to calculate the LIME parameters from atomistic simulations of a DPPC/DSPS bilayer in explicit solvent at neutral and low pH. In

the model, 14, 17 and 21 coarse-grained sites are used to represent DPPC, DSPS and 21PC, respectively. Each of these coarse-grained sites is classified as 1 of 10 types. LIME/DMD simulations of equimolar bilayers show the following: 21PC/DSPS bilayers separate slightly faster at low pH than at neutral pH, but DPPC/DSPS bilayers separate at approximately the same rate at neutral and low pH, 21PC/DSPS bilayers separate slightly more than DPPC/DSPS bilayers. Our results also show that at low pH equimolar DPPC/DSPS bilayers without surface area restrictions separate faster than those with restrictions but surface area restrictions on equimolar low pH 21PC/DSPS bilayers did not affect the separation rate. Simulations of DPPC/DSPS and 21PC/DSPS bilayers with different molar ratios of PC:PS lipids showed that the higher the concentration of PS lipids, the faster the separation rate. Simulations of DPPC/DSPS and 21PC/DSPS liposomes containing doxorubicin showed domain formation in both types of liposomes. However, no drug molecules escaped from either type of liposome after 1.5 billion collisions.

In Chapter 4 we describe the expansion of LIME to model the interaction between nanoparticles and lipid membranes. In this work we do not model any specific type of nanoparticle; instead we run several "proof of concept" simulations to investigate the interaction between nanoparticles with different physical properties (size, density and hydrophobicity) and lipid bilayers. Our simulations showed that hydrophobic nanoparticles do not undergo the wrapping process. Instead, they directly penetrated the lipid bilayers and remained stable in the hydrophobic core of the membranes. We also found that the wrapping of hydrophilic nanoparticles was size dependent. According to our results, a membrane will not wrap hydrophilic nanoparticles with a diameter less than 20Å. Instead we found that

these nanoparticles become embedded on the surface of the lipid bilayers among the hydrophilic head groups of the lipid molecules. Finally, our simulations showed that the mass per volume of a nanoparticle did not significantly affect its interaction with a lipid bilayer.

Chapter 5 provides a description of initial work performed to develop a new intermediate-resolution implicit-solvent model for DNA molecules, which we call "DIME" for DNA Intermediate Resolution Model. The same multiscale modeling approach used to develop LIME was followed to calculate parameters for DIME. The parameters for this model are obtained by collecting data from an atomistic simulation of a Dickerson dodecamer duplex with explicit solvent and counterions. A single coarse-grained site is used to represent each sugar, phosphate and base molecule in this model. Each of these coarse-grained sites is classified as 1 of 6 types for sugar, phosphate, cytosine, guanine, adenine and thymine. Similar to LIME, DIME was designed for use with discontinuous molecular dynamics simulations. We are able to use this new model to show the spontaneous hybridization that occurs when two Dickerson-Drew dodecamer strands are placed in an initial random configuration. In addition, we discuss the results of alternative method used to calculate the square-well width between coarse-grained sites that involves the second virial coefficient. Simulations run using this second set of parameters did not result in the spontaneous formation of a double helix even though the helix remained stable at low temperatures.

In Chapter 6 we discuss the future work that we hope to accomplish with LIME and DIME.

Chapter 2 is adapted from the following publication:

Curtis, E.; Hall, C. *J. Phys. Chem. B.*, 2013, *117*, 5019-5030.

1.3 References

1. Karve, S.; Bandekar, A.; Ali, M.; Sofou, S. *Biomaterials*. **2010**, *31*, 4409 – 4416.
2. Schroeder, A.; Levins, C.; Cortez, C.; Langer, R.; Anderson, D. *Journal of Internal Medicine*. **2009**, *267*, 9 – 21.
3. Almeida, A.; Souto, E. *Adv. Drug Delivery Rev*. **2007**, *59*, 478-479.
4. Schena, M.; Shalon, D.; Davis, R.; Brown, P. *Science*, 1995, *270*, 467 – 470.
5. DeRisi, J.; Penland, L.; Brown, P.; Bittner, M.; Meltzer, P.; Ray, M.; Chen, Y.; Su, Y.; Trent, J. *Nature Genetics*, 1996, *14*, 457 – 460.
6. Lockhart, D.; Winzeler, E. *Nature*, 2000, *405*, 827 – 836.
7. Chon, H.; Lancaster, J. *Cancer Control*, 2011, *18*, 8 – 15.
8. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.; Sipos, B.; Birney, E. *Nature*, 2013, *494*, 77-80.
9. Hurley, L. *Nature Reviews Cancer*, 2002, *2*, 188 – 200.
10. Palchaudhuri, R.; Hergenrother, P. *Current Opinion in Biotechnology*, 2007, *18*, 497 – 503.
11. Koster, D.; Palle, K.; Bot, E.; Gjornsti, M.; Dekker, N. *Nature*, 2007, *448*, 213 – 217.

CHAPTER 2

Molecular Dynamics Simulations of DPPC Bilayers Using "LIME," a New Coarse-grained Model

2.1 Introduction

The lipid bilayer, the primary constituent of cellular and intracellular membranes in all living organisms, plays a central role in many biological processes including cell signaling and protein function.[(1),(2),(3)] In addition to its physiological significance, lipid bilayers are now being used to create devices for targeted delivery of proteins, nucleic acids and drugs in the treatment of a wide variety of diseases.[(4),(5),(6)] Significant progress has been made by scientists working to use these structures to develop therapeutic agents.[(7),(8),(9),(10),(11),(12)] A tool that would allow researchers to visualize the structure and function of the lipid bilayer on a molecular level could help enhance the rate of advancement in these areas. For example, we are now using the model discussed in this manuscript to study the release of drug molecules from liposomes composed of lipid mixtures as a result of a change in pH. [(13),(14),(15)]

In this paper we take a multiscale modeling approach to develop a new implicit-solvent intermediate-resolution lipid model, "LIME" for Lipid Intermediate Resolution Model, that enables molecular dynamics simulation of the self-assembly of a lipid bilayer. The model system chosen for study is the lipid 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) in water. We show that discontinuous molecular dynamics (DMD) simulations using the LIME forcefield accurately reproduce the structural properties of a DPPC bilayer,

including the area per lipid, bilayer thickness, bond order and mass density profiles. This model is the culmination of a systematic program of research aimed at developing simulation tools based on coarse-grained lipid models that are fast enough to simulate self assembly of large structures yet have accuracy that is comparable to that found in atomistic simulations. By using multiscale modeling, which translates the atomistic details from well established force fields, GROMOS96 53a6 in this case, into coarse grained simulations, we avoid the pitfalls associated with fitting many molecular parameters to the limited data on lipid systems.

Molecular dynamics studies of phospholipid bilayers can be divided roughly into two categories: high-resolution models and low-resolution models. High-resolution or atomistic models are based on a realistic representation of membrane geometry and energetics and typically account for the motion of every atom on the membrane and every solvent atom. Atomistic simulations of lipid bilayers have been performed to study the permeation of small molecules through a lipid bilayer [6], the interaction between lipid bilayers and substrates [7], the behavior of charged and neutral bilayers [8], and a large variety of additional bilayer properties and behaviors.[(19),20,(21)] In a recent united-atom study, Kukol demonstrated that the GROMOS96 53a6 forcefield [(22)] and the Kukol DPPC3 topology could be used with GROMACS [(23),(24)] to reproduce the experimental area per lipid of a preformed DPPC bilayer for a united-atom system composed of 128 lipids and 3655 water molecules with 3% accuracy without assuming constant surface area or including surface pressure.[(25)] , one drawback associated with atomistic models is that the detail that makes these models so realistic and appealing also makes them extremely computationally intensive

and prevents them from examining large conformational changes or long time scales. For example, the 90 ns atomistic simulation of 1017 DPPC lipids and 106,563 water molecules by de Vries et al. in 2004, which shows the spontaneous formation of a DPPC vesicle, was run on four or eight processors at a rate of only 1 ps per processor CPU hour (1.7 GHz Intel Pentium IV processors).[(26)]

Coarse-grained models of lipids have been developed in order to reduce simulation time so as to access longer time scales than are achievable in atomistic simulations. In these models clusters of atoms are grouped together into single sites to reduce the number of events that require calculation.[(27] A popular explicit-solvent coarse-grained model for lipids is MARTINI which was developed by Marrink et. al. and has been used to simulate the spontaneous aggregation of a DPPC bilayer.[(28)] In this model, an average of four atoms are represented by a single interaction site, 10 different types of interaction sites are defined, and the interaction strengths between any two sites are assigned one of five values.[(28)] The Marrink predictions of the DPPC area per head group, bending modulus, area compressibility, lipid lateral diffusion coefficient and water permeation rate closely matched the experimentally measured quantities.[(28)] Marrink et. al improved the MARTINI force field (creating version 2.0) by increasing the number of types of possible interaction sites from 10 to 18 and increasing the number of interaction strength levels from 5 to 10.[(29)] MARTINI (version 2.0) was applied to model molecular raft formation in model membranes.[(29),(30)] Although we have looked, we have been unable to locate any information about the computational speed of the MARTINI model for lipid systems. Orsi and coworkers developed a coarse-grained explicit solvent model for DMPC lipids which

represents the 118 atoms of DMPC by 10 coarse-grained sites; the predicted structure, elasticity, electrostatics and dynamics of a DMPC bilayer quantitatively matched experimental data.[(27)] In this model, the spherical units representing the headgroup choline and phosphate groups interact via the Lennard Jones potential and the glycerol and hydrocarbon groups of the lipids are modeled as soft uniaxial ellipsoids through the Gay-Berne potential.[(27)] The Orsi model was extended to dioleylphosphatidylcholine (DOPC) by adopting a 12 site coarse-grained representation.[(31)] Their simulations in 2010 of the formation by 128 DOPC lipids of a defect-free bilayer required approximately 2.5 days; those of the formation by 128 DMPC lipids of a defect-free bilayer with an embedded water pore required approximately 25 days (Intel 2.8 GHz processors in serial).[(31)] Subsequently, Orsi and Essex developed the electrostatics-based "ELBA" 1.0 force field for coarse-grained models of lipid membranes which explicitly represents charges and dipoles.[(32)] In this new model the Gay-Berne components are replaced with Lennard Jones potentials. DOPC, DOPE and DSPC are each represented by 15 spherical CG sites and water is treated explicitly. The ELBA 1.0 force-field was found to accurately reproduce several of the experimentally observed physical properties of single-species lipid bilayers composed of DOPC, DOPE or DSPC in the liquid crystal phase and DSPC in the gel phase. ELBA 1.0 was later refined to become ELBA 1.1 to correctly reproduce the hexagonal inverse phase for DOPE-water systems.[(33)] ELBA 1.1 was also used to simulate mixed DOPC-DOPE bilayers at various compositions and to calculate the first reported values for the lateral pressure and electrical potential profiles for mixed DOPC-DOPE bilayers.[(33)] In a recent study, DeNicola and co-workers developed an explicit-solvent coarse-grained

model for phospholipids for use with hybrid particle field molecular dynamics simulations.[(34)] In this work, the MARTINI coarse-grained mapping scheme is used to represent DPPC and the model parameters are optimized so that the coarse-grained model reproduces the structural properties of the reference particle-particle simulations.[(34)]

Some coarse-grained models ignore the motion of solvent atoms to further enhance the computational efficiency associated with coarse-graining. Instead, the effect of solvent atoms is included implicitly through the use of effective potentials, or potentials of mean force. Recently, Wang and Deserno presented an implicit solvent, coarse-grained model for POPC bilayers derived using a multiscale modeling approach based on structure-matching methodology.[(35)] In this model the 134 atoms in each POPC lipid molecule are represented by 16 coarse-grained sites of 8 different types. The coarse-grained potentials were optimized iteratively to reproduce radial distribution functions and the area per lipid of the bilayer obtained from all-atom simulations performed with the molecular dynamics program NAMD [(36)] and the fully atomistic CHARM27 [(37)] parameters. In order to promote lipid aggregation in this model, it was necessary to introduce additional cohesive interaction potentials between the alkyl tails and between the interfacial head group sites. The strength of the cohesive interaction potentials were chosen to promote bilayer stability, to match RDFs for the coarse-grained and atomistic simulations, and to optimize the lateral stress profile; without it, the bilayer falls apart. This model was used to simulate the self-assembly of 288 POPC lipids into a bilayer from a random lipid dispersion that quantitatively matched experimental bilayer properties. A defect-free bilayer formed in approximately 32 CPU hours (Xeon E5430 2.66 GHz chips in parallel with infiniband connection).[(35)] A

similar approach was taken by Lyubartsev who constructed a coarse-grained, implicit-solvent lipid model containing 10 coarse grained sites to represent the 118 atoms of DMPC; the parameters were optimized to reproduce the radial distribution functions from all-atom molecular dynamics simulations performed with the MDyna*Mix* [(38)] package and the all-atomic CHARMM27 [(20),(37)] force field.[(39)] Simulations of DMPC lipids using this model show the formation of bicelles and vesicles starting from a disordered system of lipids.[(39)] Izvekov et. al. used a multiscale coarse-graining approach to develop a model in which 11 different coarse-grained types are used to represent DMPC and cholesterol molecules.[(40)] The Izvekov model accurately reproduced the structural and elastic properties of a DMPC lipid bilayer and was used to simulate pre-formed DMPC/cholesterol liposomes. However, simulations of a DMPC/cholesterol system starting from a random dispersion did not form a bilayer. Instead, this system assembled into aggregates composed of DMPC and cholesterol and aggregates composed primarily of cholesterol.[(40)]

In this paper, we describe the development of an implicit-solvent, coarse-grained model, LIME, derived using a multi-scale modeling approach that enables simulations of large numbers of phospholipids in aqueous solution. The number of coarse-grained sites per lipid, 14, is very similar to the coarse-grained phospholipid representation of DPPC in other models,[(28),(31),(35),(39)] but the interactions are not. Instead the interactions between coarse-grained sites are represented by hard sphere and square well potentials as opposed to Lennard Jones potentials, thereby allowing us to use discontinuous molecular dynamics, a fast alternative to traditional molecular dynamics. The multiscale modeling procedure used to determine the model parameters involves the following steps. The trajectory data from a 20

ns GROMACS simulation using the GROMOS96 53a6 united-atom forcefield for a system containing 30 DPPC phospholipids is coarse-grained into 14 sites. United-atom simulations are essentially the same as atomistic simulations with the exception that hydrogen atoms bonded to carbon atoms are represented as a single site. Radial distribution functions (RDF) between all bonded and non-bonded pairs of coarse-grained sites are calculated and used to determine LIME geometrical and energetic parameters. The hard sphere diameter between non-bonded coarse-grained sites is estimated to be the smallest distance at which the RDF takes a non-zero value. The RDFs were also used to estimate the square well width, and the minimum and maximum bond lengths between bonded pairs. The relative stiffness of each lipid is maintained by imposing pseudobonds, which limit the bond length fluctuations to the values observed in the GROMACS simulations. Interaction energies between non-bonded coarse-grained sites are determined by calculating the potential of mean force using a one-step Boltzmann inversion scheme.[(41),(42)] In the model each coarse-grained site has its own realistic mass.

Highlights of our results include the following; The model successfully simulates the spontaneous assembly of a DPPC bilayer composed of 256 lipids in less than 4 CPU hours starting from a random initial configuration, which is approximately an order of magnitude faster than the fastest reported coarse-grained implicit solvent model. The area per lipid of our bilayer is within 2% of the value calculated from our GROMACS simulation data and the literature value. The thickness of our bilayer is within 5-7% of the literature value and within 4-8% of the thickness of the bilayer in our GROMACS simulations. The orientational order parameter of the alkyl tail bonds in LIME are in excellent agreement with those

calculated from the GROMACS model. The mass density profiles of the LIME and GROMACS models closely match each other. Finally, LIME is able to simulate transbilayer flip-flop in which a lipid flips from the bottom leaflet of a membrane to the top leaflet.

2.2 Theoretical Methods

In LIME six different coarse-grained types (I – VI) are used to represent the 130 atoms that make up a DPPC molecule. DPPC is composed of a polar head group that includes a choline, phosphate and two ester linkages, and two nonpolar hydrophobic acyl tails. **Figure 2.1** illustrates the coarse-graining of a DPPC molecule from 50 united-atom (**Figure 2.1a**) to the 14 coarse-grained sites in the LIME representation. This figure and all other figures depicting lipid molecules throughout the paper were generated with Visual Molecular Dynamics (VMD).[(43) Each coarse-grained site that represents a unique set of atoms is assigned a different coarse-grained type. Each coarse-grained type is represented by a different color in **Figure 2.1b**. **Table 2.1** lists the atoms included in each coarse-grained site, the "type" assigned to each coarse-grained site and the mass of each coarse-grained site. Types I and II represent the choline entity and the phosphate group, respectively. Ester coarse-grained sites 3 and 9 are assigned types III and IV, respectively. The coarse-grained sites in the hydrocarbon tails (excluding the terminal sites) are assigned type V. Finally, the terminal tail coarse-grained sites are classified as type VI. We considered treating the terminal tail beads (sites 8 & 14) as the same type as the non-terminal tail beads (sites 4-7 & 10-13) because type V differs from type VI only by a single hydrogen atom. However, we found that the epsilons for type V&V pairs (ε = -0.050 eV) are significantly

different than the parameters for type VI&VI pairs (ε = -0.070 eV). This is probably because of the differences in connectivity between the two types. Therefore, we felt that it was important that the different groups have their own unique types.

In addition to coarse-graining and treating solvent implicitly, we employ discontinuous molecular dynamics (DMD) simulation to further increase the speed of our code. DMD [(44)] is a very fast alternative to traditional molecular dynamics simulation that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials. For this reason, all of the inter- and intra- molecular interactions in our lipid model are represented by a combination of hard-sphere and square well potentials, as opposed to the Lennard Jones, Coulombic and harmonic potentials found in traditional molecular dynamics simulations. Unlike continuous potentials, such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide. This enables the exact (as opposed to the numerical) solution of the collision dynamics. This imparts great speed to the algorithm and allows sampling of much wider regions of conformational space, longer time scales and larger systems than in traditional molecular dynamics. Molecules are given an initial random configuration that satisfies both excluded volume and angular constraints. Initial velocities are chosen randomly from a Maxwell-Boltzmann distribution about the desired temperature. Particle trajectories are followed by calculating the time between each collision and advancing the simulation to the next event. Types of events include a collision between two spheres, a bond event when the distance between two bonded spheres reaches a minimum or maximum limit, and a square well event when two spheres enter (capture), unsuccessfully attempt to escape (bounce) or successfully

leave (dissociation) a square-well attraction.[(44),(45),(46),(47)] The simulations are performed with the number of particles, the temperature and the volume held constant. The temperature is maintained constant using the Andersen thermostat, which uses ghost collisions with randomly selected particles in the system to maintain the Maxwell-Boltzmann velocity distribution about the desired temperature.[(48)] The simulations were run at constant volume as this is the most straightforward ensemble for use in DMD. In the Conclusion section we discuss the possibility of simulating lipids in the NPT ensemble using a combination of a hybrid Monte Carlo/DMD approach.

Data used to calculate the coarse-grained model parameters was obtained by running explicit-solvent NPT ensemble united-atom simulations on a system containing 30 DPPC phospholipids and 8655 water molecules. The GROMACS simulation package [(23),4], version 4.5.4, was used with the GROMOS96 53a6 forcefield [(22)] and the Kukol DPPC3 topology, which has been shown previously to accurately reproduce the experimental area per lipid, lateral self diffusion constant and deuterium order parameters for the acyl chains in DPPC bilayers in solution.[(25)] The initial configuration of this system was random in a box with equal sides of length 100.0 Å. The Berendsen thermostat [(49)] was used to keep the temperature constant at 325K throughout the simulation with a time constant of 0.1ps. The simulation was run for 20 ns with a time step of 0.002 ps for approximately 48 CPU hours. Periodic boundary conditions were applied and the pressure was maintained at 1.0 bar. Throughout the GROMACS simulation the coordinates of each atom were written to an output trajectory file every 1 ps. These output files were used to calculate the centers of mass for each of the 14 coarse-grained sites on the 30 phospholipids.

We chose to run atomistic simulations of a small system to gather data for use in calculating coarse-grained parameters because we did not want to restrict the movement of the lipids during the atomistic simulation. It is not uncommon to gather data for a coarse-grained model from a small system. For example, Lyubartsev performed atomistic simulations of only 16 DMPC lipids to obtain data for use in calculating coarse-grained parameters.[(39)] To ensure that our atomistic simulation of 30 DPPC lipids was not too small or at too low a density, we compared the radial distributions we calculated from this simulation with those calculated from a GROMACS simulation of 128 DPPC lipids in a box with dimensions of 64Å x 64Å x 90Å. The sigma and lamda values calculated from each GROMACS simulation were nearly identical. Furthermore, the epsilon values calculated from each simulation differed by minimal amounts.

2.3 Results and Discussion

The LIME interaction energies were determined using a one-step Boltzmann inversion procedure inspired by the iterative Boltzmann inversion scheme, which is a popular strategy used to systematically compute potentials for coarse-grained simulations.[(35),(41),(42)] We begin by reminding the reader of the iterative Boltzmann inversion scheme, an approach based on the idea that the effective potential, or potential of mean force, U(r), between two molecules in a sea of molecules can be obtained from the radial distribution g(r) using:

$$U(r) = -k_B T \ln[g(r)] \qquad \text{Equation 1}$$

where k_B is the Boltzmann constant and T is the temperature of the system. The iterative Boltzmann inversion scheme involves the following steps: (1) Data from an atomistic simulation is used to calculate the intermolecular radial distribution function g(r) between coarse-grained sites. (2) An initial guess for the potential of mean force, U(r), between the coarse-grained sites is determined using Equation 1. (3) A coarse-grained simulation using the initial guess for the potential of mean force is run and a new g(r) between coarse-grained sites is calculated. (4) The difference between the coarse-grained and the atomistic potentials of mean force is used to generate a correction to the coarse-grained potential of mean force. (5) This process is repeated iteratively until the coarse-grained and atomistic potentials of mean force match each other within a prescribed tolerance.[(41),(42)]

Instead of using the iterative Boltzmann inversion procedure described above, we use a simplified, one-step Boltzmann inversion to obtain the LIME interaction energies. **Figure 2.2** outlines this approach, the procedure is the following: (1) the average radial distribution function between two intermolecular coarse-grained sites is determined, (2) the potential of mean force is calculated using **Equation 1,** and (3) the minimum value of the potential of mean force between the coarse-grained sites, ε, is chosen to be the depth of the square well potential. Mathematically this can be expressed as:

$$\varepsilon = -k_B T \ln[g(r)_{MAX}] \qquad \text{Equation 2}$$

where $g(r)_{MAX}$ is the maximum value of g(r) in the radial distribution function and T is the temperature of the system. If the ε between two coarse-grained sites is greater than -0.005 eV, the sites are assumed to have a hard-sphere interaction (ε = 0.0 eV).

The iterative Boltzmann inversion approach was not used to obtain the LIME parameters because the shape of the radial distribution function for a discontinuous potential obtained by coarse-graining the GROMACS simulation is dramatically different from the shape of the radial distribution function associated with a square-well potential. See for example the radial distribution function for sites 1 & 1 obtained during a LIME simulation shown in **Figure 2.3**. As is expected for square-well systems, there are no oscillations and there are discontinuities in g(r) at distances that correspond to the discontinuity in the potential. Thus it does not make sense to use an iterative procedure to try to match the GROMACS and coarse-grained potentials. By using the one-step Boltzmann inversion scheme we are able to develop a model that very accurately matches experimental observations for a lipid bilayer. Since we were able to obtain very good agreement between the physical properties of our LIME bilayer with experimental values we did not attempt to go beyond a one-step scheme.

The LIME hard sphere (σ_{HS}) diameters and square-well (λ) widths were determined from the radial distribution functions between pairs of non-bonded coarse-grained sites in the GROMACS simulations which were run on 30 DPPC and 8655 water molecules. As shown in **Figure 2.2a** the hard sphere diameter (σ_{HS}) for each pair of interaction sites was determined by locating the smallest non-zero separation between the two sites. The square well sphere diameter (λ) for each pair of interaction sites was determined by examining the radial distribution function between those sites. First, the local maximum (labeled **b** in **Figure 2.2a**) located at the largest distance less than 15 Å was identified. Next, the local

minimum (labeled **a** in **Figure 2.2a**) preceding this local max was identified. Finaly, λ was calculated using:

$$\lambda = a + 2d \qquad \text{Equation 3}$$

where d = (b-a). If a λ value greater than 15 Å was calculated, the procedure for calculating λ was repeated using a local maximum closer to the origin. Sample radial distributions for intermolecular coarse-grained types 1&1, 1&2 and 5&5 are provided in **Figure 2.4**. The hard sphere diameters (σ_{HS}) determined for coarse-grained types 1&1, 1&2 and 5&5 are 4.35 Å, 3.85 Å and 3.75 Å, respectively. The values of λ for coarse-grained types 1&1, 1&2 and 5&5 were found to be 12.65 Å, 9.85 Å and 11.56 Å, respectively. A complete list of all hard sphere diameters, square well widths and interaction energies is provided in **Table 2.2.**

The minimum and maximum bond lengths were determined by plotting the radial distribution functions for bonded coarse-grained sites. The minimum bond length (σ_{MIN}) was chosen as the smallest possible distance between two bonded coarse-grained sites. The maximum bond length (σ_{MAX}) was chosen as the largest distance for which a non-zero g(r) was observed. Sample distributions for intramolecular coarse-grained types 1&2, 3&5, and 5&6 used to determine σ_{MIN} and σ_{MAX} for bonds and the resulting values of σ_{MIN} and σ_{MAX} are provided in **Figure 2.5**.

The relative stiffness of the lipid molecule is maintained by imposing pseudobonds, which limit the fluctuation of coarse-grained sites to the angles and torsional angles observed during the GROMACS simulations. Bond angles were maintained by imposing pseudobonds between all next nearest neighboring sites along the chain. Torsional angles were maintained

with pseudobonds between next next nearest neighboring sites along the chain. Bond distributions for intramolecular sites calculated from the GROMACS simulations were used to determine the minimum and maximum values for the pseudobond lengths in LIME. The minimum pseudobond length was determined by finding the smallest distance at which the intramolecular bond distribution function exceeds 30% of its maximum value. The maximum pseudobond length was determined by finding the smallest distance, larger than the distance at which the bond distribution function maximum occurs, where the intramolecular bond distribution function falls below 30% of its maximum value. Pseudobonds were also added between intramolecular coarse-grained sites 5&11, 6&12, 7&13 and 8&14 on the same chain to restrict the separation between the tails to the distances observed during the GROMACS simulations. This was done to prevent the tails from adopting conformations that were not frequently observed during the GROMACS simulations. The minimum pseudobond lengths between intramolecular sites 5&11, 6&12, 7&13, and 8&14 were determined by locating the smallest distance at which the intramolecular bond distribution function exceeds 30% of its maximum value. The maximum pseudobond lengths between intramolecular sites 5&11, 6&12, 7&13 and 8&14 were determined by finding the smallest distance, larger than the distance at which the intramolecular bond distribution function maximum occurs, where the bond distribution function falls below 30% of its maximum value.

In all the simulations run using the LIME force field, periodic boundary conditions were implemented to eliminate any artifacts that might be caused by the box walls. In addition, all simulations were carried out in the canonical ensemble where the number of

particles, the temperature and the volume are held constant. Simulation temperature in LIME is expressed in terms of the reduced temperature, T*:

$T^* = k_B T/\varepsilon^*$ Equation 4

where k_B is Boltzmann's constant, T is the temperature, and ε* is the reference interaction strength.[(50)] The reference interaction strength, ε*, was calculated using:

$$\varepsilon^* = \frac{\sum_{ij} n_{ij}\varepsilon_{ij}}{\sum_{ij} n_{ij}}$$ Equation 5

where n_{ij} is the number of coarse-grained sites with a type i and type j interaction and ε_{ij} is the interaction energy between coarse-grained types i and j. The ε_{ij} values were obtained from the GROMACS simulations at T=325K. The resulting value for ε* calculated from **Equation 4** is 0.0363. Thus when $T^* = k_B T/\varepsilon^* = (8.6173 \times 10^{-5} eV/K)*(325K)/(0.363eV) =$ 0.77 in our DMD/LIME simulations, the lipid molecules will behave as they would at a real temperature of 325K. The Andersen thermostat is used to hold the temperature constant. In this method randomly selected particles collide infrequently with ghost particles, effectively reassigning the particle's velocity randomly so as to maintain a Maxwell-Boltzmann distribution centered at the simulation temperature.[(48)] All LIME/DMD simulations were run with a DMD software program developed in the Hall research lab called EMBLEM. This program is written in C++. The Intel compiler was used to compile this code and all other codes used in the development and analysis of LIME. All simulations were run in

serial. DMD can be run in parallel and in the future it is likely that we will parallelize our code.[(51)]

Five independent DMD simulations starting from different random configurations were run using the LIME force field to determine whether or not a bilayer could be formed starting from a random configuration of 256 DPPC phospholipids at a T* = 0.77. The lengths of the sides of the simulation cell were set to 90 Å. A bilayer was formed in all five simulations. Simulations of the 256-lipid system were run with several different box sizes to ascertain which box size should be used to evaluate bilayer properties. First, we simulated a system of 256 lipids in a box with dimensions of 100Å x 100Å x 100Å. This system formed a bilayer with a large hole in it as shown in **Figure 2.6a**. The area per lipid of this bilayer, counting only portions external to the hole, was 63.3 $Å^2$±0.1 $Å^2$. (If it had spanned the box, the area per lipid would have been 78.1 $Å^2$). Although at first glance the hole in our bilayer may appear to be a hole of vacuum, this is not the case. In the implicit-solvent approach that we are using all of the empty space is meant to represent a structureless solvent. This is a consequence of using the McMillan-Mayer solution approach in which a two-component system is mapped onto to a one-component system by integrating out the degrees of freedom of solvent and hence increasing the speed of our simulations.[(52)] Since our lipid parameters are calculated using a method that accounts for the effect of water, they are expected to behave as they would in an aqueous solution, not a vacuum. See the Conclusions section for more discussion of this issue. Next, we ran a simulation of 256 DPPC lipids in a box with dimensions of 80Å x 80Å x 80Å. The bilayer formed in this simulation did not span the box; instead portions of the bilayer crawled upwards along the

sides of the simulation box. Thus there were obviously too few lipids in the first box size ($100Å^3$) and too many lipids in the second box size ($80\ Å^3$). Since the bilayer formed in the first simulation adjusted its size naturally, unencumbered by the constraints imposed by periodic boundary conditions, we surmised that its area per lipid ($63.3\ Å^2$) was likely to be characteristic of an equilibrium structure. Hence we decided to perform all of our 256-lipid production simulations in boxes of size of 90Å x 90Å x 90Å, since this gave the best chance of having an area per lipid of $63.3\ Å^2$. **Figure 2.6b** shows an aerial image of a bilayer formed during one of those simulations.

Snapshots of one system at different time points throughout the simulation are provided in **Figure 2.7**. Each lipid is represented according to the following color scheme: purple (choline entity – type I); orange (phosphate group – type II); red (ester groups – type III and type IV); cyan (alkyl tail groups – type V and type VI). The initial random configuration of the system is shown in **Figure 2.7(a).** The lipids begin aggregating at around 15 million collisions (**Figure 2.7(b)**) and form a single disordered aggregate (**Figure 2.7(c)**) at 50 million collisions. The snapshots in **Figures 2.7(d – f)** show the aggregate as it rearranges to form a defect free bilayer. The 150 million collisions required to form the defect free bilayer shown in **Figure 2.7f** took approximately 3.8 CPU hours. The time scale for aggregate formation, 50 million collisions, is rather quick compared to the time it takes, 150 million collisions, to organize into a defect free bilayer.

It is of interest to compare the structural properties of the DPPC bilayer formed using LIME/DMD simulations with the structural properties obtained from a GROMACS simulation. Accordingly, a GROMACS simulation of a pre-formed defect-free DPPC bilayer

composed of 128 DPPC lipids was performed using the GROMOS96 53a6 forcefield in the NPT ensemble. The initial coordinates for the bilayer were obtained from the supporting information provided by Kukol.[(25)] The Berendsen thermostat was employed to maintain the temperature at 325K; the pressure was held constant at 1.0 bar.[(49)] The simulation was run for 20 ns with a time step of 0.002 ps. The bilayer remained stable throughout the simulation.

The bilayer thickness and the area per lipid of the bilayers formed during the LIME simulations closely match both experimental values and the values calculated from the GROMACS simulation described above. The LIME values for the bilayer thickness and the area per lipid were calculated by averaging the data from the five independent DMD simulations, which were all started from different initial configurations. After a bilayer formed, the simulation was continued for an additional billion collisions. Data from the billion collisions following the formation of a defect free bilayer was used to calculate the physical properties of that bilayer, including the bilayer thickness, area per lipid and bond order parameters. The bilayer thickness is defined as the distance along the direction perpendicular to the bilayer normal between the average location of phosphate groups in the top and bottom leaflets of the bilayer. The LIME/DMD value for the thickness of the DPPC bilayer is 35.7 Å $\pm$ 0.3 Å at a reduced temperature of 0.77. The experimental value for the DPPC bilayer thickness, which is measured as the distance between phosphate groups in the upper and lower leaflets of the bilayer in the electron density profile is approximately 38.0 Å at 50^{o}C.[(53),(54)] The bilayer thickness measured between two type 2 coarse-grained (phosphate) groups during the GROMACS simulations was 38.0 Å

$\pm$ 0.6 Å at 52°C. Thus, the LIME bilayer thickness is within 5-7% the experimental value and 4-8% the GROMACS bilayer thickness. The area per lipid in our LIME/DMD and GROMACS simulations was calculated by multiplying the length and width of the bilayer and dividing by half of the number of the lipids in the system (to approximate the number of lipids in each leaflet). The area per lipid for the bilayer formed by DPPC in our DMD/LIME simulations is 63.3 Å^2, which is very close to the experimental area of 63.0 $\pm$ 1.0 Å^2 at T = 323K reported by Kucerka et al.[(54)] The area per lipid calculated for the DPPC bilayer in our GROMACS simulations at a T = 325K is 64.6 $\pm$ 0.1 Å^2.

To further evaluate the structural properties of our coarse-grained model the orientational order parameter, s_{bond}, for different bonds along the chain was calculated; it is defined to be:

$$S_{bond} = \frac{1}{2}\left\langle 3\cos^2\theta - 1 \right\rangle \qquad \text{Equation 2}$$

where θ is the angle between the vector along a coarse-grained bond and the bilayer normal.[(28),(35)] S_{bond} values of 1, -1/2 and 0 correlate to bonds with a parallel, an alignment perpendicular to the bilayer normal and completely random alignment with the bilayer normal.[(28),(35)] To obtain the orientational bond order parameters from the GROMACS simulation of the DPPC bilayer, the data was coarse-grained and the bond order parameters for different site types along the chain were calculated. The GROMACS simulation was run at a temperature of 325K and the LIME/DMD simulation was run at the equivalent reduced temperature of 0.77. **Figure 2.8** shows the values of the orientational bond order parameters obtained for the LIME coarse-grained model (green line) and for the

GROMACS simulation (blue line) versus the bond numbers, which are defined in the figure insert. The orientational bond order parameters from the DMD model and the GROMACS simulations are in very close agreement. The bond order parameters range from approximately 0.26 to 0.52 in **Figure 2.8**, indicating that the bonds in both the GROMACS and the DMD simulations are not very well ordered and have between a parallel and a completely-random alignment with the bilayer normal. These bond order parameters are very similar to those calculated by Marrink and co-workers for a DPPC bilayer at a temperature of 323K; their values ranged from approximately 0.3 to 0.5.[(28)] A direct comparison cannot be made between the LIME forcefield and the model developed by Marrink and co-workers because in LIME each alkyl tail of DPPC is represented by 5 coarse-grained sites and in the Marrink model each alkyl tail of DPPC is represented by 4 coarse-grained sites.

As the temperature is cooled in our DMD system, the bilayer undergoes a phase transition from a liquid-crystalline phase to a tilted gel phase. Experimentally, as a DPPC bilayer is cooled it also undergoes a phase transition from a liquid-crystalline phase to a tilted gel phase in which the lipid tails are tilted with respect to the bilayer normal.[(55),(56)] The tilted gel phase generally has a smaller area per lipid than in the liquid-crystalline phase and the lipid tails are straighter.[(56)] The temperature at which a DPPC bilayer transitions from the liquid-crystalline phase to the gel phase is reported experimentally as 314.4K.[(57)] Three LIME/DMD simulations, each starting from a different configuration of a preformed DPPC bilayer composed of 128 lipids at $T^* = 0.77$ and at constant volume in a box with dimensions of 64Å x 64Å x 90Å were run. During each simulation the following cooling

procedure was implemented: T* was decreased from 0.77 to 0.30 at a rate of 0.01 T*/million collisions and then maintained at 0.30 for 550 million collisions. All properties were calculated using the last 100 million collisions that a bilayer was at T*=0.30. **Figure 2.9** provides snapshots of a DPPC bilayer in a DMD/LIME simulation at (a) a reduced temperature of 0.77 where a liquid crystalline phase is observed and (b) a reduced temperature of 0.30 where a tilted gel phase is observed. **Figure 2.9c** shows a cross-tilted gel phase, which we observe in some of our simulations but not in others. We are currently investigating the conditions that lead to the tilted gel phase and the cross-tilted gel phase. We believe that the cooling rate may determine whether the tilted or cross-tilted gel phase is formed. As the temperature of the DPPC bilayer is decreased the tails become more straight and rigid causing the bilayer thickness to increase from 35.7 Å $\pm$ 0.3 Å at T* = 0.77 to 40.9 Å $\pm$ 1.0 Å at T* = 0.30. Consistent with our predictions, experimental observations also show that the bilayer thickness increases with decreasing temperature. For example, the experimental bilayer thicknesses at T=323K and T=293K are reported as 38.3 Å and 44.2 Å, respectively.[(58)] We find that as the DPPC bilayer transitions to the gel phase, the area per lipid decreases; e.g. the area per lipid at T* = 0.77 and T* = 0.30 was 63.3 Å^2 and 49.6 Å^2 $\pm$ 1.4 Å^2, respectively. The decrease in area per lipid with decrease in temperature agrees with experimental observations. The experimental value of the area per lipid at T=323K and T=293K are 64.0 Å^2 and 47.9 Å^2, respectively.[(58)] These simulations were run at constant volume with box dimensions of 64Å x 64Å x 90Å. At a T*=0.77 the bilayer spans the entire x-y plane. As the bilayer is cooled it stops spanning the entire x-y plane because its area per lipid decreases. A view of the bilayer from the top of the box

shows that the bilayer is present in only a small portion of the x-y plane. The volume of the simulation box was constant with dimensions of 64Å x 64Å x 90Å as the bilayer was cooled to T*=0.30. In our constant volume production run simulations, we adjusted the box volume so that our bilayer would span the entire x-y plane throughout the simulation. We did not adjust the box volume during the cooling simulations for the bilayer in the interest of computational efficiency because this would have required us to constantly change the box volume. In the future we plan to adjust the box volume during cooling simulations. The tilted gel phase, which we observe in our LIME/DMD simulations is not usually observed in coarse-grained simulations. For example, coarse-grained simulations performed by Marrink and co-workers and by Wang and Deserno also show the formation of untilted gel phase lipid bilayers.[(56),(59)] However, atomistic simulations performed by Leekumjorn and Sum did show the formation of a tilted gel phase of DPPC.[(60)] The question naturally arises as to why we had to go to such low reduced temperatures to observe the gel phase when it is typically reached experimentally at 314.4K [(57)] when the liquid crystalline phase is simulated at a temperature of 325K. Our explanation is that the LIME epsilon values were only calculated at a temperature of 325K, which is equivalent to a LIME/DMD reduced temperature of 0.77 and that they were then assumed to be independent of temperature. In the future, a multiscale modeling procedure may be used to determine the temperature dependence of the epsilons. This would require data from atomistic simulations at various temperatures. We speculate that once the temperature dependence of the epsilon values is incorporated into our DMD/LIME simulations, it will be unnecessary to reduce T* so dramatically to observe the gel phase.

The transbilayer movement of phospholipids from one leaflet to another, known as "translocation" or "flip-flop," was measured during the DMD/LIME simulations. Translocation is thought to play an important role in numerous cellular processes including cell apoptosis and drug function.[(61)] Unless protein-mediated, this lipid migration is thought to occur very slowly, with half-lives on the order of hours.[(61),(62),(63)] We did not observe any flip flops in any of our simulations of lipid systems at T*=0.77. However, when we performed a simulation of DPPC bilayer composed of 128 lipids at a slightly higher temperature, T*=0.85, we did see one flip flop. **Figure 2.10** shows snapshots of a lipid (in yellow) that flips from the bottom leaflet of the membrane to the top leaflet of the membrane during one of these simulations. Over the course of 1 billion collisions only one lipid successfully flipped from one leaflet to another. It required approximately 19 million collisions for the lipid to complete the flip-flop.

A comparison of the mass density profile of different coarse-grained types along the bilayer normal between the LIME and GROMACS simulations is presented in **Figure 2.11**. The mass density profile is the mass per unit volume at a distance (z) from the bilayer normal. Since the GROMACS simulation from which the mass density profile was obtained used a bilayer consisting of 128 DPPC molecules, we ran a DMD/LIME simulation on a bilayer consisting of 128 lipids. The LIME/DMD simulation was started from a random initial configuration of 128 lipids and formed a defect free bilayer in 1.1 CPU hours. The mass density distributions from the DMD/LIME simulation were taken from conformations once the lipids had formed a defect free bilayer (after 41 million collisions). The density profiles for the GROMACS simulations, were obtained by

coarse-graining the GROMACS simulation result. The mass density profiles of the GROMACS and LIME simulations closely align with each other in that peaks of the mass density distributions for corresponding coarse-grained types are within 2 Å of each other.

It is useful to compare LIME with other coarse-grained lipid models that have recently been described in the literature. Most of the previously reported coarse-grained lipid models use continuous potentials, which are usually derived from atomistic or united-atom simulations to describe the forces between interacting sites. For example, Wang and Deserno report the development of a POPC lipid model that is similar to LIME in that it coarse-grains molecules and treats solvent implicitly.[(35)] In the Wang and Deserno model, each POPC molecule is represented by 16 coarse-grained sites that have one of 8 different types. In contrast to LIME however, this model uses continuous potentials and an additional cohesive interaction potential between alkyl tails and interfacial head groups to drive lipid aggregation and to prevent the bilayer from falling apart.[(35)] Another recently developed coarse-grained model for lipids is that of Orsi and co-workers.[(32)] This model was parameterized for DOPC, DOPE and DSPC lipids and represents all three lipids with 15 coarse-grained sites.[(32)] Unlike LIME however, the Orsi model accounts for water explicitly and system electrostatics which allows the membrane dipole potential to be studied.[(32)] The popular MARTINI force field, developed by Marrink and co-workers uses Lennard Jones potentials to describe the interactions between coarse-grained lipid sites.[(28),(29)] In the MARTINI model, water is treated explicitly and DPPC is represented by 12 coarse-grained sites and each site is assigned one of 4 types. This model was parameterized by systematically

reproducing partitioning free energies between polar and apolar phases of a large number of chemical compounds.[(29)] In addition, the MARTINI force field allows charged groups to interact via an electrostatic Coulombic potential.[(28),(29)] In comparison, the current version of LIME does not include any additional forces for charged particles. LIME accounts for strong and weak attraction between sites by adjusting the depth of the square-well potential appropriately. Similar to other models LIME is parameterized from simulations and reproduces the spontaneous formation of a lipid bilayer and the formation of a gel phase with a decrease in temperature. In addition, the bilayer thickness, area per lipid and orientational bond order parameters closely agree with experimental values.

2.4 Conclusion

We described the development of LIME, an intermediate-resolution, implicit-solvent, coarse-grained model for phospholipid molecules designed for use with discontinuous molecular dynamics. In LIME, each of the 14 coarse-grained sites that make up a DPPC molecule are assigned one of 6 different types: the choline entity, the phosphate group, each ester group, each alkyl tail group and each terminal tail group. The "LIME" parameters were obtained using a multiscale modeling approach in which the geometric and energetic parameters were calculated from data collected from GROMACS simulations of a system composed of 30 DPPC molecules. A one-step Boltzmann inversion approach, which is a simplified version of the iterative Boltzmann inversion scheme, was used to calculate the depth of square well interactions in the model.

The physical properties of a DPPC bilayer, which formed spontaneously in each of 5 LIME/DMD simulations that started from a different random configuration, accurately match experimental values. The area per lipid of the bilayer is within 2% of the experimentally reported value. The bilayer thickness is within 4-8% of the range of error of the value calculated from the GROMACS simulation data and within approximately 5-7% of the experimental value. The bond order parameter values calculated from the LIME/DMD simulations and the GROMACS simulation are in very good agreement. LIME also reproduces the transbilayer movement of lipids. This phenomenon, which is, found to occur infrequently experimentally was observed only once in a total of 2 billion collisions. The phase transition of a DPPC bilayer from a liquid-crystalline phase to a tilted gel phase was also simulated. As the DPPC bilayer was cooled from a liquid-crystalline phase to a gel phase the area per lipid decreased and the bilayer thickness increased which is the trend that is observed experimentally.

LIME has one distinctive characteristic compared to the other implicit solvent coarse-grained lipid models: its speed. One of the main advantages of combining coarse-graining with discontinuous molecular dynamics is the reduction in the computational time required to simulate systems of lipid molecules. The use of discontinuous molecular dynamics greatly enhances the speed of the simulations because calculations are required only when particles experience an event, as opposed to the calculations in traditional molecular dynamics that are required to determine the net force of every particle in a system at each time step. Therefore, our LIME model paired with discontinuous molecular dynamics enhances our ability to look at long time scale phenomena without sacrificing the ability to observe many of the essential

features observed at the atomistic scale (e.g. bilayer formation, phase transition, lipid translocation). In LIME the spontaneous formation of a bilayer composed of 256 lipids requires less than 4 CPU hours (Intel Xeon E5520 2.27GHz).

While the LIME/DMD approach has many advantages for simulating lipid self assembly it does have the following limitations: (1) treating solvent implicitly prevents diffusion and hydrodynamics from being taken into account, (2) approximating interactions with hard and square-well spheres forces the particles to interact sequentially rather than simultaneously as they do when the net force on each particle is calculated in traditional molecular dynamics, (3) a direct correlation can only be made between the temperature at which LIME was parameterized and its corresponding reduced temperature. Although we cannot correlate reduced temperatures other than $T^* = 0.77$ with a real temperature, we do reproduce the experimentally observed phase transition of a bilayer by decreasing the reduced temperature in our simulations. In addition to these limitations, LIME also lacks an explicit representation of electrostatics. This means we do not account for the very long-range portion of the electrostatic interactions, although we do account for its short range effects. Long range effects are known to play an important role in biomolecular systems.[(64)] We do however, account for electrostatic interactions implicitly during the parameterization process. Another limitation of LIME is that it does not measure real time during a simulation. Therefore, we cannot determine exactly how long in real time units it took for the DPPC bilayer to form, or for the lipid flip flop to occur. However, if the amount of time that it takes for a certain event to occur is known, the time scale of the DMD/LIME simulation could be approximated. Finally, we will address the use of constant volume in

our DMD/LIME simulations. The use of constant volume conditions in this work means that we do not allow the system to adjust its volume to the condition most favorable to the formation of bilayers. This leads to phenomena such as the large hole in the bilayer that we see in Figure 6a. We have gotten around this problem by performing test runs at a number of different volumes to see which volume sustains a defect-free bilayer. A more rigorous approach would be to perform the simulations at constant pressure, adjusting the volume as part of the simulation to find the state with the minimum free energy. To do this we would need to perform hybrid Monte Carlo—DMD simulation in which the volume change moves are made with Monte Carlo and particle displacement moves are made with DMD, as we have done in earlier work on chainlike systems.[(65),(66)] This would however slow down the code. We are considering this for future work.

2.5 References

1. Bennun, S.; Hoopes, M.; Xing, C.; Faller, R. *Chem. Phys. Lipids*. 2009, *159*, 59 – 66.
2. Meer, G.; Voelker, D.; Feigenson, G. *Nat. Rev. Mol. Cell Biol*. 2008, *9*, 112 – 124.
3. Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P. *Nature*. 2009, *459*, 379 – 385.
4. Karve, S.; Bandekar, A.; Ali, M.; Sofou, S. *Biomaterials*. **2010**, *31*, 4409 – 4416.
5. Schroeder, A.; Levins, C.; Cortez, C.; Langer, R.; Anderson, D. *Journal of Internal Medicine*. **2009**, *267*, 9 – 21.
6. Almeida, A.; Souto, E. *Adv. Drug Delivery Rev*. **2007**, *59*, 478-479.
7. Sharma, A.; Mayhew, E.; Straubinger, R. *Cancer Res*. **1993**, *53*, 5877 – 5881.
8. Sharma, A.; Straubinger, R. *Pharm. Res*. **1994**, *11*, 889 – 896.
9. Skubitz, K. *Cancer Investigation*. **2003**, *21*, 167 – 176.

10. Gabizon, A. *Cancer Investigation*. **2001**, *19*, 424 – 436.

11. Colbern, G.; Hiller A.; Musterer, R.; Pegg, E.; Henderson, I.; Working, P. *J. Liposome Res.* **1999**, *9,* 523 – 538.

12. Song, H.; Zhang, J.; Han, Z.; Zhang, X.; Li, Z.; Zhang, L.; Fu, M.; Lin, C.; Ma, J. *Cancer Chemother. Pharmacol*. **2006**, *57*, 591 – 598.

13. Pandit, S.; Jakobsson, E.; Scott, H.; *Biophy. J.* **2004**, *87*, 3312 – 3322.

14. Latour, R. *Biointerphases*. **2008**, *3*, 2 – 12.

15. Curtis, E.; Hall, C. Unpublished

16. Bemporad, D.; Essex, J.; Luttman, C. *J. Phys. Chem. B*. 2004, *108*, 4875 – 4884.

17. Heine, D.; Rammohan, A.; Balakrishnan, J. *Molecular Simulation*. 2007, *33*, 391 – 397.

18. Rog, T.; Murzyn, K.; Pasenkiewicz-Gierula, M. *Acta Biochim. Pol.* **2003**, *50*, 789 – 798.

19. Feller, S.; Gawrisch, K.; MacKerell, A. *J. Am. Chem. Soc.* **2002**, *124*, 318 – 326.

20. Feller, S.; MacKerell, A. *J. Phys. Chem. B.* **2000**, *104*, 7510-7515.

21. Rosso, L.; Gould, I. *J. Comput. Chem.* **2008**, *29*, 24 – 37.

22. Oostenbring, C.; Villa, A.; Mark, A.; Van Gunsteren, W. *J. Comput. Chem.* 2004, *25*, 1656 – 1675.

23. Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* 2008, *4*, 435 – 447.

24. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* 2005, *26*, 1701 – 1719.

25. Kukol, Andreas. *J. Chem. Theory Comput.* 2009, *5*, 615 – 626.

26. De Vries, A.; Mark, A.; Marrink, S. *J. Am. Chem. Soc.* 2004, *126*, 4488 – 4489.

27. Orsi, M.; Haubertin, D.; Sanderson, W.; Essex, J. *J. Phys. Chem. B*, **2008**, *112*, 802 – 815.

28. Marrink, S.; de Vries, A.; Mark, A. *J. Phys. Chem. B*. 2004, *108*, 750 – 760.

29. Marrink, S.; Risselada, H.; Yefimov, S.; Tieleman, D.; de Vries, A. *J. Phys. Chem. B*, **2007**, *111*, 7812 – 7824.

30. Risselada, H.; Marrink, S. *Proc. Natl. Acad. Sci.* **2008**, *105*, 17367 – 17372.

31. Orsi, M.; Michel, J.; Essex, J. *J. Phys. Condens. Matter*, 2010, *22*, 1 – 15.

32. Orsi, M; Essex, J. *PLoS ONE*, 2011, *6*, 1-22.

33. Orsi, M. and Essex, J. *Faraday Discuss.*, 2013. Advanced Article doi:10.1039/C2FD20110K

34. DeNicola, A.; Zhao, Y.; Kawakatsu, T.; Roccatano, D.; Milano, G. *J. Chem. Theory Comput.* 2011, *7*, 2947 – 2962.

35. Wang, Z.; Deserno, M.; *J. Phys. Chem. B,* 2010, *114*, 11207 – 11220.

36. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* 2005, *26*, 1781 – 1802.

37. MacKerell, A.; Banavali, N.; Foloppe, N. *Biopolymers*, 2000, *56*, 257 – 265.

38. Lyubartsev, A.; Laaksonen, A. *Comput. Phys. Commun.*, 2000, *128*, 565 – 589.

39. Lyubartsev, A. *Eur Biophys J*, **2005**, *35*, 53 – 61.

40. Izvekov, S.; Voth G. *J Phys Chem B.*, **2009**, *113*, 4443 – 4455.

41. Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.*, 2003, *24*, 1624 – 1636.

42. Chennamsetty, N.; Bock, H.; Gubbins, K. *J. Chem. Phys.,* 2006, *124*, 074105-1 – 074105-12.

43. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics*, 1996, *14*, 33 – 38.

44. Smith, S. W.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.,* 1997, *134*, 16 – 30.

45. Alder, BJ.; Wainwright, TE. *J. Chem. Phys.*, 1959, *31*, 459 – 466.

46. Rapaport, DC. *J. Phys. A: Math. Gen.*, 1978, *11,* L213 – L217.

47. Rapaport, DC. *J. Chem. Phys.*, 1979, *71*, 3299 – 3303.

48. Andersen, HC. *J. Chem. Phys.,* 1980, 72, 2384 – 2393.

49. Berendsen, H.; Postma, J.; Van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.*, 1984, *81*, 3684 – 3690.

50. Nguyen, H.D.; Hall, C.K. *Proc. Natl. Acad. Sci. U.S.A.*, 2004, *101*, 16180 – 16185.

51. Khan, M.; Herbordt, M. *J. Comput. Phys.*, 2011, *230*, 6563 – 6582.

52. Hill, T.L. *Introduction to Statistical Thermodynamics*; Addison Wesley Publishing Company, Inc.: Reading, MA, 1960.

53. Kucerka, N.; Tristram-Nagle, St.; Nagle, J. *Biophys. J.,* 2006, *90*, L83 – L85.

54. Kucerka, N.; Nagle, J.; Sachs, J.; Feller, S.; Pencer, J.; Jackson, A.; Katsaras, J. *Biophys. J.*, 2008, *95*, 2356 – 2367.

55. Tristra-Nagle, S.; Zhang, R.; Suter, R.; Worthington, C.; Sun, W.; Nagle, J. *Biophys. J.*, 1993, *64*, 1097 – 1109.

56. Marrink, S.; Risselada, J.; Mark, A. *Chem. Phys. Lipids*, 2005, *135*, 223 – 244.

57. Biltonen, R.; Lichtenberg, D. *Chem. Phys. Lipids*, 1993, *64*, 129 – 142.

58. Nagle, J.; Tristram-Nagle, S. *Biochim. Biophys. Acta*, 2000, *1469*, 159 – 195.

59. Wang, Z.; Deserno, M. *New J. Phys.*, 2010, *12*, 1 – 25.

60. Leekumjorn, S.; Sum, A. *Biochim. Biophys. Acta*, 2007, *1768*, 354 – 365.

61. Liu, J.; Conboy, J. *Biophys. J.*, 2005, *89*, 2522 – 2532.

62. Liu, J.; Conboy, J. *J. Am. Chem. Soc.*, 2004, *126*, 8376 – 8377.

63. John, K.; Schreiber, S.; Kubelt, J.; Herrman, A.; Muller, P. *Biophys. J.*, 2002, *83*, 3315 – 3323.

64. Norberg, J.; Nilsson, L. *Biophys. J.*, 2000, *79*, 1537-1553.

65. Kenkare, N.; Hall, C.; Khan, S. *J. Chem. Phys.,* 2000, *113*, 404 – 418.

66. Shultz, A.; Hall, C.; Genzer, J. *J. Chem. Phys.*, 2004, *120*, 2049 – 2055.

2.6 List of Tables

Table 2.1: The type, number of atoms, and mass for all of the coarse-grained sites in the LIME representation.

CG Site	CG Type	Atoms per CG Type	Mass of CG Type (amu)
1	I	$C_5H_{13}N$	87.2
2	II	PO_4	95.0
3	III	C_3HO_2	71.1
9	IV	$C_2H_2O_2$	58.0
4 – 7, 10 – 13	V	C_3H_6	42.1
8, 14	VI	C_3H_7	43.1

Table 2.2: The hard sphere diameters, square well widths and interaction energy for each pair of coarse-grained type.

Coarse-grained Type i	Coarse-grained Type j	$\sigma_{HS\ ij}$ (Å)	λ_{ij} (Å)	ε_{ij} (eV)
1	1	4.35	12.65	-0.065
1	2	3.85	9.85	-0.070
1	3	3.85	9.95	-0.050
1	4	3.75	10.25	-0.047
1	5	4.15	8.00	0.000
1	6	4.15	8.00	0.000
2	2	4.05	12.15	-0.080
2	3	3.45	13.05	-0.048
2	4	3.35	12.35	-0.030
2	5	3.75	8.15	0.000
2	6	3.65	6.40	0.000
3	3	3.65	10.45	-0.037
3	4	3.25	9.95	-0.036
3	5	3.65	10.53	-0.022
3	6	3.55	7.70	-0.015
4	4	3.15	11.65	-0.035
4	5	3.45	11.43	-0.026
4	6	3.45	10.85	-0.023
5	5	3.75	11.56	-0.050
5	6	3.75	11.66	-0.054
6	6	3.65	11.02	-0.070

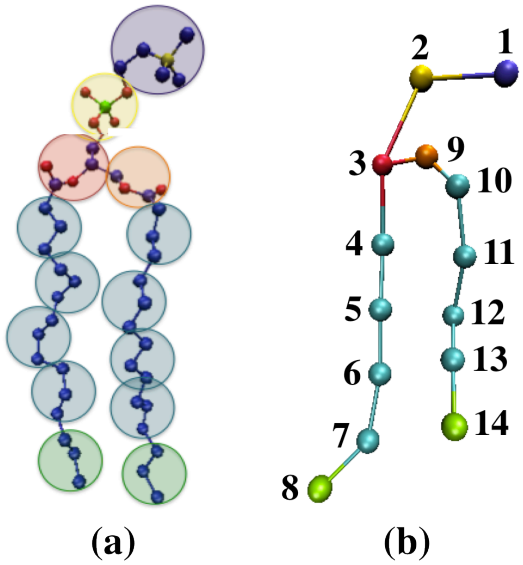


Figure 2.1: (a) United atom and (b) coarse-grained representation of DPPC. The color scheme is; purple (choline entity – type I for site 1); yellow (phosphate group – type II for site 2); red (ester group – type III for site 3); orange (ester group – type IV for site 9); cyan (alkyl tail groups – type V for sites 4-7 & 10-13); green (terminal tail groups - type VI for sites 8&14). The coarse-grained site size does not represent the actual size of each site.

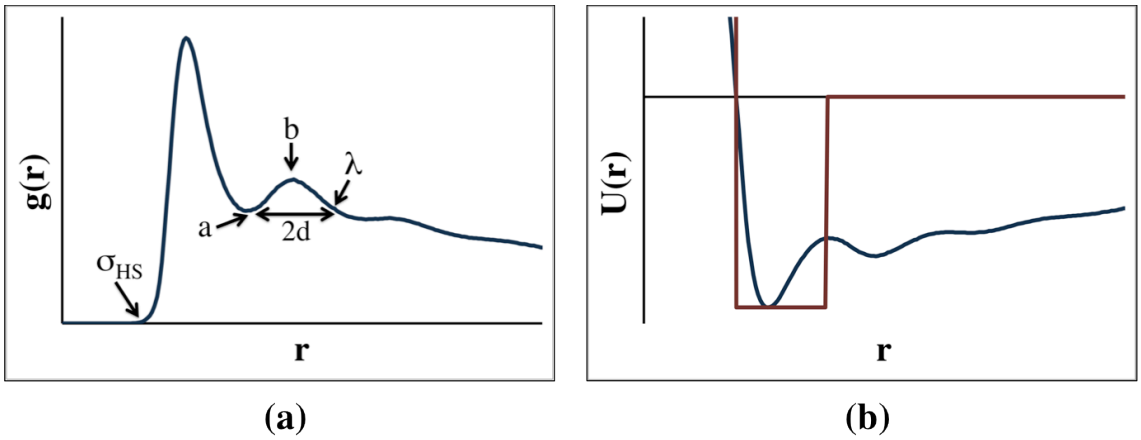


Figure 2.2: A schematic of the approach used to calculate the LIME interaction energies for two coarse-grained types: (a) the radial distribution function is calculated (b) the one-step Boltzmann inversion scheme is used to calculate the potential of mean force by inverting the RDF; ε is chosen as the minimum U(r) value (blue line); the depth of the square well potential or the interaction energy is assigned the ε value (red line)

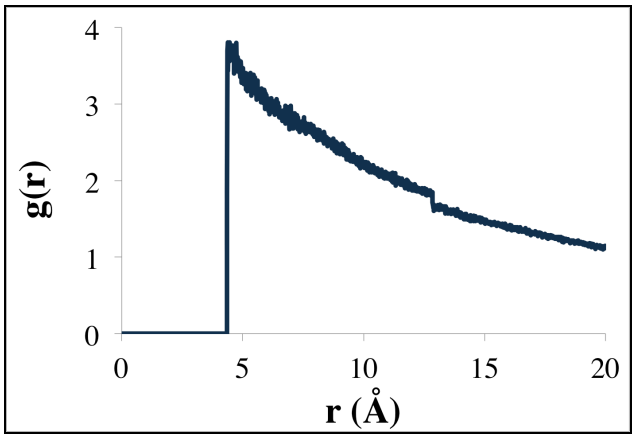


Figure 2.3: A radial distribution function for sites 1 & 1 obtained during a LIME simulation. The first non-zero value is located at the hard sphere diameter (4.75 Å) and the small discontinuity is located at the square-well width (12.55 Å). The shape of the radial distribution function differs significantly from the shape of a distribution function associated with a more traditional Lennard Jones potential.

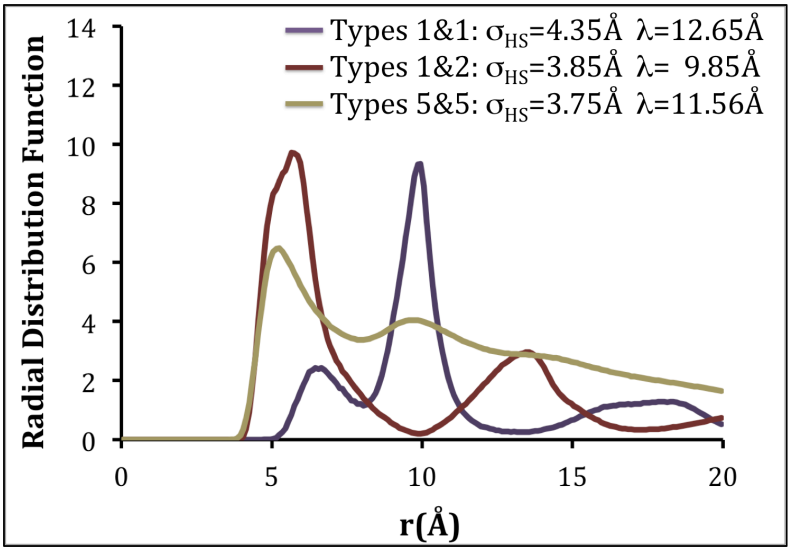


Figure 2.4: (a) The intermolecular radial distribution functions, hard-sphere diameters (σ_{HS}) and square-well diameters ($\boldsymbol{\lambda}$) for coarse-grained types 1 & 1, 1 & 2, and 5 & 5.

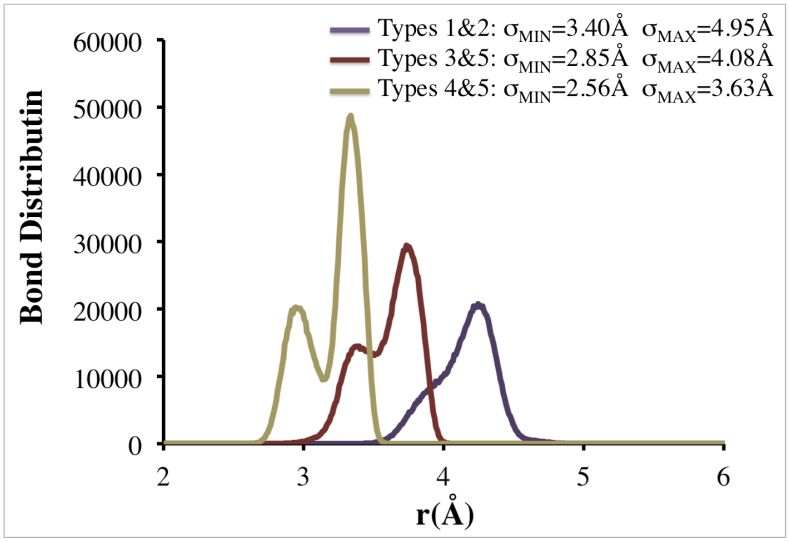
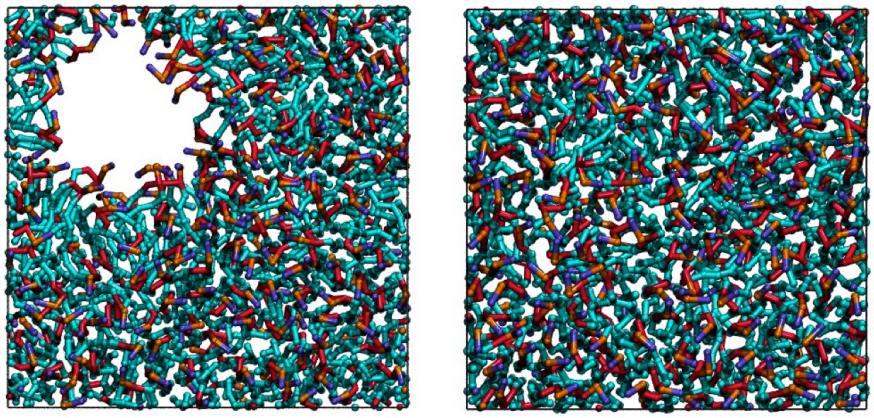


Figure 2.5: The intramolecular bond distribution functions, minimum bond lengths (σ_{MIN}) and maximum bond lengths (σ_{MAX}) for coarse-grained types 1 & 2, 3 & 5, and 5 & 6



(a) (b)

Figure 2.6: Snapshots of the areal view of a DPPC bilayer formed from 256 lipids in a box with dimensions of 100Å x 100Å x 100Å (a) and in a box with dimensions of 90Å x 90Å x 90Å (b)

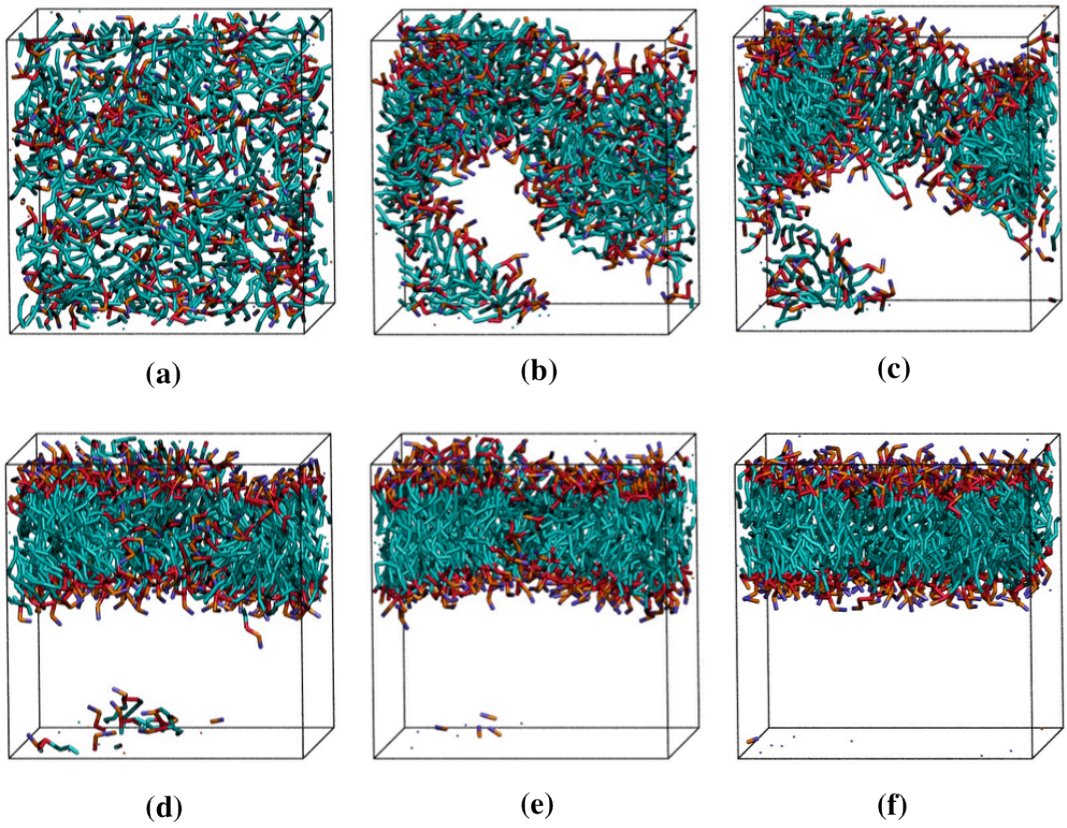


Figure 2.7: Snapshots from a simulation of DPPC spontaneous bilayer formation. The color scheme is: purple (choline entity – type I); orange (phosphate group – type II); red (ester groups – type III and type IV); cyan (alkyl tail groups – type V and type VI). (a) – (f) = 0, 15, 50, 100, 125 and 150 million collisions, respectively. The system is started from a random configuration (a) and aggregates in only 50 million collisions (d). An additional 100 million collisions are required for the aggregate to adopt the conformation of a defect-free bilayer (h).

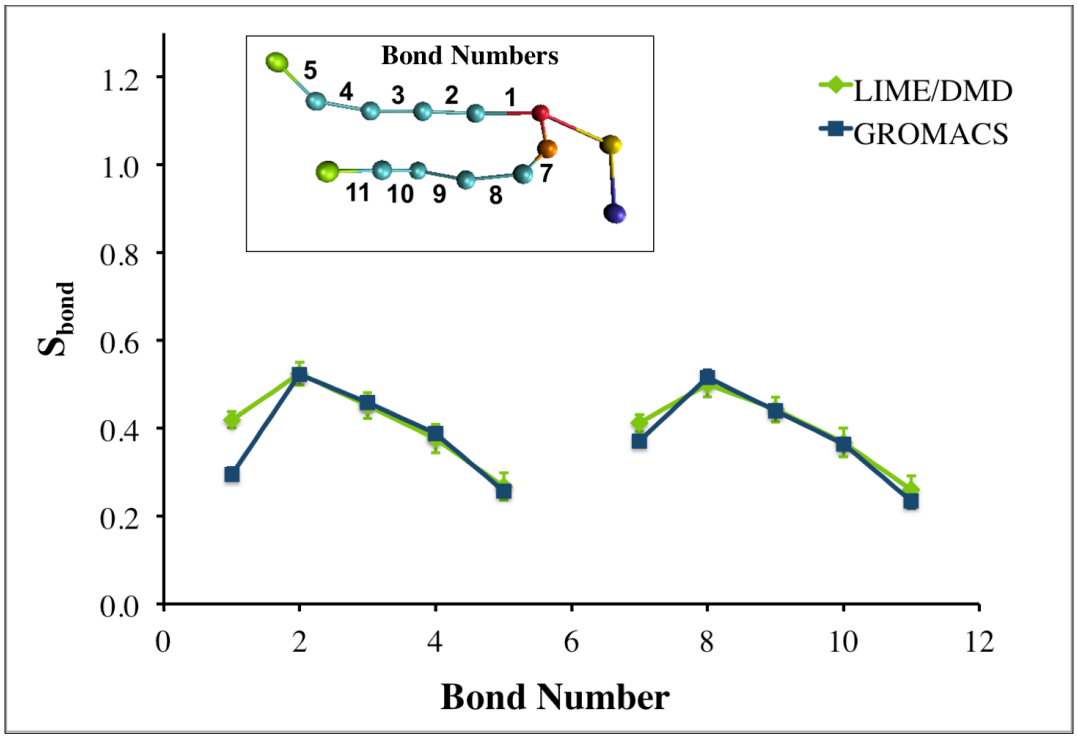


Figure 2.8: Comparison of the orientational bond order parameter S_{BOND} for intra-molecular bonds in LIME/DMD (green line) and GROMACS simulations (blue line) versus the bond number: the latter is defined in the inset.

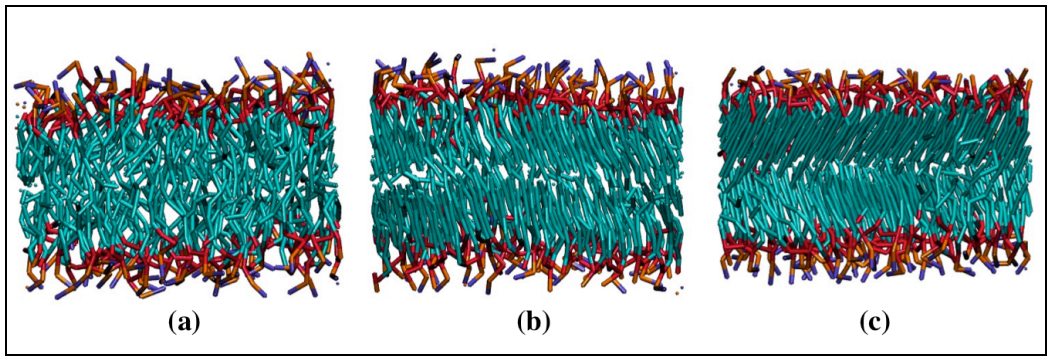


Figure 2.9: Snapshots of a lipid bilayer in DMD/LIME as the system temperature is cooled from (a) a liquid-crystalline phase at T* = 0.77, (b) a tilted gel phase at T* = 0.30 and (c) a cross-tilted gel phase which is only observed in some simulations.

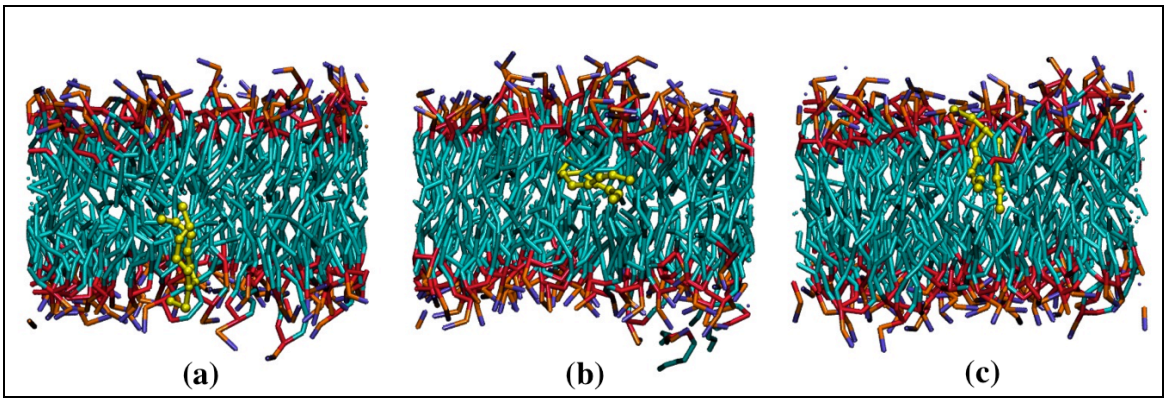


Figure 2.10: Snapshots of a lipid (spherical representation) as it flips from the bottom leaflet of a bilayer to the top leaflet. The tail beads of the lipid that flips are highlighted in yellow and the head beads are highlighted in lime. (a) – (c) = 866, 883, 885 million collisions, respectively.

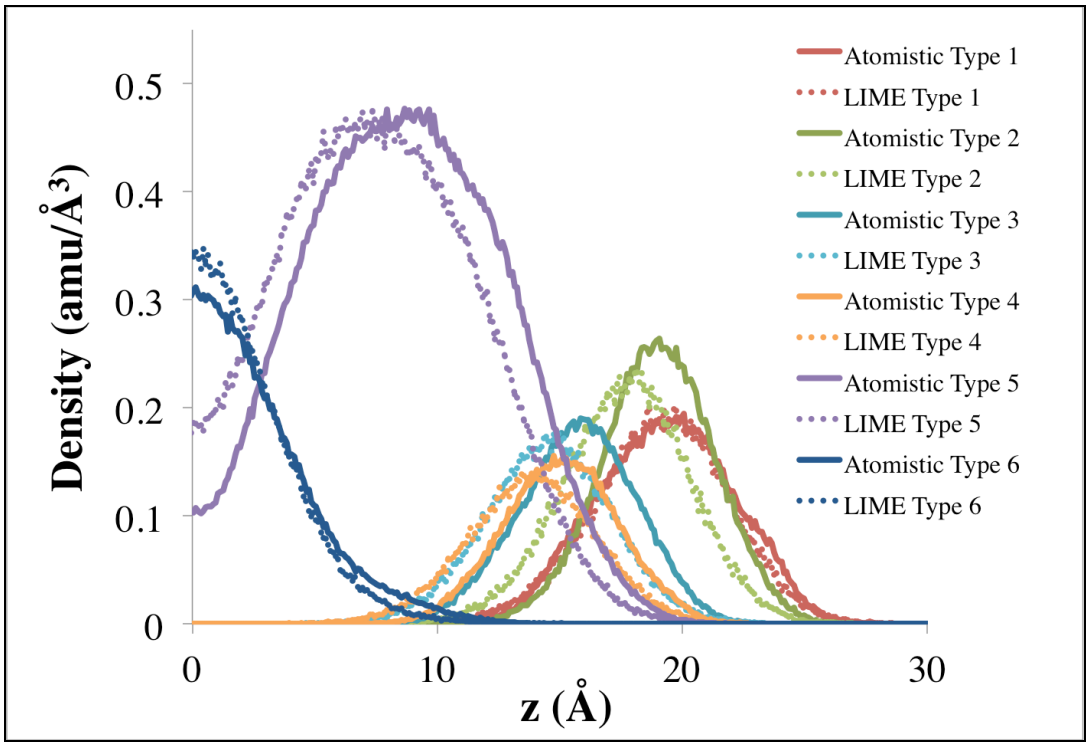


Figure 2.11: Mass density distribution of coarse-grained sites in DMD/LIME simulations (dotted lines) and GROMACS simulations (solid lines) versus the distance from the bilayer center (z = 0 Å).

CHAPTER 3

The Extension of LIME to Model the Phase Separation Behavior of Mixed Lipid Systems at Neutral and Low pH

3.1 Introduction

Phospholipids are amphiphilic molecules that spontaneously form bilayers in aqueous solution to minimize the interaction between their hydrophobic tails and to maximize the interaction between their hydrophilic heads. In recent years, scientists have been working to exploit the tunable properties of mixed lipid systems for a range of applications including imaging [1], biosensors [2,3], and membrane trafficking [4,5]. In particular, mixed lipid systems are being used as novel liposomal delivery carriers for drugs [6,7], vaccines [8], genes [9,10] and antimicrobial agents[11,12]. While scientists have successfully demonstrated that liposomal formulations improve drug efficacy [13,14], research efforts are still dedicated to optimizing their structure and function. For example, Karve and co-workers showed that liposomes composed of mixed lipid membranes could become permeable to encapsulated drug molecules in response to a decrease in pH.[15] Mixed lipid systems are also being used to create temperature-sensitive liposomes that can rapidly and efficiently release drug molecules in response to heat.[16,17,18,19] In order to fully realize the potential of these new technologies, a better understanding of the structure and interactions between lipids on a molecular level at various physiological conditions is needed. This would enable researchers to design and develop liposomes with optimal properties to release drug molecules in response to specific stimuli. An approach commonly used to gain insight into the behavior of such systems on a molecular level is computer simulation.

In a previous paper we demonstrated how computer simulations of lipid systems can complement experimental work by providing detailed molecular-level information regarding their dynamics and self-assembly.[20] In that work we introduced LIME (Lipid Intermediate Resolution Model), a new implicit solvent lipid force field for use with discontinuous molecular dynamics (DMD), that enabled the rapid simulation of 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) lipids in water. The structural properties of the DPPC bilayer formed during our LIME/DMD simulations including the area per lipid, bilayer thickness, bond order and mass density profiles were in agreement with experimental observations or atomistic simulation results.

In this paper we take a multiscale modeling approach to expand LIME to include parameters for: (1) 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), (2) 1,2-dihenarachidoyl-sn-glycero-3-phosphocholine (21PC), and (3) 1,2-distearoyl-*sn*-glycero-3-phospho-L-serine (DSPS) at 310K. DPPC and 21PC are essentially the same; they both have a choline headgroup and differ only in the lengths of their alkyl tails. We demonstrate how discontinuous molecular dynamics (DMD) simulations with this expanded LIME forcefield can be used to model the behavior of lipid systems containing large numbers of DPPC, DSPS, and 21PC and mixtures of these lipids over long time scales at a neutral pH and at a pH of 5.5. Throughout this paper we define a neutral pH as 7.0 – 7.4 and low pH as pH < 5.5. We show that LIME/DMD simulations can be used to study the phase separation of mixed lipid systems into heterogeneous domains on a molecular level.

A number of different approaches to modeling mixed lipid systems with simulations have been described in the literature. Various levels of detail are used in these models, which

can be divided roughly into two categories: high-resolution and low-resolution models. High-resolution or atomistic models are based on a realistic representation of membrane geometry and energetics and typically account for the motion of every atom including every solvent atom. Jiang and co-workers performed GROMACS [21] atomistic simulations on bilayer ribbons containing dimyristoylphosphatidylcholine (DMPC), dihexanoylphosphatidylcholine (DHPC) and didecanoyl PC (DDPC) to investigate how each species partitions between the flat and curved microenvironments of the bilayer and the role DHPC plays in stabilizing the bilayer edge.[22] In another study, Hall and co-workers performed atomistic molecular dynamics simulations to investigate the role of glycolipids in lipid rafts; systems of galactosylceramide (GalCer), cholesterol (CHOL), palmitoyloleoyl phosphatidylcholine (POPC) and palmitoyl-sphingomyelin (PSM) were simulated using GROMACS.[23] Addition of GalCer increased the thickness of the raft membranes and induced changes in the lateral diffusion of raft lipids, but did not influence the average area per lipid or the lipid conformation order.[23] In another atomistic simulation, Pandit and co-workers studied mixed bilayers containing a 5:1 ratio of DPPC and DSPS in NaCl electrolyte solutions to learn how DPPC and DSPS interact to form complexes.[24] GROMACS was used to perform all simulations based on force field parameters from the work of Berger.[24] These simulations showed that NH□□□O and CH□□□O hydrogen bonding between lipids serves as the basis of interlipid complexation and that DPPC alone is less likely to form interlipid complexes than in the presence of bound ions or DSPS.[24]

Low-resolution models, which are based on a simplified representation of molecular geometry and energetics, have also been used to study the behavior of mixed lipid systems.

For example, Illya and co-workers developed a low-resolution model designed for use with dissipative particle dynamics simulations to study the domain formation and the material properties of a two-component amphiphile bilayer membrane.[25] In this solvent-explicit model the amphiphile geometry was chosen to represent the general shape of a typical phospholipid. The amphiphile $H_3(T_6)_2$ was represented by 3 head spheres and 2 tails each composed of 6 spheres, and the amphiphile $H_3(T_8)_2$ was represented by 3 head spheres and 2 tails each composed of 8 spheres.[25] The interaction parameters for the amphiphiles were chosen to mimic the properties of lipid membranes, not calculated from atomistic simulations of amphiphiles.[25] The authors calculated the area stretch modulus and bending rigidity of the mixed amphiphilic bilayers and vesicles but could not compare their results directly to any experimental observations on a specific lipid mixture since the amphiphiles did not represent any specific lipids.

Another way to model mixed lipid systems in simulations is to use intermediate-resolution (coarse-grained) models, which represent the geometry and energetics of lipids at a level of detail that is in between high resolution and low resolution models. This allows for the simulation of specific lipids while requiring shorter simulation times than atomistic models. In coarse-grained models a single interaction site represents several atoms. This reduces the total number of sites whose trajectories must be calculated and helps to increase the speed of the simulation. One example of an intermediate-resolution explicit-solvent model is that of Faller and Marrink.[26] In this model, distearoylphosphatidylcholine (DSPC) and dilauroylphosphatidylcholine (DLPC) are represented by 14 and 10 coarse-grained sites, respectively. The coarse-grained sites interact via a Lennard Jones potential

and a Colulomb interaction is used to model electrostatics.[26] Lipid mixtures composed of DLPC and DSPC were simulated to study the phase separation that occurs between these two lipids at phase transition temperatures such that the longer lipid enters the gel phase and the shorter lipid remains in the fluid phase.[26] Simulations were performed using GROMACS with the interaction parameters previously described by Marrink and co-workers [27]. All simulations were started from a bilayer composed of randomly assigned DSPC and DLPC lipid. Results showed that the higher the ratio of the number of longer lipid (DSPC) to the number of shorter lipid (DLPC), the higher the gel-fluid transition temperature. In another intermediate resolution, explicit-solvent model, Risselada and Marrink performed simulations of the spontaneous separation of a randomized mixture composed of dipalmitoyl-phosphatidylcholine (diC_{16}-PC), dilinoleyl-phosphatidylcholine (di$C_{18:2}$-PC) and cholesterol into a liquid-ordered and a liquid-disordered phase.[28] Simulations of a planar membrane composed of approximately 2000 lipid molecules and a liposome composed of approximately 3000 lipid molecules were performed. In this model, on average, 4 heavy atoms were represented by a single coarse grained site and approximately 3 heavy atoms were used to represent the rings in cholesterol.[28] The simulations were performed with the GROMACS simulation package and the MARTINI CG force field, version 2.0.[29]

Intermediate resolution implicit-solvent models of mixed lipid systems have also been described in the literature. For example, Lu and Voth used a multiscale coarse-graining approach to develop a solvent-free coarse-grained model for a mixed bilayer composed of dioleoylphosphatidylcholine (DOPC) and dioleoylphosphatidyethanolamine (DOPE).[30] In this model DOPC and DOPE are each represented by 15 coarse-grained sites. The results of

the multiscale coarse-grained model, including radial distribution functions, bilayer thickness and order parameters, show good agreement with atomistic results.[30]

Our motivation for studying DPPC/DSPS and 21PC/DSPS lipid systems is experimental work by Sofou and co-workers aimed at developing liposome-based drug delivery devices for the treatment of cancer.[31,32] They examined the phase separation behavior of pH-triggered liposomes composed of cholesterol and two different types of phospholipids: 21PC phospholipids, which have a neutral headgroup at both neutral (pH 7.4 – 7.0) and endosomal pH (pH 5.5 – 5.0), and DSPS phospholipids which have a negatively charged headgroup at neutral pH and a neutral headgroup at endosomal pH.[31,32] They showed that at neutral pH the vesicles were much less permeable to encapsulated drug molecules than at endosomal pH. At endosomal pH, the cholesterol, 21PC and DSPS lipids allowed much more of the encapsulated drug molecules to leak out of the liposome than at neutral pH. This observation forms the basis of Sofou and coworkers proposed liposomal-based drug delivery scheme in which liposomes composed of cholesterol, 21PC and DSPS are used to transport cancer drugs throughout the body at neutral pH. When the liposomes reach the endosome in cancer cells, the decrease in pH triggers the formation of heterogeneous domains and the encapsulated drug molecules leak out.[31,32] Sofou's pH tunable liposomes have additional features including antibodies to target cancer cells.

In addition to studying the phase separation behavior of liposomes composed of cholesterol, 21PC and DSPS, Sofou and co-workers also studied liposomes composed of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and DSPS with and without cholesterol.[33]

They found that liposomes composed of DOPC, DSPS and cholesterol had a higher level of phase separation at a pH of 5.0 than at a pH of 7.0. In contrast, liposomes composed of DOPC and DSPS without cholesterol had the same level of phase separation at pH values of 7.0, 6.0 and 5.0. Sofou and co-workers did not perform any experiments on liposomes composed of DPPC/DSPS or 21PC/DSPS without cholesterol. Since the simulations presented here are essentially a proof of concept study to see if LIME/DMD can predict phase separation in mixed lipid systems, we chose to model the simplest possible mixed liposomes that were close to the experiments of Sofou and co-workers on DPPC/DSPS or 21PC/DSPS liposomes. For this reason we did not include cholesterol.

In this paper, we describe the expansion of "LIME", the implicit-solvent, coarse-grained model previously used to simulate the spontaneous formation of DPPC bilayers at 325K, to DPPC, 21PC, and DSPS lipids at 310K.[20] We analyze the extent of phase separation that occurs during simulations at both neutral and low pH in DPPC/DSPS and 21PC/DSPS bilayers containing 128 lipids with equimolar ratios. We also study the phase separation that occurs at low pH during simulations of DPPC/DSPS and 21PC/DSPS bilayers containing 128 lipids and molar ratios of 0.25:0.75and 0.75:0.25. Finally, we simulate the behavior of both DPPC/DSPS liposomes and 21PC/DSPS liposomes containing 30 doxorubicin drug molecules at low pH and molar ratios of 0.5:0.5. The LIME geometrical and energetic parameters for DPPC, 21PC and DSPS at low pH and 310K are calculated using a multi-scale modeling approach based on evaluating radial distribution functions obtained from an atomistic simulation of a fully hydrated bilayer composed of 64 DPPC lipids and 64 protonated DSPS lipids performed with the GROMACS simulation package

and the GROMOS96 53a6 forcefield.[21,34,35] The same method is used to extract parameters at neutral pH; the only difference is that the DSPS lipids are unprotonated.

Highlights of our results include the following: LIME/DMD simulations of bilayers composed of either DPPC/DSPS or 21PC/DSPS show similar phase separation rates at both neutral and low pH. This observation is supported by experimental data from the Sofou lab on the phase separation rates for liposomes composed of DOPC and DSPS at a pH of 5.0, 6.0 and 7.0.[33] Our simulation results on equimolar bilayers show the following. 21PC/DSPS bilayers separate slightly faster at low pH than at neutral pH, but DPPC/DSPS bilayers separate at approximately the same rate at neutral and low pH. 21PC/DSPS bilayers separate slightly faster than equimolar DPPC/DSPS bilayers. Low pH DPPC/DSPS bilayers without any surface area restrictions separate faster than those with restrictions but surface area restrictions on equimolar low pH 21PC/DSPS bilayers did not affect the separation rate. Simulations of DPPC/DSPS and 21PC/DSPS bilayers with different molar ratios of PC:PS lipids showed that the higher the concentration of PS lipids, the faster the separation rate. Simulations of DPPC/DSPS and 21PC/DSPS liposomes containing doxorubicin showed domain formation in both types of liposomes. However, no drug molecules escaped from either type of liposome after 1.5 billion collisions.

3.2 Methods and Model

In LIME 10 different coarse-grained types (I to X) are used to represent the groups of atoms that make up the DPPC, 21PC and DSPS molecules. DPPC and 21PC are similar lipids in that they have the same polar head group and differ only in the lengths of their alkyl

tails. They are each composed of a polar head group that includes a choline, phosphate and two ester linkages, and two nonpolar hydrophobic acyl tails. DSPS is composed of a head group that includes a carboxyl, amine and phosphate group and two ester linkages, and two nonpolar hydrophobic acyl tails. At a neutral pH, the net charge carried by the DSPS head group is negative. This negative charge causes electrostatic repulsion between the DSPS lipids which forces them to be distributed evenly throughout a bilayer membrane or liposome.[31] At a pH of approximately 5.5[33], the carboxyl group on DSPS becomes protonated and loses its negative charge. When the DSPS molecules lose their negative charge the electrostatic repulsion between these lipids stops preventing them from interacting, allowing them to form hydrogen bonds between their protonated amino groups and deprotonated phosphate groups on the head groups.[31,36] Therefore, at a low pH, DSPS lipids have a very strong attraction for each other. **Figure 3.1** shows the structure of (a) a DPPC lipid, and (b) a DSPS lipid at a neutral pH. At a low pH the oxygen atom that is circled becomes protonated.

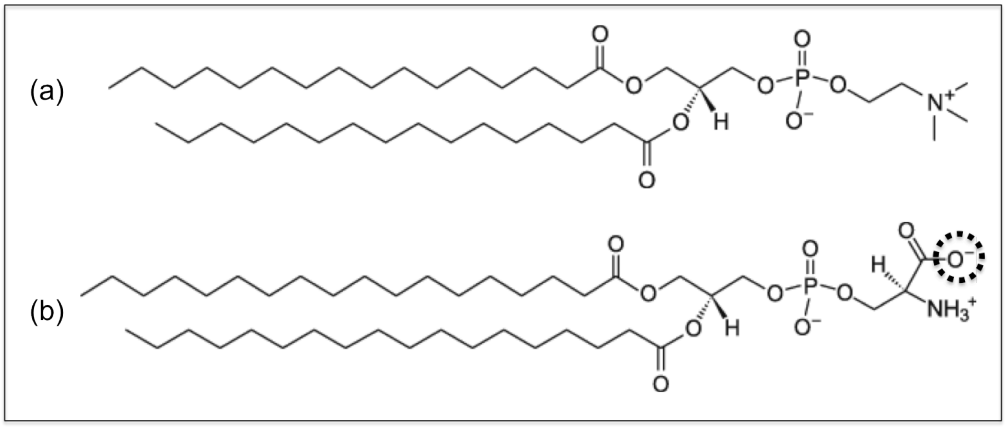


Figure 3.1: Structure of (a) DPPC and (b) DSPS at a neutral pH. At a low pH the oxygen atom in the dotted circle becomes protonated.

Figure 3.2 illustrates the coarse-graining of: (a) a DPPC molecule from 50 united-atoms to 14 coarse-grained sites, (b) a 21PC molecule from 60 united atoms to 18 coarse-grained sites, and (c) a DSPS molecule from 57 united-atoms to 17 coarse-grained sites. In LIME, each coarse-grained site is assigned a different coarse-grained type, which is represented by a unique color. This figure and all other figures depicting lipid molecules throughout the paper were generated with Visual Molecular Dynamics (VMD).[37] Type I (purple) represents the choline entity on DPPC and 21PC. Type II (yellow) represents the phosphate groups on DPPC and 21PC. The phosphate group on DSPS is represented by type IX (yellow); this is a different type than the phosphate groups on DPPC and 21PC because the energy parameters calculated for this coarse-grained site were significantly different from those calculated for the DPPC and 21PC phosphate groups. Ester coarse-grained sites are assigned types III (red) and IV (orange), respectively. Two coarse-grained types for the ester are needed because coarse-grained III has one more carbon and hydrogen atom than coarse-

grained type IV. Coarse-grained type V (cyan) is used to represent the tail sites (each contains 3 carbons and 6 hydrogens) of DPPC, 21PC and DSPS. Coarse-grained type VI (lime) is assigned to the terminal tail group of DPPC representing 3 carbon and 7 hydrogen atoms and coarse-grained type X (gray) is assigned to the terminal tail groups of 21PC and DSPS representing 2 carbon and 5 hydrogen atoms. Coarse-grained type VIIa (black) is used to represent the unprotonated carboxyl head group and type VIIb (black) is used to represent the protonated carboxyl head group at DSPS site 1. **Table 3.1** lists the atoms included in each coarse-grained site, the type assigned to each coarse-grained site and the mass of each coarse-grained site.

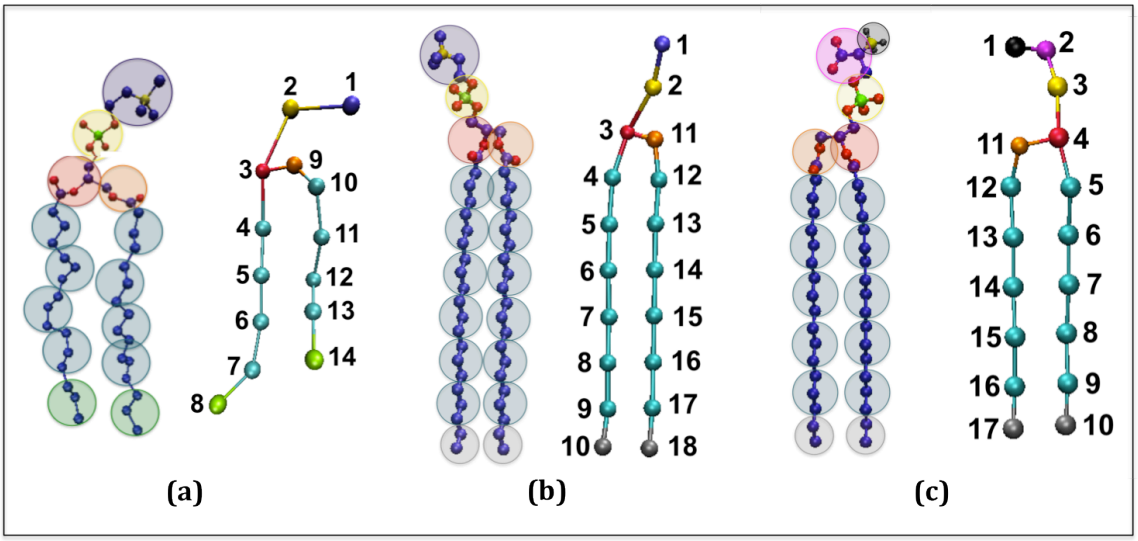


Figure 3.2: United atom and coarse-grained representations of (a) DPPC. (b 21PC. (c) DSPS. The color scheme is; purple (choline entity – type I for DPPC site 1 and 21PC site 1); yellow (phosphate group – type II for DPPC site 2, 21PC site 2); red (ester group – type III for DPPC site 3, 21PC site 3 and DSPS site 4); orange (ester group – type IV for DPPC site 9, 21PC site 11 and DSPS site 11); cyan (alkyl tail groups – type V for DPPC sites 4-7 & 10-13, 21PC sites 4-9 & 12-17 and DSPS sites 5-9 & 12-16); green (terminal tail groups – type VI for DPPC sites 8&14); black (negatively charged carboxyl group – type VII a and protonated carboxyl group – type VII b for DSPS site 1); magenta (amine group – type VII for DSPS site 2); yellow (phosphate group – type IX for DSPS site 3); grey (terminal tail groups – type X for 21PC sites 10&18 and DSPS sites 10&17). The protonated carboxyl group located at DSPS site 1 (not shown) is assigned type VIII.

Table 3.1: The type, number of atoms, and mass for all of the coarse-grained sites in the LIME representation.

CG Type	CG Type Color	CG Site	Atoms per CG Type	Mass of CG Type (amu)
I	Purple	DPPC 1, 21PC 1	$C_5H_{13}N$	87.2
II	Yellow	DPPC 2, 21PC 2	PO_4	95.0
III	Red	DPPC 3, 21PC 3, DSPS 4	C_3HO_2	71.1
IV	Orange	DPPC 9, 21PC 11, DSPS 11	$C_2H_2O_2$	58.0
V	Cyan	DPPC 4-7&10-13 21PC 4-9&12-17 DSPS 5-9&12-16	C_3H_6	42.1
VI	Lime	DPPC 8&14	C_3H_7	43.1
VII a	Black	DSPS 1	$CO0^-$	44.0
VII b	Black	DSPS 1	COOH	45.0
VIII	Magenta	DSPS 2	$C_2H_3NH_3$	44.1
IX	Yellow	DSPS 3	PO_4	95.0
X	Gray	21PC 10&18, DSPS 10&17	C_2H_5	29.1

The simulation method used in this paper is the discontinuous molecular dynamics (DMD) algorithm, which is a very fast alternative to traditional molecular dynamics simulation that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere and square-well potentials.[38,39] For this reason, all of the inter- and intra- molecular interactions in our lipid model are represented by a combination of hard-sphere and square-well potentials, as opposed to the Lennard Jones, Coulombic and harmonic potentials found in traditional molecular dynamics simulations. A hard sphere is an impenetrable, solid sphere; a square-well is a hard sphere surrounded by an attractive well.

Expressions for the hard sphere (HS) and square well (SW) potentials between spheres *i* and *j* are, respectively:

$$u_{ij}^{HS}(r) = \begin{cases} \infty & r \le \sigma_{ij} \\ 0 & r > \sigma_{ij} \end{cases} \qquad \text{Equation 1}$$

$$u_{ij}^{SW}(r) = \begin{cases} \infty & r \le \sigma_{ij} \\ -\varepsilon_{ij} & \sigma_{ij} < r \le \lambda_{ij} \\ 0 & r > \lambda_{ij} \end{cases} \qquad \text{Equation 2}$$

where r is the distance between spheres, σ_{ij} is the hard sphere diameters, λ_{ij} is the well diameter and ε_{ij} is the well depth. These parameters define the strength of the interaction between the coarse-grained spheres. Unlike continuous potentials, such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide. This makes the simulation an efficient event-scheduling algorithm, which allows sampling of longer time scales and larger systems than traditional molecular dynamics.

A DMD simulation proceeds in the following way. Molecules are initially placed in a random configuration consistent with excluded volume and angular constraints. Initial velocities are assigned based on a Maxwell-Boltzmann distribution about the desired simulation temperature. Particle trajectories are followed by calculating the time between each collision and advancing the simulation to the next event. Types of events include a collision between two hard spheres, a bond event when the distance between two bonded spheres reaches a minimum or maximum limit, and square well events when two spheres enter (capture), unsuccessfully attempt to escape (bounce) or successfully leave (dissociation) a square well.[38,39,40,41]

In all LIME/DMD simulations the simulation temperature is expressed in terms of the reduced temperature T*:

$T^* = k_B T/\varepsilon^*$ Equation 3

where k_B is Boltzmann's constant, T is the temperature, and ε* is the reference interaction strength.[42] The reference interaction strength, ε*, was calculated using:

$$\varepsilon^* = \frac{\sum_{ij} n_{ij}\varepsilon_{ij}}{\sum_{ij} n_{ij}}$$ Equation 4

where n_{ij} is the total number of pairs of coarse-grained sites of type i (on one molecule) and type j (on another molecule) and ε_{ij} is the interaction energy between coarse-grained types i and j. The neutral and low pH ε_{ij} values were obtained from separate GROMACS simulations at neutral and low pH at T=310K. The resulting value for ε* calculated from **Equation 4** was 0.033 for both the neutral and low pH GROMACS simulations. Therefore, when $T^*=k_B T/\varepsilon^*=(8.6173\times10^{-5}\text{eV/K})*(310\text{K})/(0.033\text{eV}) = 0.8$ in our neutral and low pH DMD/LIME simulations, the lipid molecules behave as they would at a real temperature of 310K. In our previous work, LIME/DMD simulations at a value of T*=0.8 corresponded to a real temperature of 325K because the GROMACS simulations were performed at 325K.[20] In this work, T* = 0.8 corresponds to 310K because we are using interaction energies obtained from a GROMACS simulation at 310K. Thus, these interaction energies are different than those used in our previous work. The temperature is maintained constant using the Andersen thermostat, which uses ghost collisions with randomly selected particles

in the system to maintain the Maxwell-Boltzmann velocity distribution about the desired temperature.[43] Periodic boundary conditions are employed for all DMD/LIME simulations. In this work, all the simulations were performed in the canonical ensemble (NVT).

The coarse-grained parameters in LIME were obtained using a multiscale modeling technique. Multiscale modeling is a method in which several atoms are grouped into a single coarse-grained site and the parameters used to represent the behavior of this coarse-grained site are extracted from atomistic simulations. Data used to calculate the LIME parameters was obtained by running united-atom explicit-solvent simulations at T=310K using the GROMACS simulation package [1,21] version 4.5.4 along with the GROMOS96 53a6 [34] forcefield. Since the Sofou Lab conducts experiments at 310K to study how liposomes would behave in the body, we parameterized our model from GROMACS simulations at 310K. The Kukol DPPC3 topology [35] was used to model DPPC in all GROMACS simulations. In this topology, the partial charges for the lipid head group are taken from Chiu et al. [44], who calculated these values using the GAUSSIAN 92 program. The DSPS topology used to model DSPS in all GROMACS simulations was based on the Kukol DPPC3 topology. For DSPS at a neutral pH, the partial charges of the phosphatidylserine headgroups were replaced by the partial charges calculated by Pandit and Berkowitz with Gaussian 98.[45] DSPS molecules protonated at their terminal carboxyl sites, were used in all low GROMACS simulations. The partial charges of the protonated carboxyl group on the DSPS molecule were obtained using GAUSSIAN 09.[46]

The LIME low pH intermolecular σ_{ij}, λ_{ij}, and ε_{ij} between coarse-grained types I-VI, VIIb and VIII-X were obtained by analyzing results from a low pH GROMACS simulation of a mixed bilayer composed of 64 DPPC lipids and 64 DSPS lipids in explicit solvent. The neutral pH intermolecular values for σ_{ij}, λ_{ij}, and ε_{ij} between coarse-grained types I-VI, VIIa and VIII-X were calculated from a neutral pH GROMACS simulation composed of 64 DPPC lipids and 64 non-protonated DSPS lipids. Since the neutral pH σ_{ij} values were not significantly different than the low pH σ_{ij} values, we decided to use the low pH σ_{ij} values for simulations at both pH values. Overall, the neutral pH and low pH λ_{ij}, and ε_{ij} values were similar, however, there were several coarse-grained types for which these parameters varied by more than 25%. For this work we decided to use the low pH λ_{ij}, and ε_{ij} values calculated from the low pH GROMACS simulation for low pH LIME/DMD simulations The neutral pH λ_{ij}, and ε_{ij} values calculated from the neutral pH GROMACS simulation were used for the neutral pH LIME/DMD simulations. Each GROMACS simulation was started from an initially homogeneous bilayer in a box with dimensions of 64 Å x 64 Å x 90 Å and was run at a temperature of 310K and a pressure of 1.0 bar for 20 ns. Throughout both the neutral and low pH simulations the bilayers remained intact. There was no evidence of phase separation between the two different lipids likely because the simulation time scale is short to the time scale for phase separation. The net charge of all GROMACS simulations was neutralized with the addition of NaCl. The Berendsen thermostat[47] was used to keep the temperature constant with a time constant of 0.1 ps. Throughout each GROMACS simulation the coordinates of each atom were written to an output trajectory file every 1 ps. These output files were used to calculate the centers of mass for each of the coarse-grained

sites. The LIME intramolecular σ_{ij}, λ_{ij}, and ε_{ij} values were set to the same values as the intermolecular values.

The multiscale modeling procedure that we used to obtain the new LIME parameters for DPPC, 21PC and DSPS is described in detail in our previous work.[20] In this multiscale modeling procedure the LIME parameters are calculated from data collected from the GROMACS simulations. The depth of square well interactions is calculated using a one-step Boltzmann inversion approach.

The minimum and maximum bond lengths in LIME were also determined from the GROMACS simulation data. The minimum bond length (σ_{MIN}) was chosen as the smallest possible distance between two bonded coarse-grained sites. The maximum bond length (σ_{MAX}) between coarse-grained sites was chosen as the largest possible distance for which a non-zero g(r) was observed. The distribution of σ_{MIN} and σ_{MAX} values for each set of coarse-grained types was evaluated and the most restrictive σ_{MIN} and σ_{MAX} was selected for each pair of coarse-grained types. The most restrictive σ_{MIN} and σ_{MAX} permitted the smallest amount of bond fluctuation for a pair of coarse-grained types. The bond and pseudobond values calculated from the neutral pH GROMACS simulations were very similar to those calculated from the low pH GROMACS simulations. Therefore, we choose to approximate the σ_{MIN} and σ_{MAX} values for interactions between coarse-grained types at a neutral pH with those calculated at a low pH.

In LIME the stiffness of each lipid molecule is maintained by imposing pseudobonds, which limit the fluctuation of coarse-grained sites to the angles and torsional angles observed during the GROMACS simulations. Bond angles are maintained by imposing pseudobonds

between all next-nearest neighboring sites. Torsional angles are maintained with pseudobonds between all next-next-nearest neighboring sites. The minimum and maximum pseudobond lengths were determined from the radial distribution functions for the coarse-grained sites. The minimum pseudobond length for a pair of coarse-grained sites was chosen as the smallest possible distance for which a non-zero g(r) was detected. The maximum pseudobond length for a pair of coarse-grained sites was chosen as the largest distance at which a non-zero g(r) was observed. The pseudobonds that were calculated for each pair of coarse-grained sites were evaluated and the most restrictive pseudobonds for each pair of coarse-grained types were selected.

EMBLEM, a C++ program developed in the Hall research lab was used to run all LIME/DMD simulations. EMBLEM uses the discontinuous molecular dynamics (DMD) algorithm to simulate the behavior of any type of molecule or mixture of molecules. Any DMD force field parameters can be used with EMBLEM. An Intel compiler was used to compile this code. EMBLEM is run with several efficiency techniques. Work is in progress to parallelize portions of EMBLEM, however, all of the simulations run for this work were done in serial.

3.3 Results

LIME/DMD simulations were run to study the behavior of bilayers composed of DPPC/DSPS and of 21PC/DSPS lipids. Simulations were run on ten systems, referred to here as Systems 1 through 10, which differed in the lipid species considered, the molar ratio of lipids, the pH and box length. The ten systems are summarized in **Table 3.2**. Three

replicate simulations, each composed of a total of 128 lipids, were run at a T* = 0.8 for each system. Each simulation was started from a preformed bilayer in which the lipids were randomly placed into an area of 61Å x 61Å. In simulations on Systems 1 through 8 the length of the simulation cell in the plane of the bilayer was 61Å x 61Å. In simulations on Systems 9 and 10, the bilayers which spanned an area of 61Å x 61Å, were placed in the center of a box with dimensions of 120Å x 120Å x 120Å. This was done to prevent the bilayer from interacting with its periodic images. While interacting with its periodic images would not force the bilayer to occupy a certain area, it might promote the lipids to maintain an area that would maximize their interaction with each other. In this section we use PC to refer to either a DPPC or 21PC molecule and PS to refer to a DSPS molecule.

Table 3.2: The type of lipids, molar ratio of lipids, pH and bilayer plane box lengths for each set of simulation parameters.

System Number	**Lipids**	**Molar Ratio**	**pH**	**Bilayer Plane Box Lengths ($Å^2$)**
1	DPPC:DSPS	0.5:0.5	neutral	61 x 61
2	DPPC:DSPS	0.5:0.5	low	61 x 61
3	21PC:DSPS	0.5:0.5	neutral	61 x 61
4	21PC:DSPS	0.5:0.5	low	61 x 61
5	DPPC:DSPS	0.25:0.75	low	61 x 61
6	DPPC:DSPS	0.75:0.25	low	61 x 61
7	21PC:DSPS	0.25:0.75	low	61 x 61
8	21PC:DSPS	0.75:0.25	low	61 x 61
9	DPPC:DSPS	0.5:0.5	low	120 x 120
10	21PC:DSPS	0.5:0.5	low	120 x 120

During each simulation the total number of DSPS lipids with a DSPS nearest neighbor was calculated to provide a measure of the rate and extent of phase separation between the PC and PS lipids. For this calculation, a DSPS molecule was deemed to be nearest neighbors with another DSPS molecule if the nearest neighbor to the DSPS head group (the center of mass of coarse-grained site VIIa at neutral pH or VIIb at low pH) was any coarse-grained site on another DSPS molecule. As DSPS lipids separate into domains, the total number of DSPS lipids with a DSPS nearest neighbor should increase.

Equimolar Lipid Mixtures at Low and Neutral pH

Simulations on Systems 1, 2, 3, and 4 were run to investigate the behavior of bilayers composed of equimolar ratios of DPPC:DSPS and 21PC:DSPS, respectively, at both a neutral and a low pH. **Figure 3.3** provides snapshots (aerial images) of the bilayer configurations at 1 million and 1 billion collisions for simulations on Systems 1 through 4. **Figure 3.3 (a, c, e, g)** show that after only 1 million collisions a small amount of separation between PC and PS lipids can be detected. **Figure 3.3 (b, d, f, h)** show that after 1 billion collisions the PC and PS lipids in simulations from Systems 1, 2, 3 and 4 have separated into domains. Based on these images, it is difficult to detect if there is a difference between the phase separation achieved by the PC and PS lipids at neutral and low pH. In fact the extent of phase separation between the PC and PS lipids in **Figures 3.3b**, **3.3d**, **3.3f**, **3.3h** appears to be the same. In addition, the DPPC/DSPS and 21PC/DSPS systems appear to separate to the same extent.

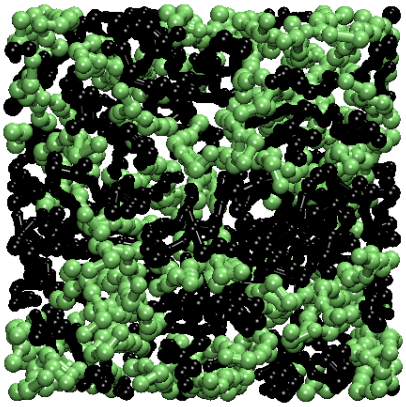
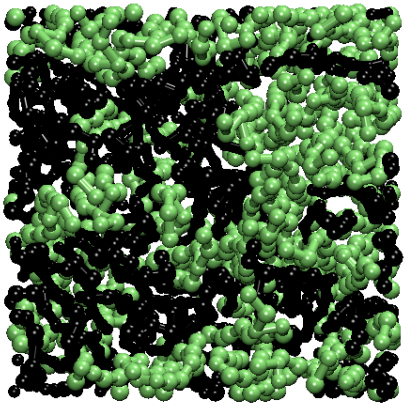
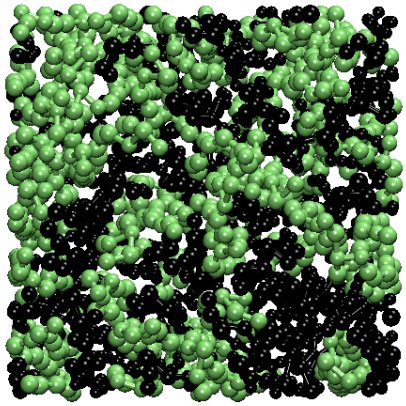
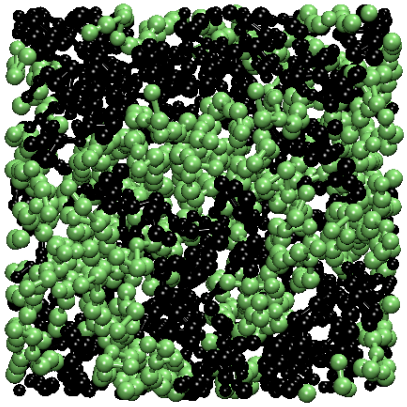
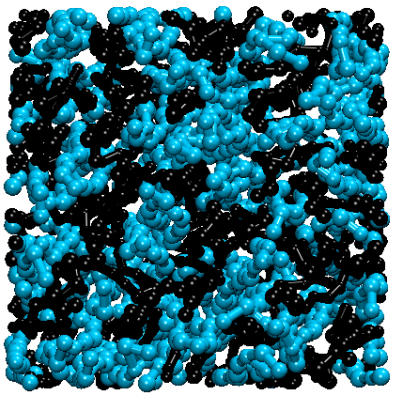
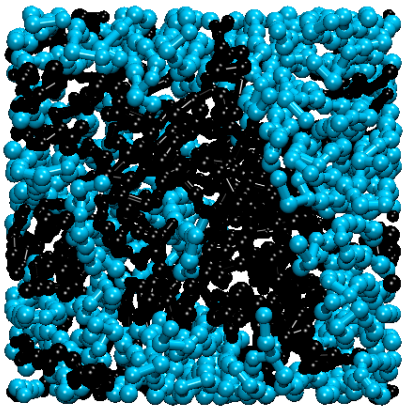
Figure 3.4 displays the average number of DSPS lipids whose closest neighbor is a DSPS lipid for Systems 1 through 4 as a function of time measured in terms of the number of

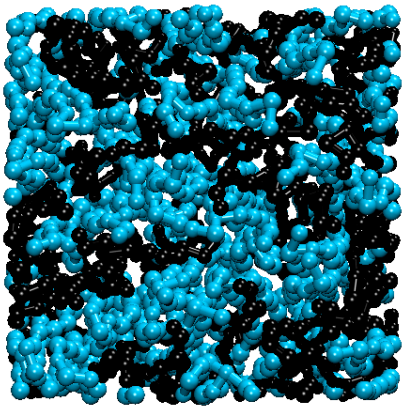
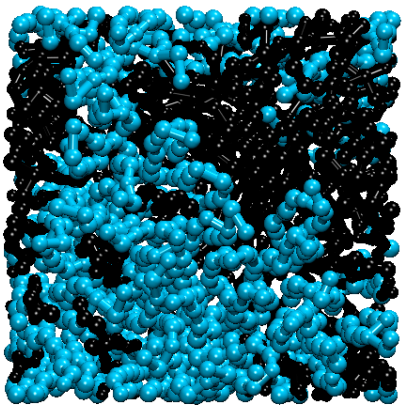
collisions. The closest neighbor to the DSPS head group (coarse-grained site VIIa at neutral pH or VIIb at low pH) was considered for this calculation. The data displayed is averaged over the 3 replicates run for each system. The figure shows that the average number of DSPS lipids whose closest neighbor is a DSPS lipid increases at approximately the same rate for Systems 3 (21PC:DSPS, neutral pH) and 4 (21PC:DSPS, low pH) until approximately 200 million collisions. After that the number of DSPS lipids with a DSPS nearest neighbor increases at a faster rate for System 4 (21PC:DSPS, low pH) than for System 3 (21PC:DSPS neutral pH). **Figure 3.4** also shows that the average number of DSPS lipids whose closest neighbor is a DSPS lipid is approximately the same for System 2(DPPC:DSPS, low pH) and System 1 (DPPC:DSPS, neutral pH) suggesting that these systems are undergoing phase separation at similar rates. This is consistent with the experimentally observed phase separation for DOPC/DSPS systems at both neutral and low pH. Finally, Systems 3 and 4 have higher numbers of DSPS lipids whose closest neighbor is a DSPS lipid than Systems 1 and 2. This suggests that systems run with the longer PC lipid (21PC) separate into heterogeneous domains faster than those run with the shorter PC lipid (DPPC).

An interesting event occurred during one of the replicate simulations of System 4. When reviewing snapshot images of this simulation we discovered that by 1 billion collisions, the majority of DSPS lipids had changed their alignment with the z-axis from vertical to a horizontal i.e., perpendicular to the z-axis. **Figure 3.5** provides snapshots of the side profile of the bilayer in this simulation at **(a)** 750 million collisions and **(b)** 1 billion collisions. In **Figure 3.5a** the 21PC (cyan) and DSPS (black) lipids are oriented vertically in alignment with the z-axis. In **Figure 3.5b** the 21PC lipids remain in the vertical orientation,

however, the majority of the DSPS lipids have adopted a horizontal alignment, perpendicular to the z-axis. We believe that the change in the alignment of the DSPS lipids is reflected in a sharp increase in the number of DSPS lipids with a DSPS nearest neighbor for System 4 observed for this run at around 800 million collisions in **Figure 3.4**. Therefore, it makes sense that adopting the new horizontal alignment would result in an increase in the number of DSPS head groups that interact. In addition, it is also likely that this realignment caused the dramatic increase observed for the error bars in **Figure 3.4** after approximately 800 million collisions since this phenomena was only observed in one of the three replicates run for System 4. This realignment is not seen in any of the neutral pH simulations or in the low pH simulation of DPPC:DSPS (System 2). We plan to investigate why this realignment of the DSPS lipids occurs. It is possible that some of our interaction energies at low pH are too large or our square-well widths are too wide, resulting in this phenomenon. It is also possible that this realignment is an artifact of the small system size that we are using to study the phase separation in bilayers.

Figure 3.3: Snapshots (a, c, e and g) of the bilayer configurations are after 1 million collisions and snapshots (b, d, g, h) of the bilayer configurations are after 1 billion collisions for Systems 1-4. The color scheme is: black (DSPS lipids), lime (DPPC lipids), blue (21PC lipids)

System Numbers	Snapshot (Arial Image) at 1 Million Collisions	Snapshot (Arial Image) at 1 Billion Collisions
1 DPPC:DSPS 0.5:0.5 neutral pH	 **(a)**	 **(b)**
2 DPPC:DSPS 0.5:0.5 low pH	 **(c)**	 **(d)**
3 21PC:DSPS 0.5:0.5 neutral pH		

	(e)	(f)
4 21PC:DSPS 0.5:0.5 low pH	 **(g)**	 **(h)**

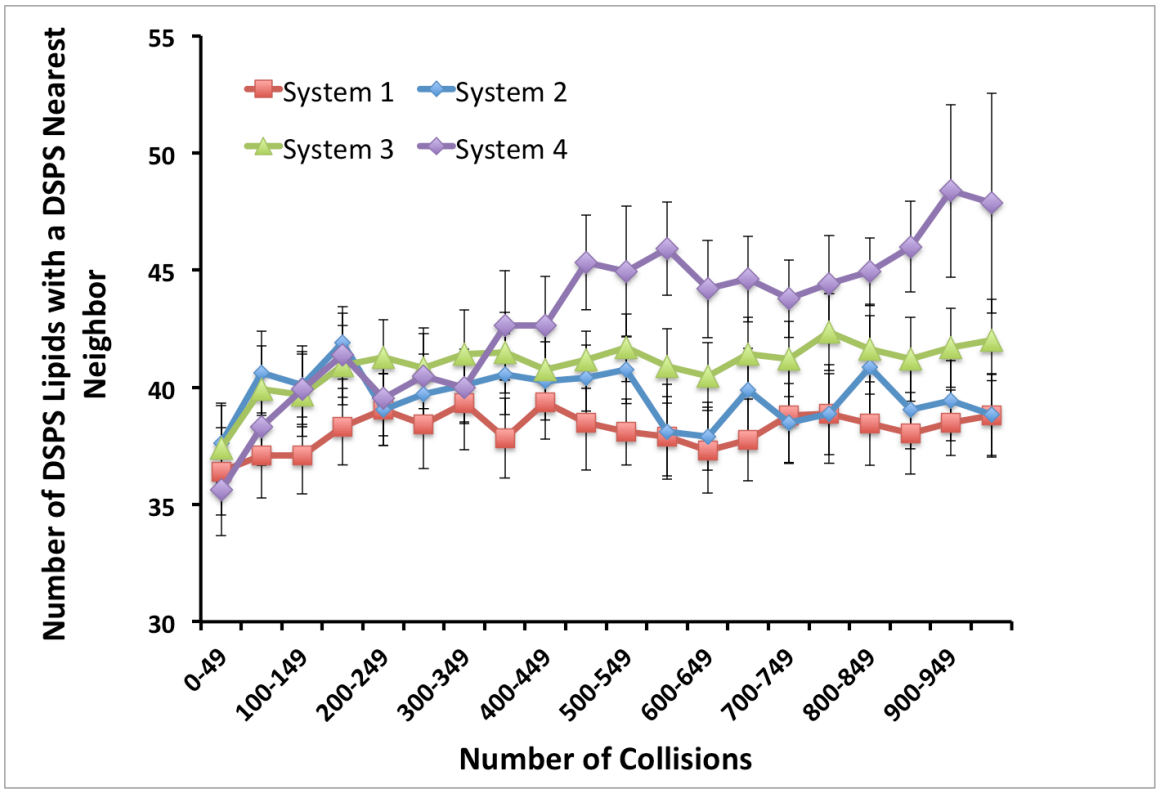


Figure 3.4: The number of DSPS lipids with a DSPS nearest neighbor versus the collision time for Systems 1 through 4. The data displayed is time averaged in 50 million collision increments from 0 to 1000 million collisions. In addition, the data was averaged for the 3 replicates run for each system.

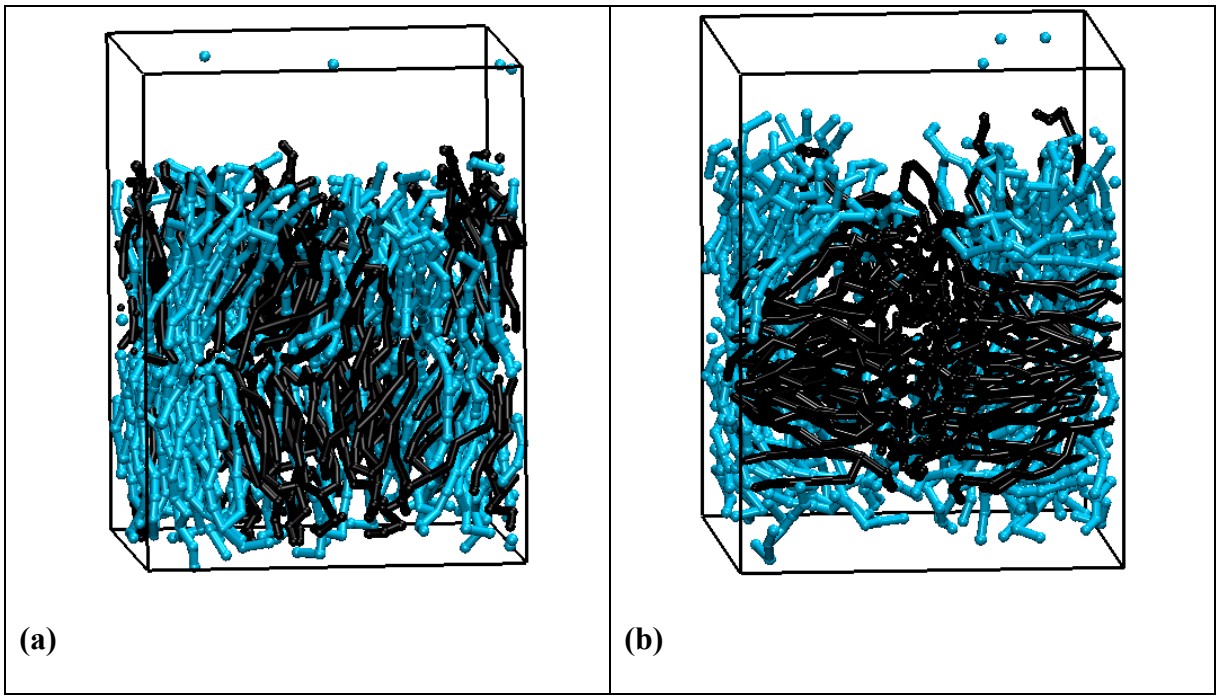


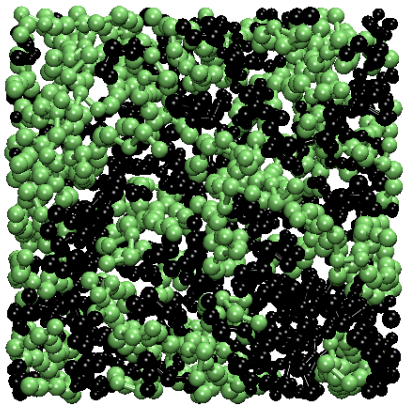
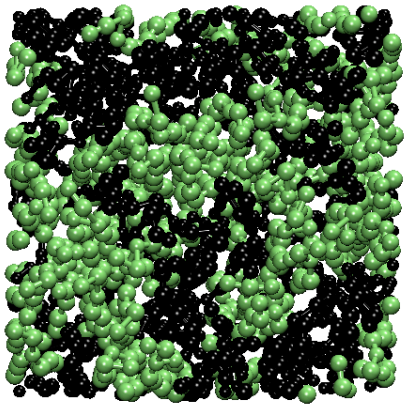
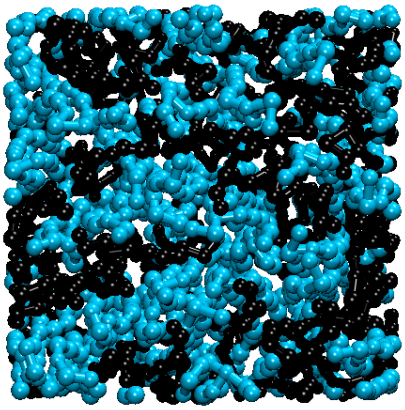
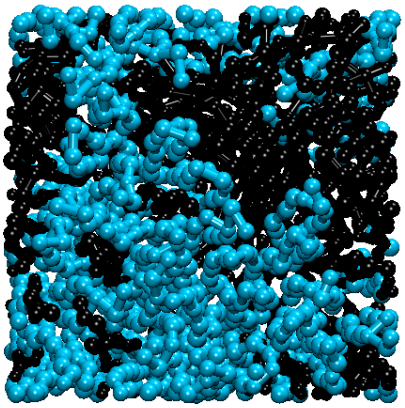
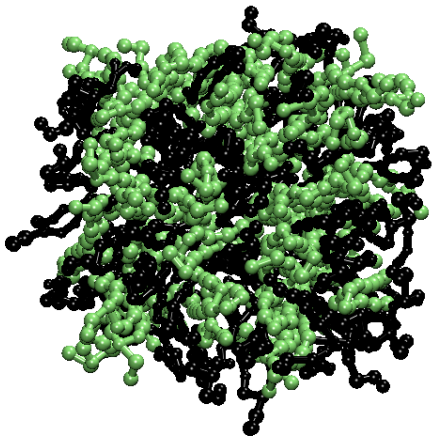
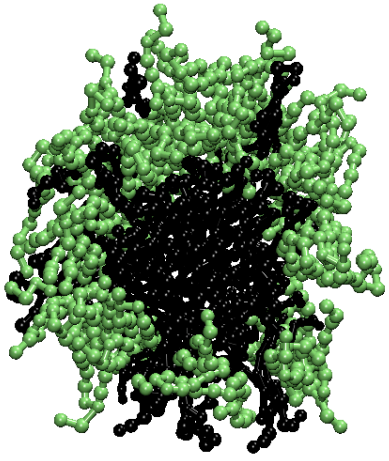
Figure 3.5: Snapshots of the side profile of replicate 3 from System 4 at **(a)** 750 million collisions and **(b)** 1 billion collisions. The color scheme is: DSPS lipids (black) and 21PC lipids (cyan).

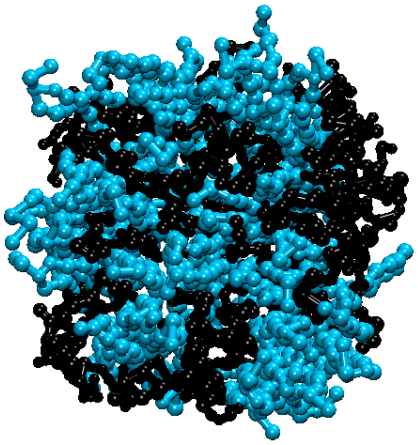
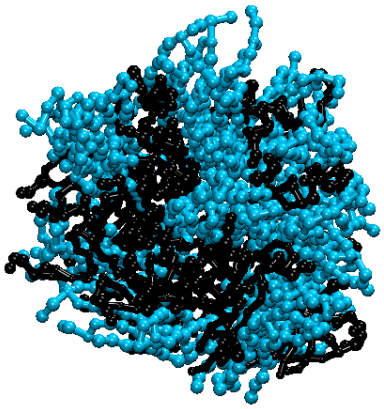
Equimolar Lipid Mixtures at Low pH with Different Surface Area Constraints

In order to learn how bilayer surface area constraints associated with periodic boundary conditions impacts the phase separation in DPPC:DSPS and 21PC:DSPS systems, we compare the results of simulations of Systems 2, 4, 9 and 10. The goal was to determine if allowing the bilayer to interact with its periodic image had an effect on the rate at which the PC and PS lipids separated into different domains. Simulations on Systems 2, 4, 9 and 10 contained bilayers with equimolar ratios of DPPC:DSPS and 21PC:DSPS lipids, respectively. Each of these simulations were started from initial bilayers with an area of 61Å x 61Å. The box lengths for simulations on Systems 2 and 4 were 61Å x 61Å while, the box lengths for

simulations on System 9 and 10 were 120Å x 120Å. Therefore, in simulations of Systems 2 and 4 the bilayers can interact with their periodic images, whereas in simulations of Systems 9 and 10 they cannot. **Figure 3.6** provides snapshots (aerial images) after 1 million and 1 billion collisions from one of the replicates of the simulations that were run on Systems 2, 4, 9 and 10. At 1 billion collisions the area per lipid for Systems 2, 4, 9 and 10 was 58.1Å^2, 58.1Å^2, 35.2Å^2, and 35.2Å^2, respectively. Therefore, when the bilayers are not promoted to interact with their periodic images, the surface area they adopt decreases. The bilayer in System 10 (no periodic boundary conditions) appears to separate into heterogeneous domains extremely fast as can be seen by comparing the images of System 10 at 1 million collisions (**Figure 3.6g**) and 1 billion collisions (**Figure 3.6h**). In comparison, the bilayers in Systems 2, 4 and 9 appear to separate more slowly than the bilayer in System 10. The images from 1 million collisions shown in **Figure 3.6a** (System 2), **3.6c** (System 4) and **3.6e** (System 9) have a greater number of domains (meaning less phase separation) than the images of these systems at 1 billion collisions shown in **Figure 3.6b** (System 2), **3.6d** (System 4) and **3.6f** (System 9). Overall, System 9 appears to have separated to the greatest extent forming only 1 DPPC and 1 DSPS domain.

Figure 3.6: Snapshots (aerial images) of the bilayer formed at 1 billion collisions in simulations of Systems 2 (a and b), 4 (c and d), 9 (e and f) and 10(g and h). The color scheme is: black (DSPS lipids), lime (DPPC lipids), blue (21PC lipids).

System Numbers	Snapshot (Arial Image) at 1 Million Collisions	Snapshot (Arial Image) at 1 Billion Collisions
2 DPPC:DSPS 0.5:0.5 low pH Box Lengths: 61Åx61Å	 **(a)**	 **(b)**
4 21PC:DSPS 0.5:0.5 low pH Box Lengths: 61Åx61Å	 **(c)**	 **(d)**
9 DPPC:DSPS 0.5:0.5 low pH Box Lengths: 120Åx120Å		

	(e)	**(f)**
10 21PC:DSPS 0.5:0.5 low pH Box Lengths: 120Åx120Å	 **(g)**	 **(h)**

To further investigate how surface area constraints associated with periodic boundary interactions impacted the phase separation in DPPC/DSPS and 21PC/DSPS systems, we prepared **Figure 3.7**, which displays the number of DSPS lipids with a DSPS nearest neighbor. According to the data in **Figure 3.7**, the number of DSPS lipids with a DSPS nearest neighbor increases most rapidly for System 9 (DPPC:DSPS, low pH, no surface area constraints). However, after about 300 million collisions the number of DSPS lipids with a DSPS neighbor remains stable for System 9. In addition, until approximately 900 million collisions, System 9 has the highest number of DSPS lipids with a DSPS nearest neighbor. At approximately 900 million collisions, System 4 (21PC:DSPS, low pH, surface area constraints) and System 9 (DPPC:DSPS, low pH, no surface area constraints) have approximately the same number of DSPS lipids with a DSPS nearest neighbor. However, the snapshots displayed in Figure 6 do not support this. Instead, the snapshots displayed in **Figure 3.6**, show that System 9 separated to a larger extent than Systems 2, 4 and 10. The previously-discussed unusual phenomenon that occurred in one of the replicates of System 4 in which the DSPS lipids changed their orientation with the z-axis most likely inflated the number of DSPS lipids with a DSPS nearest neighbor. In **Figure 3.7**, System 2 (DPPC:DSPS, low pH, surface area constraints) and System 10 (21PC:DSPS, low pH, no surface area constraints) have similar amounts of DSPS lipids with a DSPS nearest neighbor. This is in not line with the extent of phase separation for these systems that can be seen in the snapshots provided in **Figure 3.6** which show a higher level of separation achieved at 1 billion collisions for System 10 in comparison to System 2.

In summary, simulations of DPPC/DSPS bilayers with and without surface area constraints imposed by periodic boundary conditions showed that at low pH, DPPC/DSPS bilayers without surface area constraints separated to a larger extent than DPPC/DSPS bilayers with surface area constraints. According to the number of DSPS lipids with a DSPS nearest neighbor, 21PC:DSPS bilayers with surface area constraints separated to a greater extent than 21PC:DSPS bilayers without surface area constraints. However, we suspect that this might be due to replicate 3 from System 4(21PC:DSPS, low pH, surface area constraints), which may have elevated the number of DSPS lipids with a DSPS nearest neighbor for this system. According to images of the bilayers, approximately the same level of phase separation was achieved for 21PC:DSPS bilayers with and without surface area constraints.

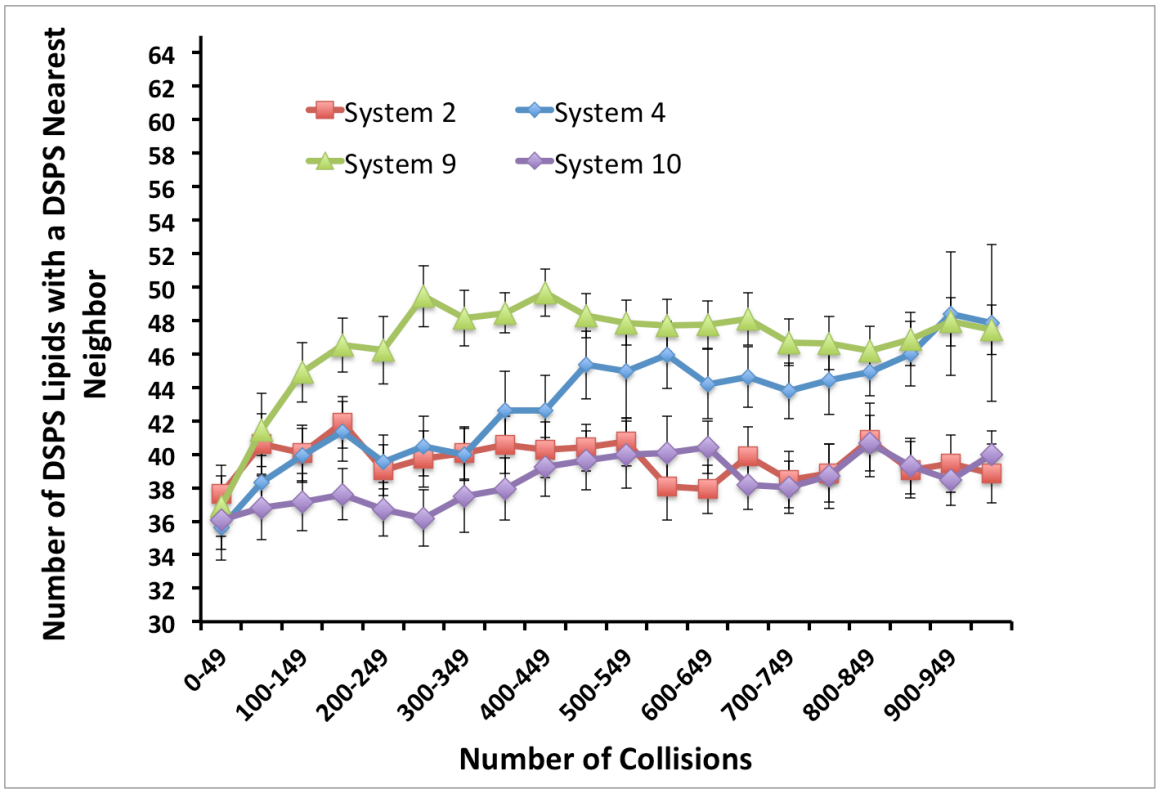


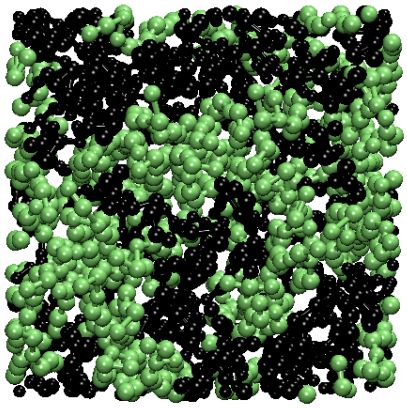
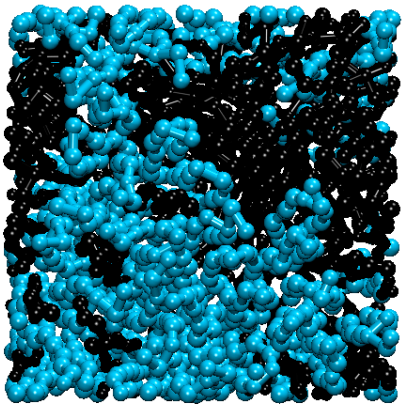
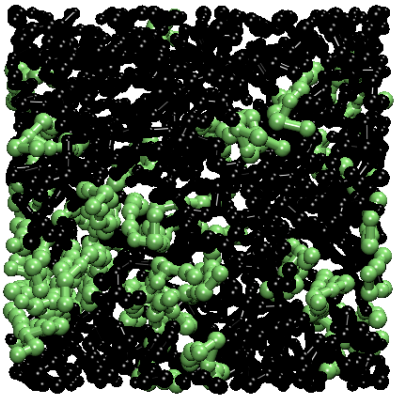
Figure 3.7: The number of DSPS lipids with a DSPS nearest neighbor verses the collision time for Systems 2, 4, 9 and 10. The values for the y-axis were averaged for each of the 3 replicates run for each system number.

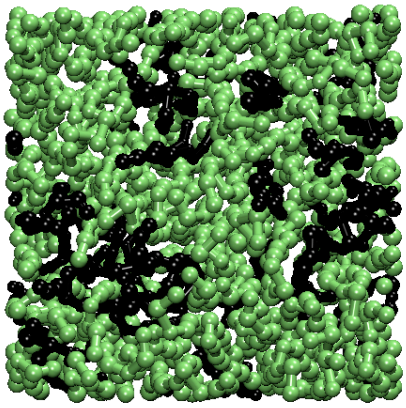
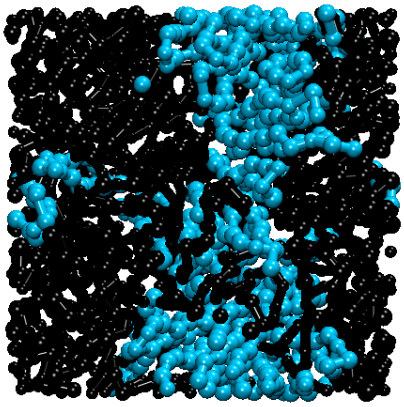
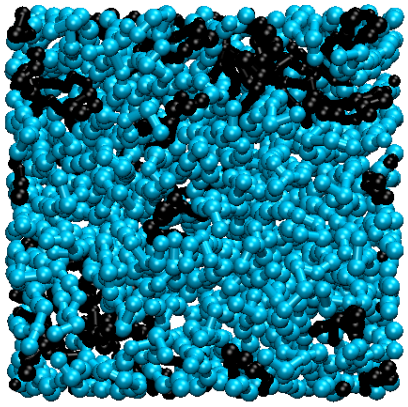
Comparing Different Ratios of Lipid Mixtures

Simulations were performed to investigate the rate at which bilayers composed of different ratios of PC and PS lipids separate at low pH. Simulations on Systems 2, 5 and 6 contained bilayers with DPPC/DSPS molar ratios of 0.25:0.75, 0.5:0.5 and 0.75:0.25, respectively. Simulations on Systems 4, 7 and 8 contained bilayers with 21PC/DSPS molar ratios of 0.25:0.75, 0.5:0.5 and 0.75:0.25, respectively. **Figure 3.8** provides snapshots of bilayers at 1 billion collisions from simulations on Systems 2, 4, 5, 6, 7 and 8. Each snapshot

shows that by 1 billion collisions the PC and PS lipids have separated into different domains. Snapshots from DPPC:DSPS simulations in **Figure 3.8a** (molar ratio 0.5:0.5), **3.8c** (molar ratio 0.25:0.75) and **3.8d (**molar ratio 0.75:0.25) all show evidence of phase separation between the DPPC and DSPS lipids. Similarly, the phase separation between 21PC and DSPS lipids can be seen in snapshots from 21PC:DSPS simulations in **Figure 3.8b** (molar ratio 0.5:0.5), **3.8e** (molar ratio 0.25:0.75) and **3.8f** (molar ratio 0.75:0.25). One important trend that is apparent in the snapshots is that as the ratio of PS to PC lipids increases, the number of domains formed and the size of those domains increases. Snapshots at 1 billion collisions of systems with a PC:PS mole fraction of 0.25:0.75 (Systems 5 and 7) have larger domains than snapshots of systems with a PC:PS mole fraction of 0.75:0.25 (Systems 6 and 8). This observation could be due to the fact that in simulations with a larger ratio of DSPS lipids, DSPS lipids have a greater chance of interacting with another DSPS lipid. Since DSPS lipids have a very strong attraction for each other it would help these lipids drive the domain separation.

Figure 3.8: Snapshots (aerial images) of bilayers at 1 billion collisions for Systems 2, 4, 5, 6, 7, and 8. The color scheme is: black (DSPS lipids), lime (DPPC lipids), blue (21PC lipids).

System Numbers	Snapshot (Arial Image) at 1 Billion Collisions
2 DPPC:DSPS 0.5:0.5 low pH Box Lengths: 61Åx61Å	 **(a)**
4 21PC:DSPS 0.5:0.5 low pH Box Lengths: 61Åx61Å	 **(b)**
5 DPPC:DSPS 0.25:0.75 low pH Box Lengths: 61Åx61Å	

	(c)
6 DPPC:DSPS 0.75:0.25 low pH Box Lengths: 61Åx61Å	 **(d)**
7 21PC:DSPS 0.25:0.75 low pH Box Lengths: 61Åx61Å	 **(e)**
8 21PC:DSPS 0.75:0.25 low pH Box Lengths: 61Åx61Å	 **(f)**

To quantify the rate at which bilayers composed of different ratios of PC and PS lipids separate into domains at low pH we calculated the number of DSPS lipids with a DSPS nearest neighbor versus collision time for System 2, 4, 5, 6, 7, and 8. For this calculation, the data from all 3 replicates of each system were averaged. These values are displayed in **Figure 3.9**. For the y-axis values we divided the number of DSPS lipids with DSPS nearest neighbors by the total number of possible DSPS neighbors in **Figure 3.9**. This is different than the y-axis values used in **Figures 3.4** and **3.7**. This was done because, depending on the system, each DSPS lipid could have one of 32, 64 or 96 DSPS nearest neighbors. According to **Figure 3.9**, as the ratio of PS lipids to PC lipids in a system increases, the number of DSPS lipids with a DSPS nearest neighbor also increases. The results also show that the number of DSPS lipids with a DSPS nearest neighbor was very similar for Systems 6 and 8 (PC:PS molar ratio of 0.75:0.25), Systems 2 and 4 (PC:PS molar ratio of 0.5:0.5) and Systems 5 and 7 (PC:PS molar ratio 0.25:0.75). However, for each molar ratio that was studied, systems with 21PC had slightly more DSPS molecules with a DSPS nearest neighbor than systems with DPPC. Therefore, the longer PC lipid (21PC) may promote phase separation to a lesser degree than the shorter PC lipid (DPPC).

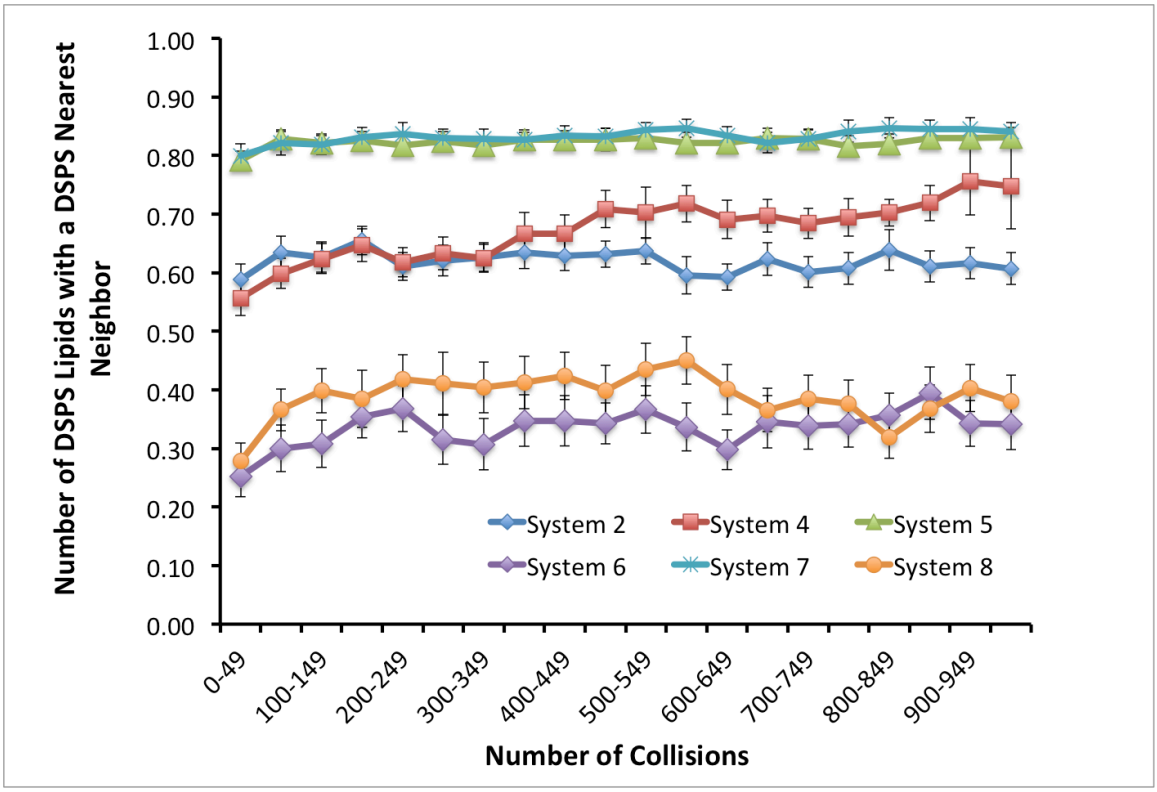


Figure 3.9: The number of DSPS lipids with a DSPS nearest neighbor verses the collision time for Systems 2, 4, 5, 6, 7 and 8. The values for the y-axis were averaged for each of the 3 replicates run for each system number.

Simulations of DPPC/DSPS and 21PC/DSPS Liposomes Containing Doxorubicin

Simulations of DPPC/DSPS and 21PC/DSPS liposomes containing doxorubicin were run to study whether or not the length of the PC lipid affected the rate of drug release from the liposomes. The DPPC/DSPS liposomes were composed of 1000 DPPC and 1000 DSPS lipids. The initial liposome configuration was generated by randomly inserting either DPPC or DSPS lipids into the inner or outer leaflets of the liposome. The initial outer radius of the DPPC/DSPS liposome was 100Å. The 21PC/DSPS liposomes were composed of 1000 21PC and 1000 DSPS lipids and had an initial outer radius of 105Å. The inner core of both

liposomes was loaded with 30 doxorubicin molecules. An atomistic representation of doxorubicin is used to model the drug geometry in these simulations. In this representation, 67 sites are used to represent each doxorubicin molecule. These sites are connected in such a way that the molecule is relatively inflexible. All interactions between lipids and doxorubicin molecules are modeled as hard sphere potentials. Thus we are accounting for the drug geometry but not the energetics; The intermolecular σ values between all doxorubicin and lipid coarse-grained sites were approximated as 1.5Å. All interactions between doxorubicin molecules were modeled as hard sphere potentials with σ values of 1.0Å. These approximation values of σ were used so that we could simulate the behavior of drug molecules in our lipid systems before running atomistic simulations and performing calculations to determine coarse-grained parameters for doxorubicin. In the future, we plan to run atomistic simulations to determine square-well parameters for lipid/doxorubicin and doxorubicin/doxorubicin interactions. **Figure 3.10** provides snapshots from the DPPC/DSPS/doxorubicin and the 21PC/DSPS/doxorubicin simulations at 1 million collisions and at 1.5 billion collisions. The images shown in this figure are cross-sections of the liposomes that are run in the simulations; this provides a better view of the drug molecules inside the liposomes. After 1.5 billion collisions, no drug molecules have escaped from the DPPC/DSPS or the 21PC/DSPS liposomes. However, two drug molecules have moved from the inner core of the 21PC/DSPS liposome to the bilayer of the liposome. In addition, there appears to be more domains of PC and PS lipids in the liposomes at 1.5 billion collisions than the liposomes at 1 million collisions. As the simulation progresses we hope to see some of the drug molecules escape.

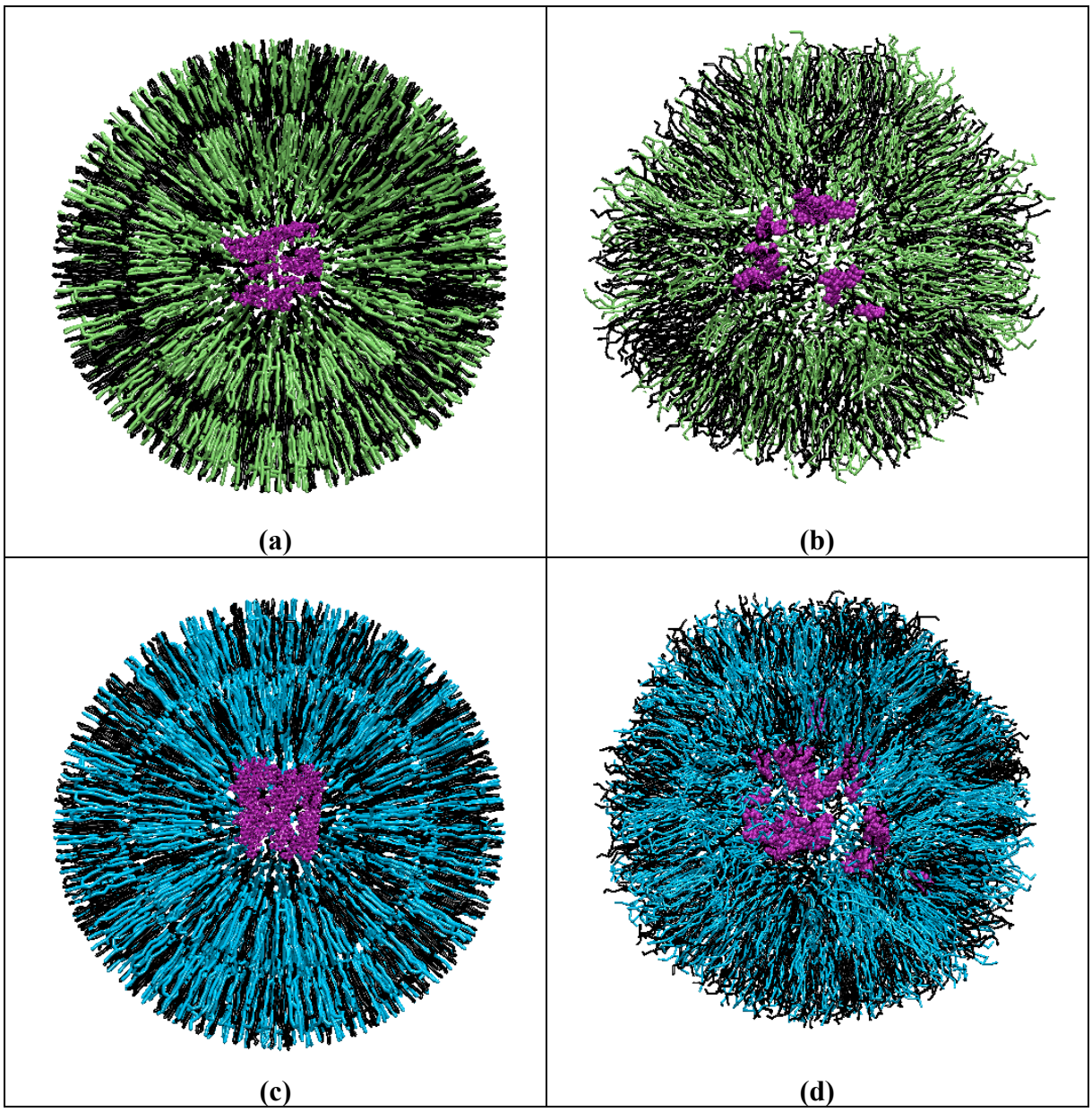


Figure 3.10: Snapshots of the DPPC/DSPS/doxorubicin liposome after (a) 1 million collisions and (b) 1.5 billion collisions. Snapshots of the 21PC/DSPS/doxorubicin liposome after (c) 1 million collisions and (d) 1.5 billion collisions. The color scheme is: black (DSPS lipids), lime (DPPC lipids), blue (21PC lipids), purple (doxorubicin)

3.4 Conclusion and Discussion

We described the extension of LIME, which was original designed for a single type of lipid, DPPC, at a temperature of 325K to be applicable to DPPC, DSPS and 21PC lipids at a temperature of 310K. LIME is an intermediate-resolution, implicit-solvent coarse-grained model for phospholipid molecules, which is designed for use with DMD. One of the main advantages of combining LIME with DMD is that it makes it possible to sample much wider regions of conformation space and longer time scales than in traditional molecular dynamics. This feature of LIME/DMD was extremely important for the simulation of the large liposomes described in this work. By treating solvent implicitly, coarse-graining and using a discontinuous potential we can simulate the behavior of mixed lipid systems at a much faster rate than in traditional molecular dynamics simulations.

The expanded version of LIME was used to simulate the behavior of bilayers composed of DPPC/DSPS and 21PC/DSPS at both neutral and low pH. According to our results DPPC/DSPS bilayers separated to the same extent at both a neutral and low pH. In addition, snapshots of 21PC/DSPS systems at neutral and low pH after 1 billion collisions show approximately the same extent of phase separation. However, for these 21PC/DSPS systems the number of DSPS lipids with a DSPS nearest neighbor was slightly higher at low pH than at neutral pH. This may be due to the event that took place in one of the replicates of System 4, which was a simulation of a 21PC/DSPS system at low pH. During this event the DSPS lipids changed their orientation after approximately 800 million collisions allowing more DSPS lipids to interact. This event may have inflated the number of DSPS lipids with a DSPS nearest neighbor that was calculated for System 4. If this event did not occur, we

might have calculated approximately the same number of DSPS lipids with a DSPS nearest neighbor for 21PC/DSPS systems at both neutral and low pH. While we do not have experimental data regarding the phase separation rate of bilayers composed of DPPC/DSPS or 21PC/DSPS experimental results by Sofou and coworkers have shown that liposomes composed similar lipids of (DOPC/DSPS) separate at approximately the same rate at both a neutral and low pH. In comparison, liposomes composed of DOPC/DSPS/cholesterol show significantly different separation rates when in neutral and low pH. Therefore, cholesterol plays a major role in inhibiting the rate of phase separation at neutral and low pH. In the future, we plan to calculate LIME parameters for cholesterol and to add cholesterol to our simulations. We believe that the PC/PS lipids would prefer to separate into heterogeneous domains at both neutral and low pH because the interaction strength between the DSPS head groups is much stronger at low pH at neutral pH but that cholesterol acts to prevent the separation. However, since there is less of a driving force (lower attraction energy between DSPS lipids) at a neutral pH than at low pH, it is likely easier for cholesterol to block the separation of PC and PS lipids at a neutral pH than at a low pH.

We also investigated the rate of phase separation in DPPC/DSPS and 21PC/DSPS bilayers with and without surface area constraints promoted (but not forced) by interactions with period boundary conditions. In experiments, liposomes composed of PC/PS lipids in a solution containing salt ions cannot simply adjust to any desired size for the following reason: as PC/PS lipids begin to separate into heterogeneous domains the lipids try to pack closer together. As they do this, the overall liposome size shrinks. During this process water molecules can escape from the inner core of the liposome, while salt molecules cannot. This

results in an osmotic pressure that prevents the liposome from shrinking to a large extent. According to our simulations, DPPC/DSPS bilayers without a restricted surface area will separate faster than bilayers with a restricted surface area. In comparison, bilayers composed of 21PC/DSPS separated at approximately the same rate at low pH in simulations with and without restrictions on their surface area.

We also investigated the rate at which bilayers composed of DPPC/DSPS or 21PC/DSPS at molar ratios of 0.25:0.75, 0.5:0.5 and 0.75:0.25 separated at low pH. Our results showed that the higher the concentration DSPS in the DPPC/DSPS or 21PC/DSPS bilayers, the greater the separation, meaning less domains of DSPS and DPPC with more lipids in each domain. For each molar ratio that was studied, systems with 21PC consistently had more DSPS molecules with a DSPS nearest neighbor than systems with DPPC. This may indicate that the longer PC lipid (21PC) may promote phase separation slightly more than the shorter PC lipid (DPPC).

Simulations of DPPC/DSPS and 21PC/DSPS liposomes containing doxorubicin were run to study whether or not the length of the PC lipid affected the rate of drug release from the liposomes. We did not see any drug molecules escape from the liposomes after 1.5 billion collisions. However, we do see PC and PS domain formation in both the DPPC/DSPS and the 21PC/DSPS liposomes. In addition, two of the doxorubicin molecules have migrated from the center core of the 21PC/DSPS liposome to the bilayer portion, which may indicate that they are in the process of escaping. We plan to continue to monitor the phase separation and drug encapsulation of these systems.

While our simulations were inspired by experiments done in the Sofou lab research, we have not, as yet, included the specific molecular embellishments associated with the Sofou liposomes, including the targeting ligand, fusion peptides and cholesterol.[15,31,32] The multiscale modeling procedure that we used to calculate LIME parameters for DPPC, 21PC and DSPS could, however, be used to calculate parameters for these additional molecules. In the future we plan to expand our model to include these molecules.

It is useful to discuss some of the limitations associated with DMD. One drawback of using DMD is that it is difficult to correlate collision times with real time. This is the result of the fact that DMD simulations are event-driven rather than time-driven and that solvent is modeled implicitly. If information is available regarding the experimental time required for a certain process or event to occur, this can be correlated to the simulation time required for the same process or event to occur, allowing us to estimate a correlation between simulation and real time. Another limitation of our model is that we can only correlate a reduced temperature of 0.8 with a real temperature of 310K. This is because we calculated our LIME parameters from GROMACS simulations performed at 310K. If we change the reduced temperature in our simulations we would not be able to calculate the associated real temperature. If desired, a multiscale modeling procedure could be used to determine the temperature dependence of the interaction energies in our model. This procedure would require data from atomistic simulations at various temperatures. We did not attempt to build a correlation between real temperature and reduced temperature because we only wanted to simulate the behavior of liposomes at one temperature (310K). Finally, although our LIME/DMD simulations are currently run using the NVT ensemble, we could perform NPT simulations in the future. To

do this we would need to perform hybrid Monte Carlo – DMD simulations in which the volume change moves are made with Monte Carlo and particle displacement moves are made with DMD. This type of simulation has been performed by previous members of the Hall group.[48,49] Despite these limitations we feel that our LIME/DMD model accurately models the behavior of liposomes observed experimentally and that it could be a useful tool to researchers trying to study the behavior of lipid systems on a molecular level.

3.5 References

1. Strijkers, G.; Kluza, E.; Tilborg, G.; Schaft, D.; Griffioen, A.; Mulder, W.; Nicolay, K. *Angiogenesis*, 2010, *13*, 161 – 173.
2. Nikolelis, D.; Hianik, T.; Krull, U. *Electroanalysis*, 1999, *11*, 7 – 15.
3. Bally, M.; Bailey, K.; Sugihara, K.; Grieshaber, D.; Voros, J.; Stadler, B. *Small*, 2010, *6*, 2481 – 2497.
4. Lingwood, D.; Simons, K. *Science*, 2010, *327,* 46 – 50.
5. Mukherjee, S.; Maxfield, F. *Traffic*, 2000, *1*, 203 – 211.
6. Liu, Y.; Li, K.; Pan, J.; Liu, B.; Feng, S. *Biomaterials*, 2010, *31*, 330 – 338.
7. Smet, M.; Langereis, S.; Bosch, S.; Grull, H. *Journal of Controlled Release*, 2010, *143*, 120 – 127.
8. Schwendener, R.; Ludewig, B.; Cerny, A.; Engler, O. *Liposomes*, 2010, *605*, 163 – 175.
9. Shi G.; Guo, W.; Stephenson, S.; Lee, R. *Journal of Controlled Release*, 2002, *80*, 309 – 319.
10. Templeton, N.; Lasic, D.; Frederick, P.; Strey, H.; Roberts, D.; Pavlakis, G. *Nature Biotechnology*, 1997*, 15*, 647 – 652.
11. Drulis-Kawa, Z.; Dorotkiewics-Jach, A. *International Journal of Parmaceutics*, 2010*, 387*, 187 – 198.

12. Zhang, L.; Pornpattananangkyl, D.; Hu, C.; Huang, C. *Current Medicinal Chemistry*, 2010, *17*, 585 – 594.

13. Needham, D.; Anyarambhatia, G.; Kong, G.; *Cancer Res*, 2000, *60*, 1197 – 1201.

14. Park, J.; Fong, P.; Lu, J.; Russell, K.; Booth, C.; Saltzmann, W.; Fahmy, T. *Nanomedicine: Nanotechnology, Biology and Medicine*, 2009, *5*, 410 – 418.

15. Karve, S.; Kempegowda, G.; Sofou, S. *Langmuir*, 2008, *24*, 5679-5688.

16. Ferrara, K.; Borden, M.; Zhang, H. *Acc Chem Res.*, 2009, *42*, 881 – 892.

17. Yatvin, M.; Weinstein, J.; Dennis, W.; Blumenthal, R. *Science*, 1978, *202*, 1290 – 1293.

18. Chem, Q.; Krol, A.; Wright, A.; Needham, D.; Dewhirst, M.; Yuan, F. *Int. J. Hyperthermia*, 2008, *24*, 475 – 482.

19. Smet, M.; Heijman, E.; Langereis, S.; Hijnen, N.; Grull, H. *Journal of Controlled Release*, 2011, *150*, 102 – 110.

20. Curtis, E.; Hall, C. *J. Phys. Chem. B.*, 2013, *117*, 5019 – 5030.

21. Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., Berendsen, H.: **GROMACS: Fast, Flexible, and Free**; *Journal of Computational Chemistry*, **26**, 1701 – 1719 (2005)

22. Jiang, Y.; Wang, H.; Kindt, J. *Biophysical Journal*, 2010**,** *98*, 2895 – 2903.

23. Hall, A.; Rog, T.; Karttunen, M.; Vattulainen, I. *J. Phys. Chem. B*, 2010, *114*, 7797 – 7807.

24. Pandit, S. Bostick, D. and Berkowitz, M: **Mixed Bilayer Containing Dipalmitoylphosphatidylcholine and Dipalmitoylphosphatidylserine: Lipid Complexation, Ion Binding, and Electrostatics**; *Biophysical Journal*, **85**, 3120-3131 (2003)

25. Illya, G., Lipowsky, R. and Shillcock, J.: **Two-component Membrane Material Properties and Domain Formation from Dissipative Particle Dynamics**; *The Journal of Chemical Physics*, **125**, 114710-1-9 (2006)

26. Faller, R. and Marrink, S.: **Simulation of Domain Formation in DLPC-DSPC Mixed Bilayers**; *Langmuir*, **20**, 7686-7693 (2004)

27. Marrink, S., de Vries, A., and Mark, A.: **Coarse Grained Model for Semiquantitative Lipid Simulations**; *J. Phys. Chem. B*, **108**, 750 – 760 (2004)

28. Risselada, H. and Marrink, S.: **The Molecular Face of Lipid Rafts in Model Membranes**; *PNAS*, **105**, 17367-17372 (2008)

29. Marrink, S., Risselada, H., Yefimov S., Tieleman, D., de Vries, A.: **The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations**; *J. Phys. Chem. B*, **111**, 7812-7824 (2007)

30. Lu, L. and Voth, G.: **Systematic Coarse-Graining of a Multi-Component Lipid Bilayer**; *J. Phys. Chem. B***, 113**, 1501-1510 (2009)

31. Karve, S.; Bandekar, A.; Ali, M.; Sofou, S. *Biomaterials*, 2010, *31*, 4409 – 4416.

32. Bandekar, A.; Zhu, C.; Gomez, A.; Menzenski, M.; Sempkowski, M.; Sofou, S. *Mol. Pharmaceutics*, 2013, *10*, 152 – 160.

33. Bandekar, A.; Sofou, S. *Langmuir*, 2012, *28*, 4113 – 4122.

34. Oostenbring, C., Villa, A., Mark, A., Van Gunsteren, W.: **A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6**, *Journal of Computational Chemistry*, **25**, 1656 – 1675 (2004)

35. Kukol, Andreas.: **Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins**, *J. Chem. Theory Comput.*, **5**, 615 – 626 (2009)

36. Kempegowda, G.; Karve, S.; Bandekar, A.; Adhikari, A.; Khaimchayev, T.; Sofou, S. *Langmuir*, 2009, *25*, 8144 – 8151.

37. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics*, 1996, *14*, 33 – 38.

38. Smith, S. W.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.,* 1997, *134*, 16 – 30.

39. Alder, B., and Wainwright, T. 1959. Studies in Molecular Dyanmics I. General Method. Journal of Chemical Physics, **31**, 459-466

40. Rapaport, DC. *J. Phys. A.*: *Math. Gen.*, 1978, *11*, L213 – L217.

41. Rapaport, DC. *J. Chem. Phys.*, 1979, *71*, 3299 – 3303.

42. Nguyen, H.D.; Hall, C.K. *Proc. Natl. Acad. Sci. U.S.A.*, 2004*, 101*, 16180 – 16185.

43. Andersen, HC. *J. Chem. Phys.*, 1980, *72*, 2384-2393.

44. Chiu, S., Clark, M., Balaji, V., Subramaniam, S., Scott, H., Jakobsson, E.: **Incorporation of Surface Tension into Molecular Dynamics Simulation of an Interface: A Fluid Phase Lipid Bilayer Membrane**; *Biophysical Journal*, **69**, 1230 – 1245 (1995)

45. Pandit, S. and Berkowitz, M.: **Molecular Dynamics Simulation of Dipalmitoylphosphatidylserine Bilayer with Na^+ Counterions**; *Biophysical Journal*, **92**, 1818 – 1827 (2002)

46. Gaussian 09, Revision **C.02**, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

47. Berendsen, H.; Postma, J.; Van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.*, 1984, *81*, 3684 – 3690.

48. Kenkare, N.; Hall, C.; Khan, S. *J. Chem. Phys.,* 2000, *113*, 404 – 418.

49. Shultz, A.; Hall, C.; Genzer, J. *J. Chem. Phys.*, 2004, *120*, 2049 – 2055.

CHAPTER 4

Modeling the Interaction Between Hydrophilic and Hydrophobic Nanoparticles with Bilayer Membranes Using LIME, an Intermediate-Implicit Solvent Model Designed for Use with Discontinuous Molecular Dynamics

4.1 Introduction

The subject of this paper is the interaction between nanoparticles and biomembranes and the attendant wrapping or penetration that follows from this interaction.. Motivation for this study comes from the increasing prevalence of nanoparticles in our everyday lives, the use of nanoparticles to deliver drugs, proteins, and antimicrobials into cells and concerns about nanoparticle toxicity. For a nanoparticle to be wrapped by, or penetrate, a cell membrane , specific (ligand-receptor) and nonspecific (surface charge, hydrophobicity, size and shape) binding interactions must overcome the resistive forces associated with membrane stretching and elasticity.[1] Experimental studies have been performed to investigate the role that both specific and nonspecific interactions play in the cellular uptake of nanoparticles [1,2,3,4,5], however, the complexity and diversity of nanoparticle types that currently exist make it very difficult to completely explore the behavior of all nanoparticle/membrane systems. Computer simulation is a tool that could be used to aid experimentalists by helping them to visualize the molecular motions that contribute to both the wrapping and direct penetration process that occurs at nanoparticle/membrane interfaces.

Our goal is to develop a computational model that provides molecular-level insights into, and facilitates the exploration of, the interaction between biomembranes and

nanoparticles with different geometric and energetic properties. In this paper we demonstrate how the combination of discontinuous molecular dynamics simulations (DMD) and our previously-developed LIME forcefield [6] can be used to model the interaction between lipid membranes and nanoparticles of different sizes, densities and hydrophobicities. We show that LIME/DMD simulations can be used to study the wrapping of hydrophilic nanoparticles of size range 5-100 Å by a lipid membrane and the mechanism by which a hydrophobic nanoparticle penetrates the inner core of a bilayer.

A number of experimental studies have been conducted to examine the interaction between nanoparticles and bilayer membranes. Chithrani and co-workers investigated the intracellular uptake of gold nanoparticles with different sizes and shapes.[3] Rod shaped nanoparticles with dimensions of 40x14nm and 74x14nm and spherical nanoparticles with diameters of 14, 30, 50, 74 and 100 nm were studied. The cells were incubated with the gold nanoparticles for 6 hours. Subsequently, the concentration of Au that had accumulated in the cells was measured. The cellular uptake of spherical nanoparticles exhibited a maximum as a function of nanoparticle size; it was greatest for particles with diameters of 30nm and 50nm and lowest for particles with diameters of 14nm, 74nm and 100nm. Cellular uptake for rod shaped particles was lower than that for spherical particles.[3] The authors speculated that the difference in the uptake between the various sizes and shapes of nanoparticles could be due to surface curvature and the amount and type of proteins absorbed onto the nanoparticle surface.

Win and co-workers investigated the effect of particle size and surface coating on the cellular uptake of polymeric nanoparticles intended for the oral delivery of anticancer

drugs.[4]. The authors evaluated the cellular uptake of 50nm, 100nm, 200nm 500nm and 1000nm polystyrene nanoparticles by Caco-cells. The 100nm and 200nm nanoparticles had the best cellular uptake, whereas the 50nm nanoparticles had the smallest amount of cellular uptake. Experiments were also performed to study the cellular uptake of polystyrene (PS) nanoparticles and poly(lactic-co-glycolic acid) (PLGA) nanoparticles coated with polyvinyl alcohol (PVA) or vitamin E TPGS. The Vitamin E TPGS-coated PLGA nanoparticles had better cellular uptake than the PS nanoparticles and the PVA-coated nanoparticles.[4] In another study Verma and co-workers compared the cell-membrane penetration achieved by two different nanoparticles with the same size, shape and ratio of hydrophobic to hydrophilic molecules.[5] The nanoparticles differed only in the arrangement of the surface hydrophilic and hydrophobic groups. One type of nanoparticle was coated with striations of alternating anionic and hydrophobic groups and the other type was coated with random distribution of anionic and hydrophobic groups. They found that the striated nanoparticles were able to pass directly through cell membranes and did not undergo endocytosis or pinocytosis to reach the cytosol. In contrast the nanoparticles with the random distribution of anionic and hydrophobic groups were almost completely blocked from cell entry.[5] An alternative, albeit indirect way to quantify cellular uptake of different types of nanoparticles is to measure the cytotoxicity that can accompany this process. For example, Pan and co-workers studied the cytotoxicity of gold nanoparticles with diameters ranging in size from 0.8nm to 15nm in four cell lines.[2] They also tested the toxicity of very small (diameter <0.8nm) gold particles (gold thiomalate). They found that nanoparticles with diameters in the size range from 1 – 2nm were more toxic to all four of the cell lines tested than the very small

gold nanoparticles (gold thiomalte) or the larger 15nm particles. The authors speculate that the nanoparticle toxicity was a result of endocytosis, however, their experimental methods did not allow them to determine an exact cause of cell death.[2]

In addition to the experimental work that has been performed to examine nanoparticles and membranes, various approaches to modeling the interaction between nanoparticles and membranes with simulations have been described in the literature. The levels of detail used to represent the molecules in these models fall roughly in two main categories: high-resolution and low-resolution. High-resolution or atomistic models represent the geometry and energetics of all molecules realistically and typically account for the motion of every atom including every solvent atom. One example of a high-resolution model is that developed by Bedrov and co-workers to investigate the interaction and passive transport of C_{60} fullerenes into lipid membranes composed of di-myristoyl-phosphatidylcholine (DMPC).[7] The system studied in these atomistic simulations consists of a DMPC bilayer composed of 52 lipid molecules and 1800 water molecules. The atomistic simulations were performed using the Lucretius molecular dynamics simulation package [7] along with the lipid force field parameters from CHARMM27.[8] The free energy and the diffusivity of a fullerene was obtained as a function of its position within the membrane; these properties were used to calculate the membrane permeability.[7]

In contrast to high-resolution models, low-resolution models, which are also known as coarse-grained models, are based on a simplified representation of molecular geometry and energetics. In a coarse-grained model a single interaction site is used to represent the behavior of a group of several atoms. This reduces the total number of sites whose

trajectories must be calculated, thereby increasing the speed of the simulation. One example of a low-resolution model used to describe nanoparticle membrane interactions is that developed by Vacha and coworkers to study the passive endocytosis of ligand-coated nanoparticles of different sizes, shapes, coverage and membrane-binding strength.[9] For this work the authors used the implicit-solvent model for phospholipid membranes developed by Cooke et al.[10] In this model, three spheres are used to represent each phospholipid molecule: a hydrophilic sphere to represent the phospholipid headgroup and two hydrophobic spheres to represent the two phospholipid tails. The nanoparticles are composed of several spheres that are the same size as the hydrophilic headgroup sphere, most of which are hydrophilic. All simulations were performed using the ESPRESSO molecular dynamics package.[11] Vacha and coworkers demonstrated that larger spherical particles experienced endocytosis more easily than smaller particles. The authors explain that this observation is a result of the more favorable compromise between bending rigidity and surface adhesive energy for the larger nanoparticles than for the smaller particles. The results also show that the simulation rate for passive endocytosis is higher for spherocylindrical particles than it is for spheres and that endocytosis is suppressed for particles with sharp edges.[9]

Another example of the use of a coarse-grained model to determine the effect of a nanoparticle's size on its translocation across a lipid bilayer is work by Lin and co-workers.[12] The nanoparticles in these explicit–solvent simulations are hydrophobic and range in size from 1.284 nm to 2.912 nm.; the lipid chosen for study is DPPC. All simulations were run using GROMACS 3.3.3 [13] with the MARTINI force field developed

by Marrink et al.[14,15] Results showed that the time required for a nanoparticle to translocate to different positions in a DPPC bilayer decreased with the size of the nanoparticle.[12] Yang and Ma also used coarse-grained computer simulations based on dissipative particle dynamics (DPD) to simulate the translocation of nanoparticles with different shapes across a lipid bilayer.[16] The simulations predicted the translocation of nanoparticles through the lipid membrane but not the endocytosis of the nanoparticles by the membrane. The lipid molecules contained two hydrophilic head spheres and five hydrophobic tail. The nanoparticles were constructed by arranging hydrophilic DPD spheres in the desired geometrical shape. The nanoparticles studied had the following geometries: (1) disks of radius 0.8nm and 1.6nm, (2) ellipsoids approximately 1.6nm wide and 6.4nm long, (3) cylinders with a radius of 0.8nm or 2.0nm and (4) pushpin-shaped particles measuring 2.0nm at their widest point and 0.8 nm in height. The authors concluded that the nanoparticle shape and initial orientation significantly affect the interaction between the nanoparticle and the lipid bilayer.

In this paper, we use an implicit-solvent intermediate-resolution model for lipid molecules, which we call "LIME," with discontinuous molecular dynamics (DMD), a fast alternative to traditional molecular dynamics simulation, to model the interaction between both hydrophilic and hydrophobic nanoparticles and DPPC bilayer membranes. The LIME geometric and energetic parameters for the DPPC lipids were obtained using a multiscale modeling approach as described in our previous paper. In multiscale modeling atoms are grouped into coarse-grained sites and the geometric and energetic parameters for these coarse-grained sites are extracted from atomistic simulations in explicit solvent. The

nanoparticle is modeled as a single sphere, essentially a generic nanoparticle, rather than as a cluster of spheres as other investigators have done. This is in keeping with our vision of this work as "proof of method" simulations, which could eventually evolve into examinations of more specific nanoparticle-membrane systems. Two types of nanoparticles are examined, hydrophilic and hydrophobic. The hydrophilic nanoparticles have square-well interactions with hydrophilic lipid sites and the hydrophobic nanoparticles have square-well interactions with hydrophobic lipid sites in our model. We investigate the extent to which hydrophilic nanoparticles with diameters from 5-100 Å are wrapped by a DPPC membrane and the extent to which hydrophobic nanoparticles with diameters from 5-20 Å penetrate the membrane. We chose not to study nanoparticles larger than 100 Å because this would require extremely large bilayers to provide enough lipid molecules to fully wrap the nanoparticles, which would in turn require significant computational resources and time. We also examine how the nanoparticle mass per volume affects the wrapping process membranes. Although not physically realistic, in this work we have decoupled particle volume from interaction strength. We selected a reasonable square-well width on the order of 5 Å, and interaction energies of -2.0eV for interactions between hydrophilic nanoparticles and hydrophilic lipid molecules and for interactions between hydrophobic nanoparticles and hydrophobic lipid molecules. In the future, we will utilize our multiscale modeling approach to obtain realistic square-well depths and widths for nanoparticles in the size range we found to be important from this work.

Highlights of our results include the following. Our model demonstrates the major role that nanoparticle size plays in the membrane wrapping process. We find that

hydrophilic nanoparticles with a diameter less than 20Å are not wrapped by bilayers; instead they become embedded in the bilayer's surface where they can interact with the hydrophilic head groups of the lipid molecules. Hydrophilic nanoparticles with diameters between 20Å and 100Å do became wrapped by the bilayer membrane. Hydrophobic nanoparticles with diameters of 5Å and 20Å do not undergo the wrapping process; instead they directly penetrate the membrane and remain within the inner hydrophobic core of the bilayers. The mass per volume of the nanoparticles had little effect on their interaction with the bilayer membranes. This was the case for both hydrophilic and hydrophobic nanoparticles with a mass per volume density within 0.12 – 0.67 amu/Å^3.

4.2 Methods and Model

To simulate the DPPC molecules in this work, LIME, an intermediate resolution implicit-solvent model for lipid molecules [6] developed for use with discontinuous molecular dynamics was employed. In LIME each DPPC molecule is represented by 14 coarse-grained sites one of six unique coarse-grained types (I-VI). A detailed description of the coarse-grained parameters used to describe each DPPC molecule is provided in our previous work.[6] **Figure 4.1** illustrates the coarse-grained representation of: (a) a DPPC molecule and (b) a nanoparticle. Types I and II represent the choline entity and the phosphate group, respectively. Ester coarse-grained sites 3 and 9 are assigned types II and IV, respectively. The coarse-grained sites in the hydrocarbon tails (excluding the terminal sites) are assigned type V. Finally, the terminal tail coarse-grained sites are classified as type

VI. Each nanoparticle in our simulations is represented by one single coarse-grained site and is assigned the coarse-grained type VII.

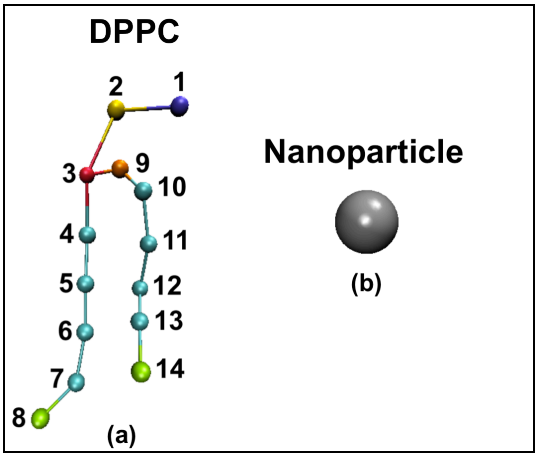


Figure 4.1: (a) Coarse-grained representation of DPPC (b) Coarse-grained representation of a nanoparticle. The color scheme is; purple (choline entity – type I for DPPC site 1); yellow (phosphate group – type II for DPPC site 2); red (ester group – type III for DPPC site 3); orange (ester group – type IV for DPPC site 9); cyan (alkyl tail groups – type V for DPPC sites 4-7&10-13); green (terminal tail groups – type VI for DPPC sites 8&14); gray (nanoparticle – type VII for nanoparticle site 1). The size of the DPPC coarse-grained sites and the nanoparticle are not drawn to scale.

The discontinuous molecular dynamics (DMD) algorithm, a very fast alternative to traditional molecular dynamics simulation, is the simulation method used for this work.[17,18] In DMD simulations, particles interact via a combination of hard-sphere and square well-potentials which means that the forces on particles need only be calculated when discontinuities in the potential are encountered. This allows for faster simulations than traditional molecular dynamics, enabling examination of larger systems and longer time scales. A hard sphere is an impenetrable, solid sphere; a square-well is a hard sphere

surrounded by an attractive well. Expressions for the hard sphere (HS) and square well (SW) potentials between spheres *i* and *j* are, respectively:

$$u_{ij}^{HS}(r) = \begin{cases} \infty & r \leq \sigma_{ij} \\ 0 & r > \sigma_{ij} \end{cases} \qquad \text{Equation 1}$$

$$u_{ij}^{SW}(r) = \begin{cases} \infty & r \leq \sigma_{ij} \\ -\varepsilon_{ij} & \sigma_{ij} < r \leq \sigma\lambda_{ij} \\ 0 & r > \lambda_{ij} \end{cases} \qquad \text{Equation 2}$$

where r is the distance between spheres, σ_{ij} is the hard sphere diameters, $\sigma\lambda_{ij}$ is the well diameter and ε_{ij} is the well depth. In our DMD simulations , the initial velocities assigned to coarse-grained sites are based on a Maxwell-Boltzmann distribution about the desired simulation temperature. The particle trajectories are then followed by calculating the time between each collision and advancing the simulation to the next event. Types of events include a collision between two hard spheres, a bond event when the distance between two bonded spheres reaches a minimum or maximum limit, and square well events when two spheres enter (capture), unsuccessfully attempt to escape (bounce) or successfully leave (dissociation) a square well.[19,20,21,22]

In all LIME/DMD simulations the simulation temperature is expressed in terms of the reduced temperature T*:

$$T^* = k_B T/\varepsilon^* \qquad \text{Equation 3}$$

where k_B is Boltzmann's constant, T is the temperature, and ε* is the reference interaction strength.[23] The same reference interaction strength, ε* = 0.0363, that was used previously for our simulations of DPPC lipids at 325K, was used for the simulations described in this

work. A detailed description of the procedure used to calculate this value is provided in our previous work.[6] Therefore, T* = 0.77 in all our of nanoparticle/bilayer simulations.

The LIME σ_{ij}, $\sigma\lambda_{ij}$, and ε_{ij} coarse-grained parameters for DPPC molecules were obtained using a multiscale modeling technique. In this procedure coarse-grained parameters are extracted from data collected from atomistic simulations. Data used to calculate the DPPC parameters were obtained by running united-atom explicit-solvent simulations at T=325K of 30 DPPC lipids using the GROMACS simulation package [24,25] version 4.5.4 along with the GROMOS96 53a6 forcefield.[26] Complete details of the multiscale modeling procedure used to calculate the LIME DPPC parameters is provided in our previous publication.[6] **Table 4.1** provides the values for σ_{ij}, $\sigma\lambda_{ij}$, and ε_{ij} between all DPPC coarse-grained types. In addition to the σ_{ij}, $\sigma\lambda_{ij}$, and ε_{ij} coarse-grained parameters, the GROMACS simulation data was also used to calculate the minimum and maximum bond and pseudobond lengths between coarse-grained sites. Pseudobonds are used in the model to maintain the relative stiffness of the lipid molecules by limiting the fluctuation of the coarse-grained sites to the angles and torsional angles observed during the GROMACS simulation. For coarse-grained pairs with an $\varepsilon_{ij} = 0.0$ (hard-sphere interaction), the $\sigma\lambda_{ij}$,is listed as "NA" for "not applicable" since these pairs do not interact with a square-well potential.

Table 4.1: The hard sphere diameters, square well widths and interaction energies for each pair of coarse-grained types.

Coarse-grained Type i	Coarse-grained Type j	$\sigma_{HS\,ij}$ (Å)	$\sigma\lambda_{ij}$ (Å)	ε_{ij} (eV)
I	I	4.35	12.65	-0.065
I	II	3.85	9.85	-0.070
I	III	3.85	9.95	-0.050
I	IV	3.75	10.25	-0.047
I	V	4.15	NA	0.000
I	VI	4.15	NA	0.000
II	II	4.05	12.15	-0.080
II	III	3.45	13.05	-0.048
II	IV	3.35	12.35	-0.030
II	V	3.75	NA	0.000
II	VI	3.65	NA	0.000
III	III	3.65	10.45	-0.037
III	IV	3.25	9.95	-0.036
III	V	3.65	10.53	-0.022
III	VI	3.55	7.70	-0.015
IV	IV	3.15	11.65	-0.035
IV	V	3.45	11.43	-0.026
IV	VI	3.45	10.85	-0.023
V	V	3.75	11.56	-0.050
V	VI	3.75	11.66	-0.054
VI	VI	3.65	11.02	-0.070

The σ_{ij} values between the nanoparticle and DPPC sites were chosen to allow us to study the behavior of nanoparticles with diameters ranging in size from 5 – 100 Å. We chose to study these relatively small nanoparticles in this "proof of method" work because larger nanoparticles would have required us to significantly increase the number of DPPC molecules in our simulations and hence reduce the number of collisions per CPU hour that could be achieved significantly. For this work we studied the interaction of the DPPC bilayer membrane with both hydrophilic and hydrophobic nanoparticles so that we could compare our results to data on nanoparticles with a wide range of hydrophobicities. We chose to model hydrophilic nanoparticles as spheres that only had strong interactions with the hydrophilic lipid head groups of the DPPC molecules. (Since the hydrophilic lipid head groups in LIME have stronger interaction energies with each other than with the hydrophobic tail groups,, we assumed that a hydrophilic nanoparticle would also have stronger interaction energies with the hydrophilic lipid head groups than with the hydrophobic tails.) To model these interactions the ε_{ij} values between DPPC coarse-grained sites 1 (choline entity), 2 (phosphate group) 3 (ester group) and 9 (ester group) and hydrophilic nanoparticles were each set to -2.0 eV. A value of -2.0 eV was chosen because it represents a very strong attraction (large well-depth) and we felt it would give us a good idea of the way that very hydrophilic nanoparticles would interact with a membrane. The ε_{ij} between all of the DPPC akyl tail groups (coarse-grained sites 4-8 and 10-14) and the hydrophilic nanoparticles were chosen to be zero, i.e. they interact as hard-spheres. The λ_{ij} values between hydrophilic nanoparticles and DPPC coarse-grained sites 1 (choline entity), 2 (phosphate group) 3 (ester group) and 9 (ester group) were set to a value that made the width of the square-well

interaction ($\sigma\lambda_{ij}$ - σ_{ij}) equal to 5.0 Å. Since this work was not performed to model any specific nanoparticle/bilayer system, we did not have any atomistic or experimental data to help us select a value for the width of the square-well interactions between the nanoparticle and lipid head groups. We decided to use a value of 5.0 Å for the width of these square-well interactions because it is large enough to allow the particles to interact, yet small enough to allow the lipid head-groups to maintain interactions between themselves. **Table 4.2** provides the values for ε_{ij} between all DPPC coarse-grained types and a hydrophilic nanoparticle.

Table 4.2: Interaction energies for between each pair of DPPC coarse-grained types and a hydrophilic nanoparticle.

Coarse-grained Type i	**Coarse-grained Type j**	**ε_{ij} (eV)**
I	VII	-2.0
II	VII	-2.0
III	VII	-2.0
IV	VII	-2.0
V	VII	0.0
VI	VII	0.0

In addition to studying hydrophilic nanoparticles we also ran simulations to investigate hydrophobic nanoparticle/membrane interactions. In LIME the hydrophilic lipid tails have much stronger interactions with each other than they do with the hydrophilic lipid

head groups. Therefore, we assumed that hydrophobic nanoparticles would also have stronger interactions, and prefer to interact, with, the hydrophobic lipid tails rather than with the hydrophilic lipid head groups. To model these interactions hydrophobic nanoparticles were assigned square-well interactions of strength ε_{ij} = -2,0eV with the hydrophobic tails (coarse-grained sites 4-8 and 10-14) and hard-sphere interactions with the DPPC coarse-grained sites 1 (choline entity), 2 (phosphate group) 3 (ester group) and 9 (ester group). The $\sigma\lambda_{ij}$ value between hydrophobic nanoparticles and DPPC coarse-grained sites 4-8 Å and 10-14 (hydrophobic alkyl tails) was set to a value that made the width of the square-well interaction ($\sigma\lambda_{ij}$ - σ_{ij}) equal to 5.0 Å. **Table 4.3** provides the values for ε_{ij} between all DPPC coarse-grained types and a hydrophobic nanoparticle.

Table 4.3: Interaction energies for between each pair of DPPC coarse-grained types and a hydrophobic nanoparticle.

Coarse-grained Type i	Coarse-grained Type j	ε_{ij} (eV)
I	VII	0.0
II	VII	0.0
III	VII	0.0
IV	VII	0.0
V	VII	-2.0
VI	VII	-2.0

4.3 Results and Discussion

Ten independent DMD simulations were run to study the interaction between hydrophilic nanoparticles with different physical properties and a bilayer composed of DPPC molecules. The nanoparticles in each simulation had a diameter ranging from 5 – 100 Å and a mass per volume density ranging from 0.013 – 0.67 amu/Å^3. Since nanoparticles are made of a wide variety of materials they have a wide range of densities. We decided not the study nanoparticle densities larger than 0.67 amu/Å^3 to avoid slowing down our simulations. Increasing the nanoparticle density decreases the velocity of the nanoparticles, which means it takes longer for them to travel to the surface of the bilayer membrane. Since our goal was to model the nanoparticle/bilayer interaction and not the behavior of the nanoparticle before it reaches the membrane surface we selected densities equal to or less than 0.67 amu/Å^3 The lowest mass per volume density limit that we studied , 0.013 amu/Å^{3,} was not based on the density of a specific nanoparticle. We considered this very low (and unrealistic) density because we felt it would allow us to determine whether or not density played a major role in the behavior of nanoparticles in our model. The mass per volume density of gold and silver nanoparticles is 116 and 6.3 amu/Å^3, respectively, which is significantly higher than the highest density we studied. In the future we plan to run simulations with more realistic densities. Each simulation was started from a preformed DPPC bilayer containing either 1500 or 4000 molecules. The bilayers composed of 1500 lipids were built to span an area of 218Å x 218Å in the center of a box with dimensions of 250Å x 250Å x 250Å. The bilayers composed of 4000 lipids were built to span an area of 356Å x 356Å in the center of a box with dimensions of 500Å x 500Å x 500Å. The bilayer was placed in a position where it could

not interact with its periodic boundary image in each simulation. This was done to prevent the surface tension of the bilayer from affecting its interaction with the nanoparticle. Each nanoparticle was placed approximately 5Å above the preformed bilayer at the center of the bilayer. At 5Å above the bilayer, the nanoparticle is far enough from the bilayer to prevent any overlaps, yet close enough to begin interacting with the bilayer. We chose to position nanoparticles close to the bilayers to avoid spending computational resources on simulations in which the nanoparticles did not interact with the bilayer. All simulations were run at a T* = 0.77. **Table 4.4** provides information regarding the nanoparticle diameter, nanoparticle mass, nanoparticle hydrophobicity and the number of DPPC lipids present in each simulation.

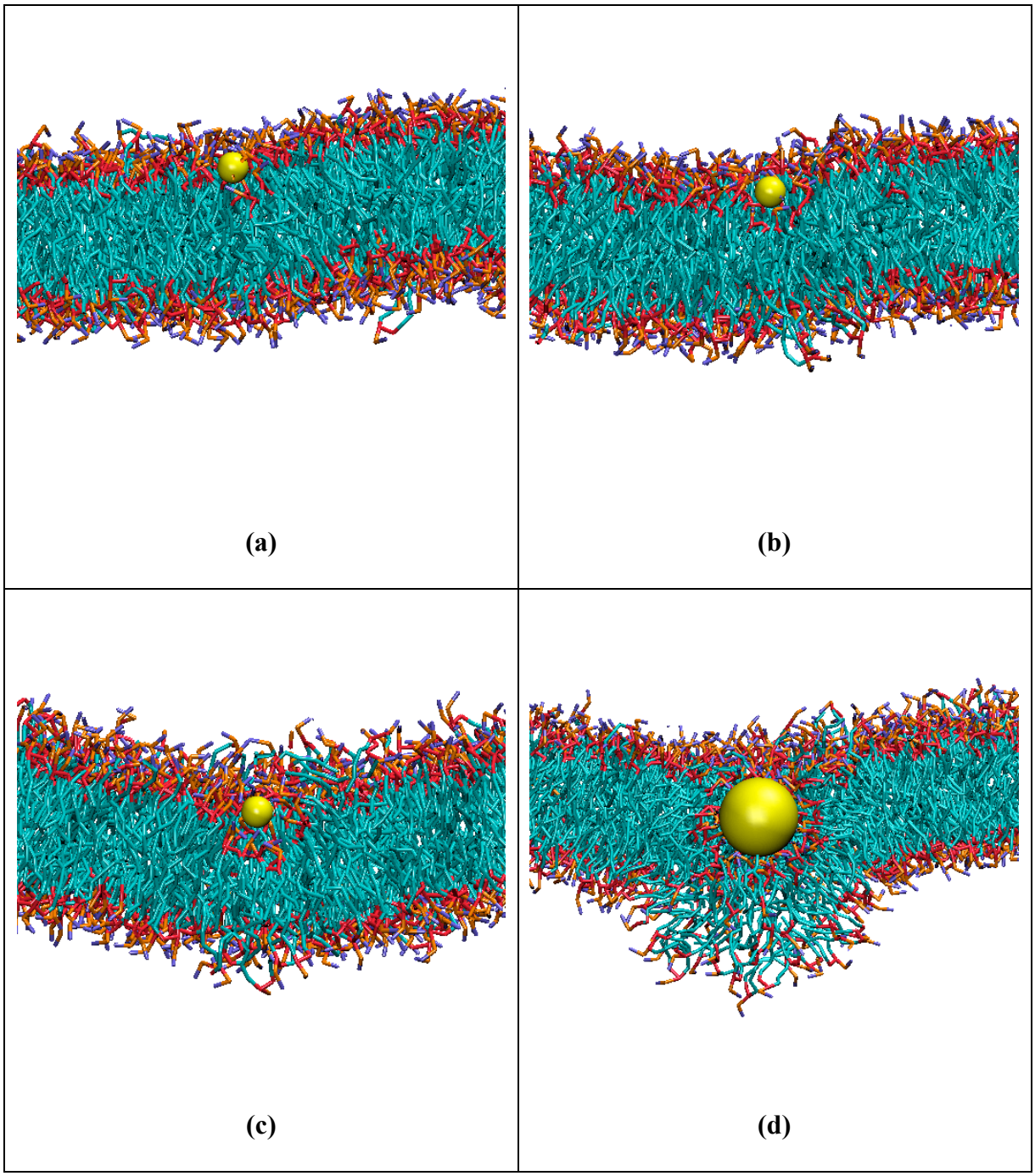
Table 4.4: The nanoparticle diameter, nanoparticle hydrophobicity, nanoparticle mass and number of DPPC lipids in each simulation.

Run #	Nanoparticle Diameter (Å)	Nanoparticle Mass (amu)	Hydrophobicity	Number of DPPC lipids
1	5	0.82	hydrophilic	1500
2	5	43.6	hydrophilic	1500
3	10	6.6	hydrophilic	1500
4	20	52.5	hydrophilic	1500
5	40	420.0	hydrophilic	1500
6	40	4200.0	hydrophilic	1500
7	60	1417.5	hydrophilic	1500
8	100	6562.5	hydrophilic	4000
9	5	43.6	hydrophobic	1500
10	20	52.5	hydrophobic	1500
11	5	0.82	hydrophobic	1500
12	40	4200.0	hydrophobic	1500

Figures 4.2 (a) – (h) provide snapshots of runs 1 – 8 (described in **Table 4.4**), respectively which describe the interaction between hydrophilic nanoparticles and lipid membranes. It is apparent that nanoparticle size plays a major role in determining whether or not a hydrophilic nanoparticle will be wrapped by a lipid bilayer. Nanoparticles with diameters of either 5 Å

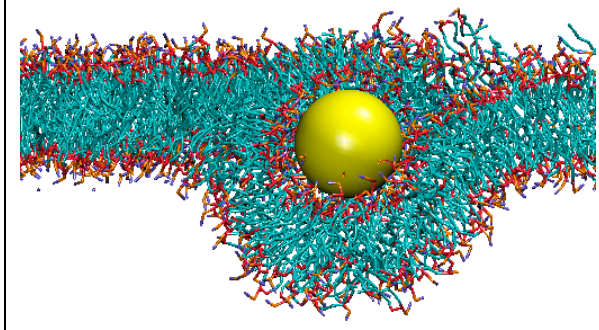
or 10 Å (**Figure 4.2a – 4.2c)** embed themselves within the hydrophilic portion of the bilayers but do not get wrapped. Nanoparticles with diameters of 20, 40, 60 and 100 Å (**Figure 4.2d – 4.2h)** do get wrapped. Two simulations were run on the 5 Å nanoparticles to determine how nanoparticle mass, 0.82amu or 43.6amu, affected the results. Interestingly the mass difference in the two nanoparticles did not change the outcome. In both of the simulations, the nanoparticles embedded themselves in the membrane next to the hydrophilic head groups of the lipids. Simulations were also run to compare the behavior of nanoparticles with a diameter of 40 Å and a mass of 420.0amu or 4200.0amu. Again, there was no significant difference between the outcomes of the two simulations. We conclude that the nanoparticle mass per volume does not significantly the rate or extent to which a hydrophilic nanoparticle is wrapped by a DPPC bilayer.

Figure 4.2: Snapshots of simulations run on the interaction between hydrophilic nanoparticles of different sizes and mass/volume with a DPPC bilayer membrane. (a) run #1, (b) run #2, (c) run #3, (d) run #4, (e) run #5, (f) run #6, (g) run #7, (h) run #8. The color scheme is: purple (DPPC choline entity), orange (DPPC phosphate group), red (DPPC ester groups), cyan (DPPC alkyl tail groups), yellow (nanoparticles).

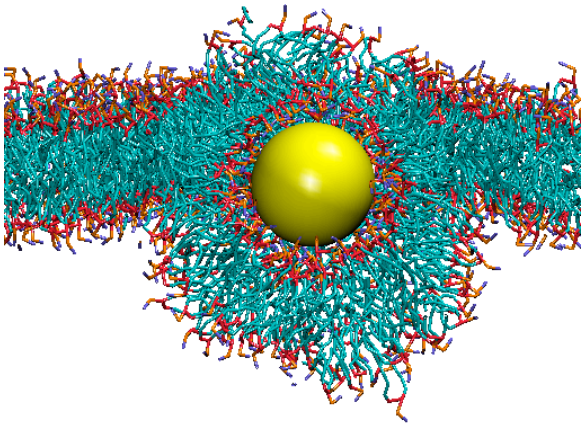


(a) **(b)**

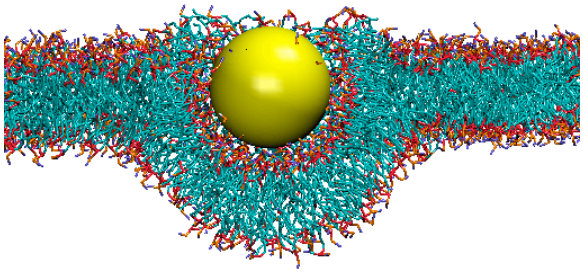
(c) **(d)**



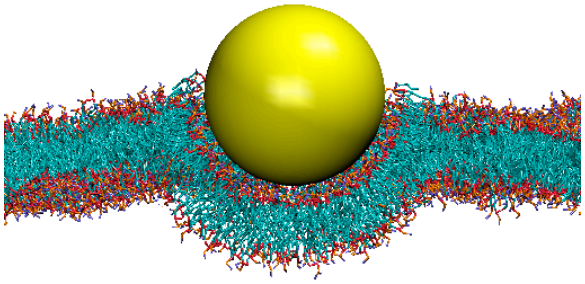
(e)



(f)



(g)



(h)

Figure 4.3 provides snapshots at different time points throughout run #6 in which a hydrophilic nanoparticle with a diameter of 40Å and mass of 4200 amu is wrapped by a bilayer. At 25 million collisions (**Figure 4.3a**) the nanoparticle reaches the surface of the membrane and is subsequently wrapped by the bilayer in **4.3b (**625 million collisions), **4.3c** (1250 million collisions) and **4.3d** (3250 million collisions). The same wrapping process was observed for all other hydrophilic nanoparticles with a diameter greater than 20Å.

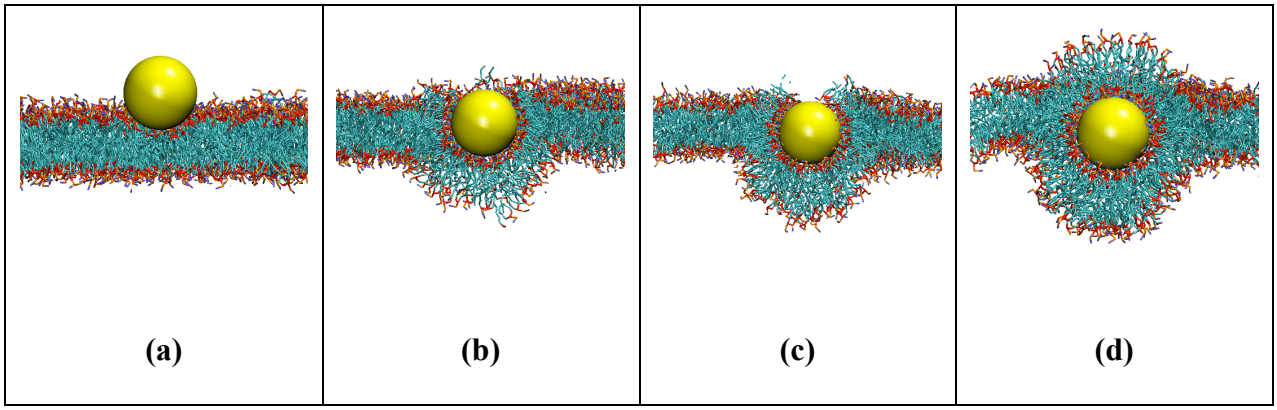


Figure 4.3: Snapshots from run #6 in which a hydrophilic nanoparticle with diameter 40Å is wrapped by a bilayer membrane. The nanoparticle (a) reaches the surface of the bilayer at 25 million collisions, the wrapping process at (b) 625 million collisions, (c) 1250 million collisions and (d) 3250 million collisions.

Figure 4.4 provides simulation snapshots of a hydrophobic nanoparticle with a diameter of 20 Å and the DPPC bilayers at three different time points in run #10. **Figure 4.4a** shows the hydrophobic nanoparticle approaching the surface of the membrane after 25 million collisions. In **Figure 4.4b** the nanoparticle has entered the lipid bilayer at 50 million collisions, however, it is still interacting with both the hydrophilic head groups and

hydrophobic tails of the DPPC lipids. **Figure 4.4c** shows the nanoparticle after it has completely embedded itself within the inner hydrophobic core of the membrane at 275 million collisions. In fact, the nanoparticle completely embeds itself within the membrane by 75 million collisions; there is no visible change in the configuration between the 75 million and 275 million collisions. Since this simulation ran at a rate of approximately 4 million collisions per hour, the time required for the nanoparticle to completely embed itself within the bilayer membrane is approximately 19 CPU hours.

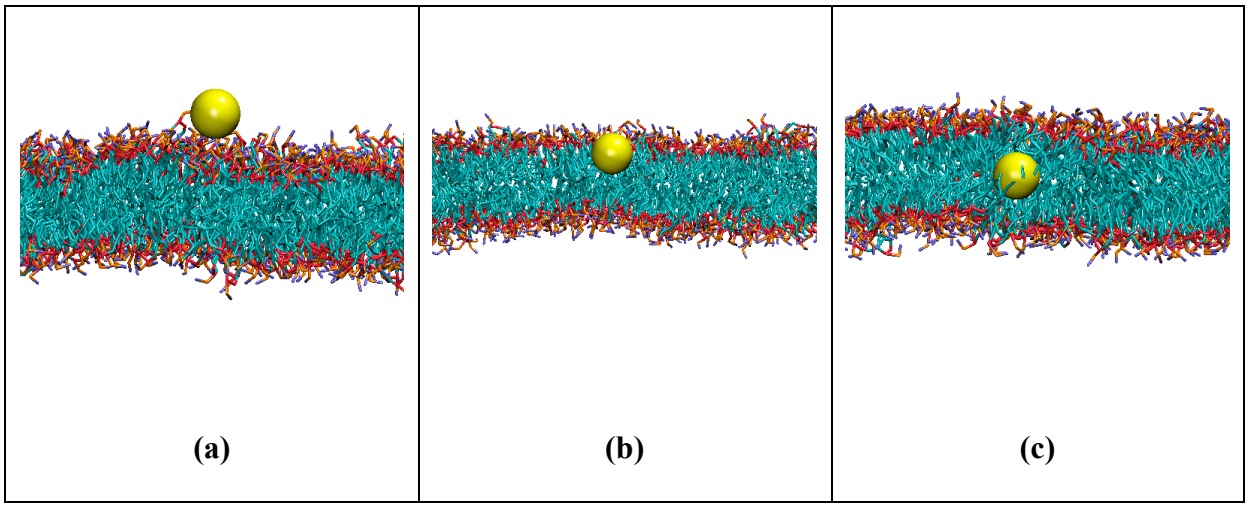


Figure 4.4: Snapshots from run #10 of a simulation of a hydrophobic nanoparticle with a diameter of 20Å and a DPPC bilayer composed of 1500 lipids. The nanoparticle (a) reaches the surface of the bilayer after 25 million collisions, (b) is embedding itself within the membrane after 50 million collisions ,and (c) is fully embedded within the inner hydrophobic core of the membrane after 275 million collisions.

Figure 4.5 provides simulation snapshots of a hydrophobic nanoparticle with a diameter of 40 Å and the DPPC bilayers at three different time points in run #12. **Figure 4.5a** shows the hydrophobic nanoparticle approaching the surface of the membrane after 50

million collisions. In **Figure 4.5b** the nanoparticle has entered the lipid bilayer at 75 million collisions, however, it is still interacting with both the hydrophilic head groups and hydrophobic tails of the DPPC lipids. **Figure 4.5c** shows the nanoparticle after it has completely embedded itself within the inner hydrophobic core of the membrane at 200 million collisions. In fact, the nanoparticle completely embeds itself within the membrane by 125 million collisions; there is no visible change in the configuration between the 125 million and 200 million collisions.

Figure 4.6 provides images from simulations of hydrophobic nanoparticles with diameters of 5Å from runs #9 and #11. The nanoparticle mass is 43.6 amu and 0.82 amu in run #9 and run #11, respectively. The way in which the nanoparticles entered the bilayers in each of these simulations appeared to be the same as in the simulations with larger nanoparticles (runs #10 and #12). Therefore, we found that nanoparticle size does not affect the way in which hydrophobic nanoparticles penetrate the bilayer membrane.

We also compared the time to entry for different sizes of nanoparticles. We measure the time to entry as the time required for a nanoparticle to completely embed itself within a bilayer membrane after reaching the surface of the bilayer. The time to entry for hydrophobic nanoparticles with a diameter of 5Å, 20Å and 40Å was approximately 10, 50 and 75 million collisions, respectively. We did not see a significant difference in the time to entry for runs #9 and #11, which were simulations of nanoparticles with diameters of 5Å and a mass of 43.6 amu and 0.82 amu, respectively. Since the 5Å nanoparticles entered the membrane so quickly it is difficult for us to determine if density plays a significant role in the

time to entry. In the future we would like to compare the time to entry for hydrophobic nanoparticles with diameters larger than 5Å and different mass per volume densities.

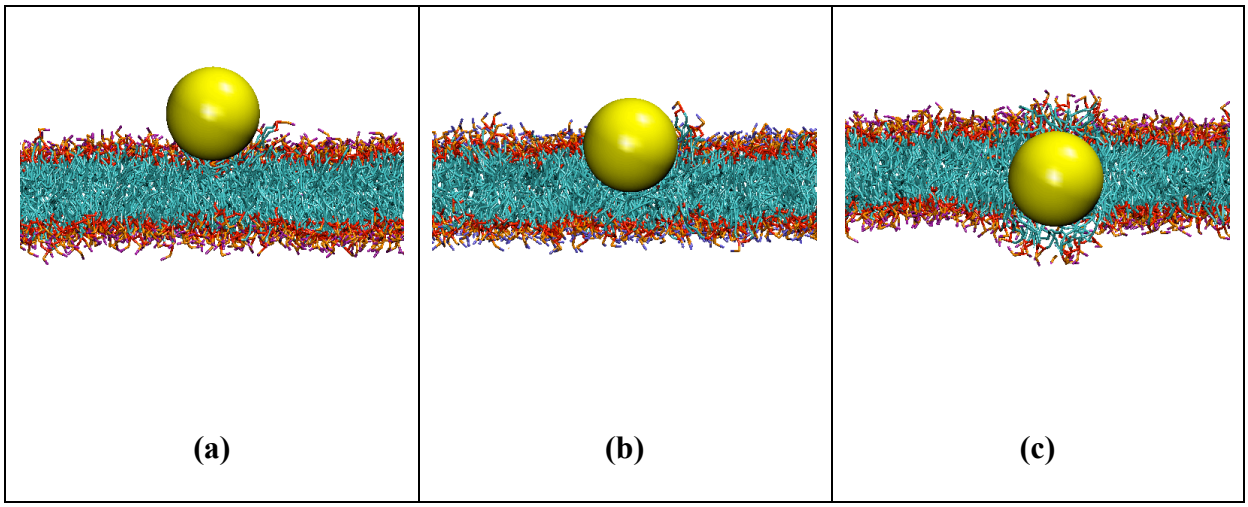


Figure 4.5: Snapshots from run #12 of a simulation of a hydrophobic nanoparticle with a diameter of 40Å and a DPPC bilayer composed of 1500 lipids. The nanoparticle (a) reaches the surface of the bilayer after 50 million collisions, (b) is embedding itself within the membrane after 75 million collisions and (c) is fully embedded within the inner hydrophobic core of the membrane after 200 million collisions.

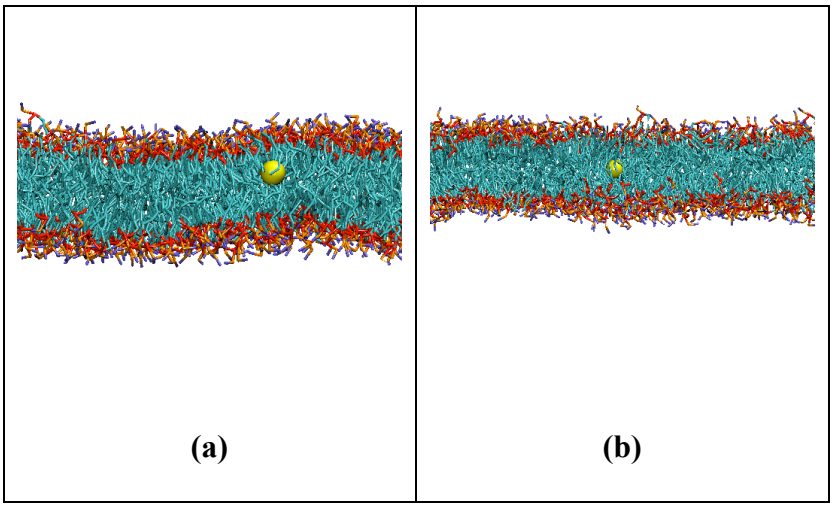


Figure 4.6: Snapshots from simulations of hydrophobic nanoparticles with a mass of 43.6 amu (a) and 0.82 amu (b).

4.4 Conclusion

We describe the extension of LIME, an intermediate-resolution, implicit-solvent, coarse-grained model for phospholipid molecules to systems containing both phospholipids and nanoparticles. For this work we chose to model the behavior of nanoparticles with a wide range of physical properties (sizes, densities and hydrophobicities) to gather preliminary data regarding the interaction between nanoparticles and lipid membranes in our LIME/DMD model. The model is generic, meant to give insights into the general biophysics associated with nanoparticle-membrane interactions. For this reason, we cannot compare our findings to any specific nanoparticle/membrane system because the parameters that we used for the nanoparticle in this model were not based on experimental work or on atomistic simulation.

We chose to study both hydrophilic and hydrophobic nanoparticles to ensure that LIME could accurately model the behavior of nanoparticles with different hydrophobicities. In this work we demonstrated that LIME can be used to model the process by which hydrophilic nanoparticles are either wrapped by a bilayer membrane or implant themselves on the surface of the membrane. We also show how hydrophobic nanoparticles spontaneously penetrate the lipid bilayer to embed themselves within the membrane core.

Since this work was essentially a proof-of-method study to see if LIME/DMD simulations were capable of mimicking the adsorption of a nanoparticle by a lipid membrane, we designed our systems to require minimal computational resources which is why we limited the diameter of our nanoparticles to the size range of 5Å - 100Å. In the future we would like to simulate systems with much larger nanoparticles (10nm – 100nm). However, as the nanoparticle size increases it will be necessary to add more lipid molecules to our system so that there are sufficient lipids to wrap around the nanoparticle. Also, as the size of our system increases the speed of the simulation will decreases and therefore it will take us more time to evaluate the behavior of the nanoparticles in these simulations. Now that we have verified the ability of our model to properly simulate nanoparticle behavior, we are ready to invest the time required to model these larger systems.

Nanoparticle size has also been shown to play a role in cellular uptake. According to Nel and co-workers a threshold radius exists for every particle that is capable of entering a cell, below which cellular uptake is reduced.[1] In addition, nanoparticles are thought to have optimal sizes which help to accelerate the wrapping process.[1,3,27] Values for the threshold radius and the optimal wrapping radius vary depending on nanoparticle properties.

We wanted to verify that our LIME/DMD simulations would show that a membrane would not wrap nanoparticles with a diameter below a critical value. According to our results, a membrane will not wrap nanoparticles with a diameter less than 20Å. Instead these nanoparticles simply embed themselves on the surface of the lipid bilayers. In addition to investigating various sizes, nanoparticles with a wide range of mass per volume (0.013 – 0.67 amu/Å^3) was studied. We saw no difference in the results for nanoparticles with different densities and the same diameter. In the future, we plan to simulate the behavior of specific nanoparticle/membrane systems. For this work, we plan to model the correct size, density and hydrophilicity of the nanoparticle.

One advantage that our LIME/DMD model has over other coarse-grained models reported in the literature is its speed. We demonstrated in our previous publication that LIME allows for the simulation of lipids at the fastest rate reported in the literature.[6] While the LIME/DMD model has many advantages for simulating the interaction between nanoparticles and lipid membranes it does have several disadvantages. Some of these limitations include: (1) electrostatics is not represented explicitly, (2) the use of an implicit solvent approach means that diffusion and hydrodynamics are not well represented, and (3) a direct correlation between reduced temperature and real temperature can only be made at the temperature at which LIME was parameterized.

4.5 References

1. Nel, A.; Madler, L.; Velegol, D.; Xia, T.; Hoek, E.; Somasundaran, P; Klaessig, F.; Castranova, V.; Thompson, M. *Nature Materials*, 2009, *8,* 543 – 557.

2. Pan, Y.; Neuss, S.; Leifert, A.; Fischler, M.; Wen, F.; Simon, U.; Schmid, G.; Brandau, W.; Jahnen-Dechent, W. *Small*, 2007, *11*, 1941 – 1949.

3. Chithrani, B.; Chazani, A.; Chan, W. *Nano Letters*, 2006, *6*, 662 – 668.

4. Win, K.; Feng, S. *Biomaterials*, 2005, *26*, 2713 – 2722.

5. Verma, A.; Uzun, O.; Hu, Y.; Hu, Y.; Han, H.; Watson, N.; Chen, S.; Irvine, D.; Stellacci, F. *Nature Materials*, 2008, *7*, 588 – 595.

6. Curtis, E.; Hall, C. *J. Phys. Chem. B.*, 2013, *117*, 5019-5030.

7. Bedrov, D.; Smith, G.; Davande, H.; Li, L. *J. Phys. Chem. B.*, 2008, *112*, 2078 – 2084.

8. Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.*, 2000, *21*, 86 – 104.

9. Vacha, R.; Martinez-Veracoechea, F.; Frankel, D. *Nano Lett.*, 2011, *11*, 5391 – 5395.

10. Cooke, I. R.; Deserno, M. *J. Chem. Phys.* 2005, *123*, 224710.

11. Limbach, H.; Arnold, A.; Mann, B.; Holm, C. *Comput. Phys. Commun.*, 2006, *174*, 704 – 727.

12. Lin, X.; Li, Y.; Gu, N. *J. of Computational and Theoretical Nanoscience*, 2010, *7*, 269 – 276.

13. Lindahl, E.; Hess, B.; Van der Spoel, D. *J. Mol. Model*., 2001, *7*, 306 – 317.

14. Marrink, S.; de Vries, A.; Mark, A. *J. Phys. Chem. B.*, 2004, *108*, 750 – 760.

15. Marrink, S.; Risselada, H.; Yefimov, S.; Tieleman, D.; de Vries, A. *J. Phys. Chem. B.*, 2007, *111*, 7812 – 7824.

16. Yang, K.; Ma, Y. *Nature Nanotechnology*, 2010, *5*, 579 – 583.

17. Smith, S.W.; Hall, C.K.; Freeman, B.D. *J. Comput. Phys.*, 1997, *134*, 26-30.

18. Alder, B.; Wainwright, T. *J. of Chemical Physics*, 1959, *31*, 459 – 466.

19. Smith, S. W.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.,* 1997, *134*, 16 – 30.

20. Alder, B., and Wainwright, T. 1959. Studies in Molecular Dyanmics I. General Method. Journal of Chemical Physics, **31**, 459-466

21. Rapaport, DC. *J. Phys. A.*: *Math. Gen.*, 1978, *11*, L213 – L217.

22. Rapaport, DC. *J. Chem. Phys.*, 1979, *71*, 3299 – 3303.

23. Nguyen, H.D.; Hall, C.K. *Proc. Natl. Acad. Sci. U.S.A.*, 2004*, 101*, 16180 – 16185.

24. Faller, R.; Marrink, A. *Langmuir*, 2004, *20*, 7686-7693.

25. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Computational Chemistry*, 2005, *26,* 1701 – 1719.

26. Oostenbring, C.; Villa, A.; Mark, A.; Van Gunsteren, W.; *J. of Computational Chemistry*, 2004, *25*, 1656 – 1675.

27. Decuzzi, P.; Ferrari, M. *Biomaterials*, 2007, *28*, 2915 – 2922.

CHAPTER 5

Discontinuous Molecular Dynamics Simulations of DNA Hybridization using "DIME," a New Coarse-Grained Implicit-Solvent Model

5.1 Introduction

The technological advances that have been made to date in the many fields of research aimed at exploiting the unique properties of DNA are astonishing. For example, DNA microarrays are used to measure gene expression levels [1,2,3,4], a wide range of drugs that function by targeting DNA molecules have been developed [5,6,7], and DNA was recently shown to provide an information storage density of approximately 2.2petabytes/g[8]. While significant progress has been made in discovering and exploiting the properties of DNA for a wide range of applications, much remains to be learned about its structure, function and properties.[9] In order to fully realize the potential of the numerous emerging DNA technologies, a tool that would provide better understanding of DNA on a molecular level is needed. Such a tool might aid scientists in their work to design selective sequences for DNA recognition, to improve our current ability to detect DNA damage and mutation, and to optimize DNA structure for specific protein or drug interactions. One popular method that is currently used to study the behavior of biomolecules including DNA on a molecular level is computer simulation.[10,11,12,13]

In this paper we use a multiscale modeling approach to develop a new implicit-solvent intermediate-resolution model for DNA molecules. We call this model "DIME" for DNA Intermediate Resolution Model. The model system chosen for this study is the

Dickerson-Drew dodecamer duplex, which contains the strand CGCGAATTCGCG and its Watson-Crick complement.[14] The multiscale modeling approach used here is very similar to the procedure we reported in our paper describing the development of the LIME force field for lipid molecules.[15] DIME was designed to facilitate the simulation of systems containing large numbers of DNA molecules over long times scale while maintaining sufficient molecular detail to reproduce system dynamics. Similar to LIME, the DIME force field is intended for use with discontinuous molecular dynamics (DMD), which is a very fast alternative to traditional molecular dynamics (MD).

Molecular dynamics studies of nucleic acid systems can be divided roughly into two categories: high-resolution and low-resolution models. High-resolution or atomistic models are based on a realistic representation of molecular geometry and energetics. The advantage of using atomistic models, which typically account for the motion of every single atom in a simulation, to investigate nucleic acid systems, is the high level of detail that can be examined. The disadvantage is that the simulations are generally restricted to stable structures, since processes such as self-assembly, transcription, replication and denaturation occur at longer time scales than those that can be feasibly modeled with current computational resource. A nice example of an atomistic simulation of DNA is work by Mukherjee and co-workers, who simulated the insertion of an anticancer drug between a pair of base pairs [16], the aim being to inhibit DNA replication leading to cell death. The simulations were carried out with the GROMACS simulation package [17] and the AMBER94 force-field.[18] The simulation results suggest a mechanism by which the drug

first binds to the minor groove of DNA (where the backbones of the two strands are close together) and then intercalates into the DNA.[16]

In contrast to high-resolution models, low-resolution or coarse-grained models use a simplified representation of molecular geometry and energetics. In low-resolution models a single coarse-grained site typically represents the behavior of a group of atoms. Knotts and co-workers present a coarse-grained model for DNA in which each nucleotide is represented by three interaction sites, one site each for the phosphate, sugar and base.[19] This model was used to reproduce salt-dependent thermal melting, the dynamics of bubble formation, and the rehybridization process from a bubble structure. (A bubble structure is a single-stranded, melted region bounded on both sides by a double-stranded region.)[19] Sharma and co-workers developed a coarse-grained model to study the DNA-histone complex.[20] Each nucleotide is represented by three coarse-grained spheres, one each for the sugar, phosphate and base molecules. The discontinuous molecular dynamics algorithm was used to simulate the interactions and dynamics of histone tails and DNA. The results show that the histone tails play an important role in stabilizing higher-order chromatin structure.[20]

In this paper we report the first stages of development of a new, intermediate-resolution, implicit-solvent model, "DIME". This new force field was derived using a multiscale-modeling approach. Multiscale modeling is a method in which groups of atoms are represented by a single coarse-grained site and the parameters for these coarse-grained sites are calculated by extracting and analyzing data from atomistic simulations. In DIME three coarse-grained sites are used to represent each nucleotide, one each for the sugar, phosphate and base. This new model was designed for use with discontinuous molecular

dynamics (DMD), which is a fast alternative to traditional molecular dynamics. In DMD, interactions between coarse-grained sites are represented by hard sphere and square well potentials as opposed to the Lennard Jones potentials used in traditional molecular dynamics simulations. The DIME geometrical and energetic parameters for DNA are calculated from radial distribution functions obtained from a 10ns GROMACS [21] simulation with the Amber99sb forcefield [22] of a fully hydrated Dickerson-Drew dodecamer duplex at 300K. In our DIME/DMD simulations, the relative stiffness of the DNA strands are maintained by imposing pseudobonds, which limit the bond length fluctuations to the values observed in the GROMACS simulations. Interaction energies between non-bonded coarse-grained sites are determined by calculating the potential of mean force using a one-step Boltzmann inversion scheme.[23,24] We also investigate an alternative procedure for calculating the square-well width for each pair of interaction sites by setting the second virial coefficient associated with the potential of mean force between coarse-grained sites calculated in the atomistic simulation equal to the second virial coefficient for the site-site square-well potential.[25] Each coarse-grained site in the model has its own realistic mass.

Highlights of our results include the following; The model successfully simulates the spontaneous formation of a double-helical structure from two initially single-stranded Dickerson-Drew dodecamer chains placed at random in a simulation box. The use of DMD with DIME allows us to observe the hybridization event within approximately 0.17 CPU hours. DIME is fast enough to allow for the study of large systems and processes that occur over longer time scales, while providing the resolution or molecular detail required to describe events such as hybridization or protein binding. In addition, an initial pre-formed

duplex structure remains stable in simulations run with parameters obtained using an alternative method to determine square-well widths.

5.2 Model and Methods

In DIME six different coarse-grained types (S, P, C, G, A, T) are used to represent each DNA strand. Types S and P represent the sugar and phosphate groups, respectively. The cytosine, guanine, adenine and thymine bases are assigned types C, G, A, and T, respectively. **Table 5.1** lists the mass and the atoms included in each coarse-grained type. For this preliminary model, terminal nucleotides are assigned the same coarse-grained types as non-terminal nucleotides.

Table 5.1: The type, number of atoms and mass for all of the coarse-grained molecules in the DIME model

Coarse-Grained Molecule	**Coarse-Grained Type**	**Atoms per Coarse-Grained Type**	**Mass of Coarse-Grained Type (amu)**
Sugar	S	5C + 2O + 7H	99.106
Phosphate	P	1P + 3O	78.970
Cytosine	C	4C + 3N + 1O + 4H	110.102
Guanine	G	5C + 5N + 1O + 4H	150.132
Adenine	A	5C + 5N + 4H	134.132
Thymine	T	5C + 2N + 2O + 5H	125.110

Figure 5.1 illustrates the coarse-graining of a Dickerson-Drew dodecamer duplex from 758 atoms to the 70 coarse-grained sites in the DIME representation. This figure and all other figures that are presented throughout the paper were generated with Visual Molecular Dynamics (VMD).[26]

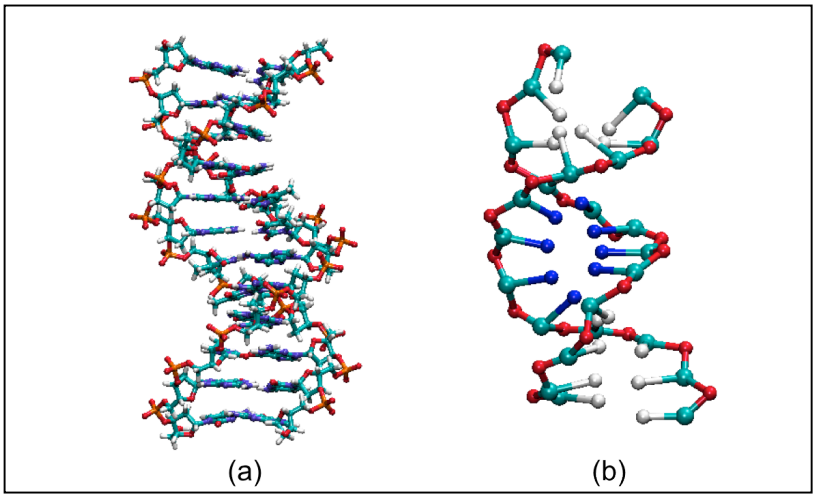


Figure 5.1: (a) Atomistic and (b) coarse-grained representation of a Dickerson-Drew dodecamer duplex. The color scheme for the coarse-grained representation is; cyan (sugar group – type S); red (phosphate group – type P); white (cytosine and guanine bases – types C&G); blue (adenine and thymine bases – types A&T). The coarse-grained size does not represent the actual size of each site.

We employ discontinuous molecular dynamics (DMD) to increase the speed of our code.[27,28] DMD is a very fast alternative to traditional molecular dynamics simulation but is applicable only to systems of molecules that interact via discontinuous potentials, e.g.,

hard-sphere and square-well potentials. For this reason, all of the inter-and intra- molecular interactions in our DNA model are represented by a combination of hard-sphere and square-well potentials, as opposed to the Lennard Jones, Coulombic and harmonic potentials found in traditional molecular dynamics simulations. Expressions for the hard sphere (HS) and square well (SW) potentials between spheres *i* and *j* are, respectively:

$$u_{ij}^{HS}(r) = \begin{cases} \infty & r \le \sigma_{ij} \\ 0 & r > \sigma_{ij} \end{cases} \qquad \text{Equation 1}$$

$$u_{ij}^{SW}(r) = \begin{cases} \infty & r \le \sigma_{ij} \\ -\varepsilon_{ij} & \sigma_{ij} < r \le \sigma\lambda_{ij} \\ 0 & r > \sigma\lambda_{ij} \end{cases} \qquad \text{Equation 2}$$

where r is the distance between spheres, σ_{ij} is the hard sphere diameters, $\sigma\lambda_{ij}$ is the well diameter and ε_{ij} is the well depth. Unlike traditional molecular dynamics simulations on systems with continuous potentials in which Newton's equation of motion must be solved at small time steps, DMD is event driven, meaning that forces only occur when particles collide. This can lead to a significant increase in speed, allowing sampling of much wider regions of conformation space, longer time scales and larger systems than in traditional molecular dynamics. Throughout the simulation the particle trajectories are followed by calculating the time between each collision and advancing the simulation to the next event. Types of events include a collision between two spheres (hard-sphere collision), a bond event when the distance between two bonded spheres reaches a minimum or maximum limit, and a square-well event when two spheres enter (capture), unsuccessfully attempt to escape (bounce) or successfully leave (dissociate) from a square-well attraction.[27,28,29,30] Our

DIME/DMD simulations are performed in the canonical ensemble ; the number of particles, temperature and volume are held constant.

Data used to calculate the coarse-grained parameters was obtained by running an explicit-solvent NPT ensemble GROMACS simulation of the Dickerson-Drew dodecamer duplex. The GROMACS simulation package [21,31], version 4.5.4, was used with the Amber99sb forcefield.[22] The initial configuration of this system was the duplex conformation in a box with equal sides of length 70.0 Å. The Berendsen thermostat [32] was used to keep the temperature constant at 300K throughout the simulation with a time constant of 0.1ps. The simulation was run for 10ns with a time step of 0.002 ps for approximately 12 hours. Periodic boundary conditions were applied and the pressure was maintained at 1.0 bar. Throughout the GROMACS simulations the coordinates of each atom were written to an output trajectory file every 1 ps.

5.3 Results and Discussion

The DIME interaction energies were determined using a one-step Boltzmann inversion procedure described in detail in our previous work.[15] This procedure was inspired by the iterative Boltzmann inversion scheme, which is a popular strategy used to systematically compute potentials for coarse-grained simulations.[23,24] The one-step Boltzmann inversion procedure involves the following steps: (1) the average radial distribution function between two non-bonded coarse-grained sites is determined, (2) the potential of mean force is calculated using ,

$U(r) = -k_B T \ln[g(r)]$ Equation 3

and (3) the minimum value of the potential of mean force between the coarse-grained sites, ε, is chosen to be the depth of the square well potential.

Mathematically, ε is expressed as:

$$\varepsilon = -k_B T \ln[g(r)_{MAX}] \qquad \text{Equation 4}$$

where $g(r)_{MAX}$ is the maximum value of g(r) in the radial distribution function and T is the temperature of the system. If the ε between two coarse-grained sites is greater than -0.005eV, the sites are assumed to have a hard-sphere interaction (ε=0.0eV).

The one-step Boltzmann inversion scheme was used to calculate the ε between all intermolecular coarse-grained sites. The ε for any pair of coarse-grained types was chosen as the lowest value of ε (largest well depth) calculated for coarse-grained site pairs of that type. For example, coarse-grained sites 14&17 are both coarse-grained type A and coarse-grained sites 20&23 are both coarse-grained type T. Therefore, coarse-grained site pairs 14&20, 14&23, 17&20, and 17&23 all represent coarse-grained pairs of type A&T. The ε values for coarse-grained sites 14&20, 14&23, 17&20, and 17&23 were -0.0847, -0.1355, -0.1374 and -0.1048eV. Therefore, we used -0.1374eV as our ε value for coarse-grained types A&T. We chose to select our ε values following this procedure because we found that there were some large discrepancies between the ε values for pairs of coarse-grained sites. The reason for this is that some of the coarse-grained sites did not get a chance to interact often during the GROMACS simulation because they were relatively far apart and therefore had less negative ε values (smaller well depths) than the coarse-grained sites of the same type that did interact frequently. This suggests that the coarse-grained sites that have the smallest ε values (largest

well depths) provide the most accurate value for the ε value between a pair of coarse-grained types. In the future we hope to eliminate large differences between ε values for coarse-grained sites of the same type by running longer GROMACS simulations.

After determining which coarse-grained sites to choose our ε value from, we used the same coarse-grained sites to calculate the hard-sphere diameter (σ_{HS}). For example, the σ_{HS} determined for coarse-grained sites 17&20 was used as our σ_{HS} for coarse-grained types A&T. The σ_{HS} between two coarse-grained sites was determined by locating the smallest non-zero separation between the two sites. The square-well ($\sigma\lambda$) width for each pair of coarse-grained types was determined following the procedure described in our previous work.[15]

Table 5.2 provides the hard-sphere diameters, square-well widths and interaction energies for each pair of intermolecular coarse-grained types. **Table 5.3** provides the hard sphere diameters, square-well widths and interaction energies for each pair of non-bonded intramolecular coarse-grained types.

Table 5.2: The hard-sphere diameters, square-well widths and interaction energies for each pair of intermolecular coarse-grained types.

CG Type i	CG Type j	Intermolecular σ_{ij} (Å)	Intermolecular ε_{ij} (eV)	Intermolecular $\sigma\lambda_{ij}$ (Å)
S	S	6.975	-0.07763	9.355
S	P	7.645	-0.06670	10.905
S	C	9.985	-0.09912	11.045
S	G	8.825	-0.10014	10.085
S	A	9.205	-0.09699	10.465
S	T	10.245	-0.09670	11.325
P	P	10.035	-0.05544	12.275
P	C	11.425	-0.07265	13.605
P	G	10.245	-0.07248	12.825
P	A	10.435	-0.07685	12.535
P	T	11.485	-0.07206	13.385
C	C	5.285	-0.09196	7.905
C	G	5.295	-0.14460	5.875
C	A	4.385	-0.09881	6.785
C	T	9.255	-0.07231	10.855
G	G	3.765	-0.10682	6.705
G	A	6.505	-0.08965	8.405
G	T	5.625	-0.07640	7.645
A	A	6.105	-0.10273	7.525
A	T	5.725	-0.13741	6.265
T	T	5.555	-0.10391	7.855

Table 5.3: The hard-sphere diameters, square-well widths and interaction energies for each pair of non-bonded intramolecular coarse-grained types.

CG Type i	CG Type j	Intramolecular σ_{ij} (Å)	Intramolecular ε_{ij} (eV)	Intramolecular $\sigma\lambda_{ij}$ (Å)
S	S	10.445	-0.07809	12.125
S	P	12.575	-0.06308	15.815
S	C	7.065	-0.07440	10.205
S	G	8.875	-0.07262	11.995
S	A	8.545	-0.07823	10.645
S	T	8.815	-0.07629	10.935
P	P	10.025	-0.06778	14.325
P	C	7.445	-0.07289	10.715
P	G	8.605	-0.08029	10.625
P	A	8.835	-0.08030	10.735
P	T	9.255	-0.07959	11.155
C	C	6.785	-0.09168	10.975
C	G	3.125	-0.14972	3.925
C	A	6.635	-0.09155	8.295
C	T	3.365	-0.12875	4.665
G	G	6.545	-0.09965	7.945
G	A	3.205	-0.14155	4.145
G	T	6.915	-0.08338	8.615
A	A	3.265	-0.14013	4.405
A	T	3.195	-0.14607	3.935
T	T	3.275	-0.13583	4.535

The minimum and maximum bond lengths between bonded coarse-grained sites were also determined from the radial distribution functions. The minimum bond length (σ_{MIN}) was chosen as the smallest possible distance between two bonded coarse-grained sites. The maximum bond length (σ_{MAX}) was chosen as the largest possible distance for which a non-zero g(r) was observed. The σ_{MIN} and σ_{MAX} for a pair of coarse-grained types was chosen as the σ_{MIN} and σ_{MAX} from the pair of coarse-grained sites with the most restrictive σ_{MIN} and

σ_{MAX} values. By most restrictive we mean the pair of σ_{MIN} and σ_{MAX} values that allow the smallest bond fluctuation. **Table 5.4** shows the σ_{MIN} and σ_{MAX} for each pair of coarse-grained types. The relative stiffness of a DNA strand was maintained by imposing pseudobonds, which limit the fluctuation of coarse-grained sites to the angles and torsional angles observed during the GROMACS simulation. Bond angles were maintained by imposing pseudobonds between all next-nearest neighboring sites. Torsional angles were maintained with pseudobonds between all next-next nearest neighboring sites. The minimum and maximum pseudobond lengths were determined from the associated radial distribution functions. The minimum pseudobond length for a pair of coarse-grained sites was chosen as the smallest possible distance for which a non-zero g(r) was detected. The maximum pseudobond length for a pair of coarse grained sites was chosen as the largest distance at which a non-zero g(r) was observed. The pair of coarse-grained sites with the most restrictive pseudobond values (smallest distance between minimum and maximum pseudobond lengths) was chosen to represent the pseudobond lengths for a pair of coarse-grained types. **Table 5.5** provides the minimum and maximum pseudobond lengths for bond angles between pairs of different coarse-grained types. **Table 5.6** provides the minimum and maximum pseudobond lengths for the torsional angles between pairs of different coarse-grained types.

Table 5.4: Minimum and maximum bond lengths for pairs of coarse-grained types

Bonds	**Minimum Bond Length (σ_{MIN}) (Å)**	**Maximum Bond Length (σ_{MAX}) (Å)**
$S_i - P_{i+1}$	3.495	3.875
$P_i - S_i$	3.945	4.605
$S_i - C_i$	4.345	4.875
$S_i - G_i$	4.915	5.475
$S_i - A_i$	4.855	5.375
$S_i - T_i$	4.465	4.945

Table 5.5: Minimum and maximum bond lengths for bond angles between different pairs of coarse-grained types

Bond Angles	**Minimum Pseudobond Length (Å)**	**Maximum Pseudobond Length (Å)**
$S_i - P_{i+1} - S_{i+1}$	5.255	6.645
$P_i - S_i - P_{i+1}$	6.075	7.715
$C_i - S_i - P_{i+1}$	6.485	7.995
$G_i - S_i - P_{i+1}$	6.915	8.445
$A_i - S_i - P_{i+1}$	6.805	8.195
$T_i - S_i - P_{i+1}$	6.725	7.915
$P_i - S_i - C_i$	5.695	8.285
$P_i - S_i - G_i$	6.225	9.105
$P_i - S_i - A_i$	6.075	8.885
$P_i - S_i - T_i$	5.615	8.065

Table 5.6: Minimum and maximum bond lengths for torsional angles between different pairs of coarse-grained types

Torsional Angles	**Minimum Pseudobond Length (Å)**	**Maximum Pseudobond Length (Å)**
$S_i - P_{i+1} - S_{i+1} - P_{i+2}$	8.595	10.095
$P_i - S_i - P_{i+1} - S_{i+1}$	8.865	10.885
$C_i - S_i - P_{i+1} - S_{i+1}$	5.525	8.705
$G_i - S_i - P_{i+1} - S_{i+1}$	5.305	8.125
$A_i - S_i - P_{i+1} - S_{i+1}$	5.395	7.905
$T_i - S_i - P_{i+1} - S_{i+1}$	5.895	8.185
$S_i - P_{i+1} - S_{i+1} - C_{i+1}$	5.495	8.185
$S_i - P_{i+1} - S_{i+1} - G_{i+1}$	5.665	8.765
$S_i - P_{i+1} - S_{i+1} - A_{i+1}$	5.375	8.325
$S_i - P_{i+1} - S_{i+1} - T_{i+1}$	5.445	7.695

Simulation temperature in LIME is expressed in terms of the reduced temperature, T*, which is defined to be:

$$T^* = k_B T/\varepsilon^* \qquad \text{Equation 5}$$

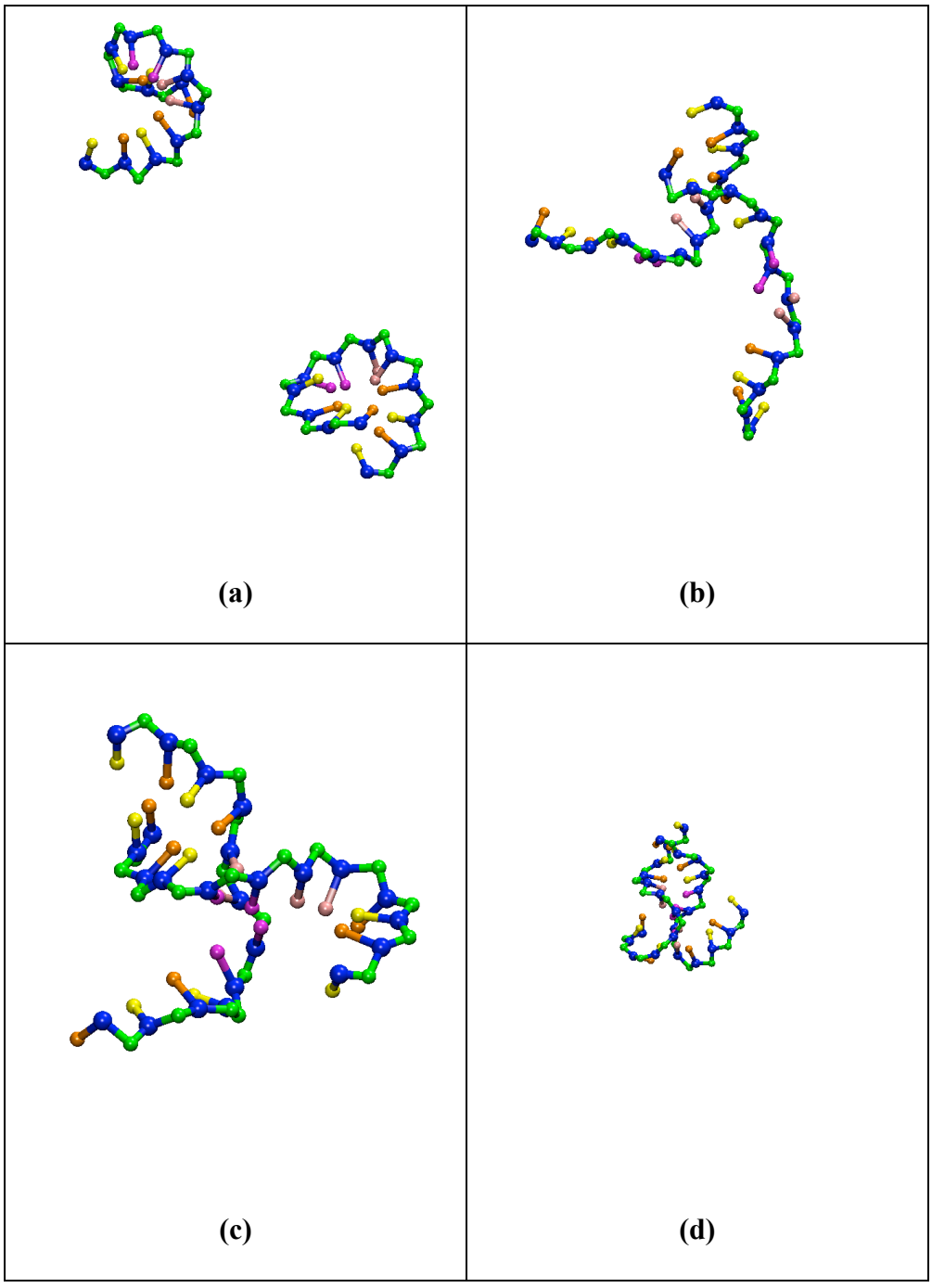
where k_B is Boltzmann's constant, T is the temperature, and ε* is the reference interaction strength.[33] The reference interaction strength, ε*, is chosen as the absolute value of the

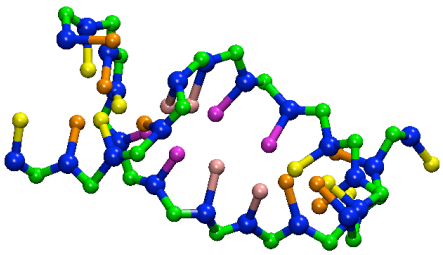
largest well-depth. For this work, ε*=0.1446eV, which is the absolute value of the largest interaction energy between site types C&G. Thus, $T^* = k_B T/\varepsilon^* = (8.6173 \times 10^{-5} eV/K)*(325K)/(0.1446eV) = 0.2$ in our DIME/DMD simulations. The Andersen thermostat is used to hold the temperature constant.[34] In this method randomly selected particles collide infrequently with ghost particles, effectively reassigning the particle's velocity randomly so as to maintain a Maxwell-Boltzmann distribution centered at the simulation temperature. All DIME/DMD simulations were run with a DMD software program developed in the Hall research lab called EMBLEM. This program is written in C++. The Intel compiler was used to compile this code and all other codes used in the development and analysis of DIME. All simulations were run in serial.

To study the DNA hybridization process, five independent DMD simulations were run starting from different random configurations of two single Dickerson-Drew dodecamer strands at a T*=0.2. The starting configurations for these stands were obtained by coarse-graining the atomistic coordinates from one of the trajectories saved from the Dickerson-Drew dodecamer duplex GROMACS simulation. The lengths of the sides of the simulation cell were set to 70Å. A double helix was formed in all five simulations. The average time required to form the double helix was approximately 0.17 CPU hours. Snapshots of one system at (a) 680,027,000 (b) 685,726,000 (c) 685,833,000 (d) 685,997,000 (e) 686,139, 000 (f) 689,497,000 collisions during the simulation are provided in **Figure 5.2**. The following color scheme is used to represent each of the coarse-grained types: S (blue), P (green), C (yellow), G (orange), A (pink), T (magenta). The two strands first come into contact at

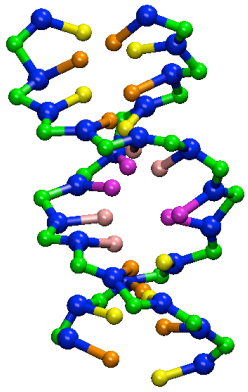
approximately 685,726,000 collisions and require approximately 3,771,000 collisions to complete the formation of the duplex structure.

Figure 5.2: Snapshots from a simulation of the spontaneous formation of a Dickerson-Drew dodecamer duplex formation. The following color scheme is used to represent each of the coarse-grained types: S (blue), P (green), C (yellow), G (orange), A (pink), T (magenta). (a) – (f) = 680,027,000, 685,726,000, 685,833,000, 685,997,000, 686,139, 000 and 689,497,000 collisions. The system is started from a random configuration. (a) After 680,027,000 collisions the strands have not yet begun to interact. (b) After 685,726,000 collisions the strands contact each other (f) After approximately 3,771,000 collisions the duplex is formed


(a)
(b)
(c)
(d)



(e)



(f)

.

Calculating Lamda Using the 2nd Virial Coefficient

We chose to investigate an alternative procedure for calculating the square-well diameter ($\sigma\lambda$) for each pair of interaction sites by setting the second virial coefficient associated with the potential of mean force between coarse-grained sites calculated in the atomistic simulations equal to the second viral coefficient for the site-site square –well potential. Recall that the second virial coefficient, $B_2(T)$ for a system of molecules interacting via a potential u(r) is given by:

$$B_2(T) = -2\pi \int_0^{r^*} \left(e^{-\beta u(r)} - 1\right) r^2 dr \qquad \text{Equation 6}$$

where *k* is the Boltzmann constant, and *T* is the temperature. If we take u(r) to be the potential of mean force calculated using equation (3) then this becomes:

$$B_2(T) = -2\pi \int_0^{r^*} (g(r) - 1) r^2 dr \qquad \text{Equation 7}$$

where g(r) is the radial distribution function, r is the radius, σ is the hard-sphere diameter also referred to as σ_{HS}, *k* is the Boltzmann constant, and *T* is the temperature.[25] If instead we take u(r) to be a square well potential as given in equation (2) we obtain the second virial coefficient for a square-well interaction:

$$B_2(T) = b_o\left\{1 - (\lambda^3 - 1)\left(e^{\beta\varepsilon} - 1\right)\right\} \qquad \text{Equation 8}$$

where $b_O = 2\pi\sigma^3/3$ and β=-1/(*k*T). We determined the value of λ by the following steps. First we numerically integrated the 2nd virial coefficient in **Equation 7** using the trapezoid rule. The g(r) for the coarse-grained sites evaluated to determine the σ_{HS} and ε values for a given pair of coarse-grained types was used. The integral was truncated after the first peak at the first value of r where g(r) was less than 1. The square–well $B_2(T)$ (Equation 8) was set equal

to the value calculated from **Equation 7**. The σ_{HS} and ε values previously calculated for a pair of coarse-grained types were used to determine λ. Once λ was calculated it was multiplied by σ_{HS} to determine the $\sigma\lambda$ for a coarse-grained type pair. The reason that we truncated our values of g(r) in evaluating Equation 7 is that the simulation values of g(r) did not approach 1 as r goes to infinity. Instead they went to 0 due to finite size effects. The reason for this is that the only molecules in our system are in the duplex: there are no bulk molecules in the simulation cell. This is why it was necessary to truncate the integral used to calculate the 2^{nd} virial coefficient before g(r) reached 0. **Table 5.6** provides the values for the intermolecular and intramolecular square-well diameters calculated from the 2^{nd} virial coefficient method.

Table 5.7: The intermolecular and intramolecular square-well diameter calculated from the 2nd virial coefficient method.

CG Type i	CG Type j	2nd Virial Intermolecular $\sigma\lambda_{ij}$ (Å)	2nd Virial Intramolecular $\sigma\lambda_{ij}$ (Å)
S	S	8.215	11.090
S	P	9.056	14.635
S	C	10.320	8.388
S	G	9.232	9.877
S	A	9.625	9.455
S	T	10.590	9.742
P	P	11.295	12.635
P	C	12.084	8.773
P	G	11.046	9.433
P	A	11.105	9.641
P	T	12.152	10.015
C	C	6.567	7.652
C	G	5.504	3.571
C	A	5.727	7.545
C	T	10.210	4.157
G	G	5.070	7.253
G	A	7.497	3.769
G	T	7.507	8.022
A	A	6.818	3.840
A	T	5.961	3.683
T	T	6.364	3.932

A test DMD simulation was run using the new $\sigma\lambda$ calculated using the 2nd virial coefficient. This simulation was started from a preformed Dickerson-Drew dodecamer duplex to test if the new $\sigma\lambda$ values would allow the duplex to remain intact. The coordinates for the initial duplex were taken from one of the DMD/DIME simulations discussed previously. The simulation cell length was 70Å and the reduced temperature was set to 0.2. The duplex did not remain intact and the two strands separated. We

decided to repeat this simulation at a lower reduced temperature of 0.05. This reduced temperature was chosen arbitrarily, to test whether or not the duplex would hold at a much lower reduced temperature than the previously tested reduced temperature of 0.2. Again, the simulation was started from a preformed Dickerson-Drew dodecamer duplex in a cell with equal sides of 70Å. This simulation was run for 20 million collisions and the duplex remained intact throughout this simulation. **Figure 5.3** shows a snapshot of the duplex structure that was maintained throughout this simulation. The color scheme is the same as in **Figure 5.2**.

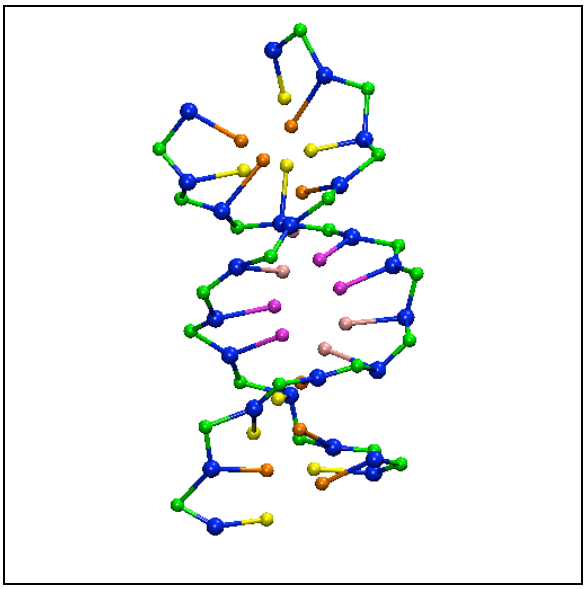


Figure 5.3: Snapshot from a simulation started from a Dickerson-Drew dodecamer duplex at T* = 0.05. The color scheme for each coarse-grained type is as follows: S (blue), P (green), C (yellow), G (orange), A (pink), T (magenta).

Next, we ran 3 replicate simulations to determine if the intermolecular and intramolecular $\sigma\lambda$ values calculated using the second virial coefficient method would result in the spontaneous formation of a duplex. The starting configuration for this simulation was two Dickerson-Drew dodecamer strands placed randomly in a cell with equal lengths of 70Å. The geometry for these strands was obtained by coarse-graining the atomistic coordinates from one of the trajectories saved from the GROMACS simulation. The simulation was run at $T^* = 0.05$. A duplex did not form during any of the replicate simulations which were each run for 1 billion collisions. Instead, the two DNA strands formed a structure, similar to the one shown in **Figure 5.4** in all three replicate simulations almost immediately after coming into contact. In this structure, some of the terminal bases on one strand appear to interact with a few of the terminal bases on the other strand. In addition, some of the bases on each strand appear to have intramolecular interactions with each other. One concern that we do have is the extremely fast rate (almost immediately after coming into contact) at which the DNA strands formed the structure shown in **Figure 5.4**. It is possible that a $T^* = 0.5$ is forcing the strands to freeze in a conformation instead of allowing them to adopt a helical duplex structure. In the future we plan to run simulations using the parameters calculated from the second virial coefficient method at a range of reduced temperatures.

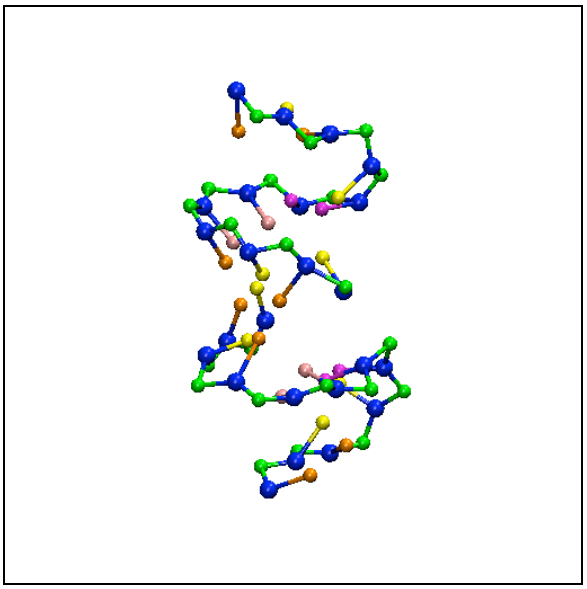


Figure 5.4: Snapshot at 300 million collisions of the structure formed during a simulation that started from an initial random configuration of two Dickerson-Drew dodecamer strands at a T*=0.05 in a cell with equal box lengths of 70.Å The color scheme for each coarse-grained type is as follows: S (blue), P (green), C (yellow), G (orange), A (pink), T (magenta).

5.4 Conclusion

We described the development of DIME, an intermediate-resolution, implicit-solvent, coarse-grained model for DNA designed for use with discontinuous molecular dynamics. In DIME, each nucleotide is represented by 3 coarse-grained sites. Each of these sites is assigned 1 of 6 different types: the sugar entity, the phosphate group, or the cytosine, guanine, adenine or thymine base. The "DIME" parameters were obtained using a multiscale modeling approach in which the geometric and energetic parameters were calculated from data collected from a GROMACS simulation of a Dickerson-Drew dodecamer duplex. A one-step Boltzmann inversion approach, which is a simplified version of the iterative

Boltzmann inversion scheme, was used to calculate the depth of the square well interactions in the model.

Two methods were used to calculate λ values for this model. In the first method the σλ values are chosen by examining the radial distribution functions between pairs of non-bonded coarse-grained sites in the GROMACS simulation of the duplex. In the second method the σλ values are calculated using the second virial coefficient. In this procedure the λ values are calculated by setting the second virial coefficient associated with the potential of mean force between coarse-grained sites calculated in the atomistic simulations equal to the second virial coefficient for the site to site square-well potential. Simulations run using the intermolecular and intramolecular λ values calculated using the first method showed the formation of a duplex starting from an random initial configuration of two Dickerson-Drew dodecamer strands. A simulation run using the σλ values calculated with the 2^{nd} virial coefficient that started from an initial duplex structure showed that the duplex structure was maintained throughout the simulation. However, preliminary simulations run using σλ values calculated with the 2^{nd} virial coefficient starting from random configurations of Dickerson-Drew dodecamer strands did not spontaneously form a duplex.

5.5 References

1. Schena, M.; Shalon, D.; Davis, R.; Brown, P. *Science*, 1995, *270*, 467 – 470.
2. DeRisi, J.; Penland, L.; Brown, P.; Bittner, M.; Meltzer, P.; Ray, M.; Chen, Y.; Su, Y.; Trent, J. *Nature Genetics*, 1996*, 14*, 457 – 460.
3. Lockhart, D.; Winzeler, E. *Nature*, 2000, *405*, 827 – 836.

4. Chon, H.; Lancaster, J. *Cancer Control*, 2011, *18*, 8 – 15.

5. Hurley, L. *Nature Reviews Cancer*, 2002, *2*, 188 – 200.

6. Palchaudhuri, R.; Hergenrother, P. *Current Opinion in Biotechnology*, 2007, *18*, 497 – 503.

7. Koster, D.; Palle, K.; Bot, E.; Gjornsti, M.; Dekker, N. *Nature*, 2007, *448*, 213 – 217.

8. Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.; Sipos, B.; Birney, E. *Nature*, 2013, *494*, 77-80.

9. Perez, A.; Luque, J.; Orozco, M. *Accounts of Chemical Research*, 2012, *45*, 196 – 205.

10. Jayaraman, A.; Hall, C.; Genzer, J. *J. Chem. Phys.*, 2007, *127*, 144912.

11. Lee, O.; Schatz, G. *J. Phys. Chem. C.*, 2009, *113*, 2316 – 2321.

12. Bueren-Calabuig, J.; Giraudon, C.; Galmarini, C.; Egly, J.; Gago, F. *Nucl. Acids Res.*, 2011, *39*, 8248 – 8257.

13. Ziebarth, J.; Wang, Y. *Biophysical Journal*, 2009*, 97*, 1971 – 1983.

14. Drew, H.; wing, R.; Rakano, R.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. *Proc. Natl. Acad. Sci.*, 1981, *78*, 2179 – 2183.

15. Curtis, E.; Hall, C. *J. Phys. Chem. B.*, 2013, *117*, 5019-5030.

16. Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. *J. Am. Chem. Soc.*, 2008, *130*, 9747 – 9755.

17. Berendsen, H.; Spoel, D.; Vandrunen, R. *Comput. Phys. Commun.*, 1995, *91,* 43 – 56.

18. Cornell, W.; Cieplak, P.; Bayly, C.; Gould, I.; Merz, K.; Ferguson, D.; Spellmeyer, D.; Fox, T.; Caldwell, J.; Kollman, P. *J. Am. Chem. Soc.*, 1995, *117*, 5179 – 5197.

19. Knotts, T.; Rathor, N.; Schwartz, D.; de Pablo, J. *The Journal of Chemical Physics*, 2007, *126*, 084901.

20. Sharma, S.; Ding, F.; Dokholyan, N. *Biophysical Journal*, 2007, *92*, 1457 – 1470.

21. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. *J. Comput. Chem.* 2005, *26*, 1701 – 1719.

22. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *PROTEINS: Structure, Function, and Bioinformatics*, 2006, *65*, 712 – 725.

23. Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.*, 2003, *24*, 1624 – 1636.

24. Chennamsetty, N.; Bock, H.; Gubbins, K. *J. Chem. Phys.,* 2006, *124*, 074105-1 – 074105-12.

25. McQuarrie, Donald A. *Statistical Mechanics*. New York: Harper & Row, 1973. Print.

26. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics*, 1996, *14*, 33 – 38.

27. Alder, BJ.; Wainwright, TE. *J. Chem. Phys.*, 1959, *31*, 459 – 466.

28. Smith, S. W.; Hall, C. K.; Freeman, B. D. *J. Comput. Phys.,* 1997, *134*, 16 – 30.

29. Rapaport, DC. *J. Chem. Phys.*, 1979, *71*, 3299 – 3303.

30. Rapaport, DC. *J. Phys. A: Math. Gen.*, 1978, *11,* L213 – L217.

31. Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* 2008, *4*, 435 – 447.

32. Berendsen, H.; Postma, J.; Van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.*, 1984, *81*, 3684 – 3690.

33. Nguyen, H.D.; Hall, C.K. *Proc. Natl. Acad. Sci. U.S.A.*, 2004, *101*, 16180 – 16185.

34. Andersen, HC. *J. Chem. Phys.,* 1980, 72, 2384 – 2393.

CHAPTER 6

Future Work

We developed new intermediate-resolution implicit-solvent models for lipids, "LIME," and DNA molecules, "DIME, " designed for use with discontinuous molecular dynamics (DMD) simulations. We applied LIME to study the spontaneous formation of lipid bilayers, the behavior of mixed lipid systems at different pH values and the interaction between membranes and nanoparticles. DIME was used to investigate the structural properties of DNA and the process by which two DNA strands hybridize in solution. In this chapter we discuss possible directions for future work based upon our findings in Chapters 2 through 5.

6.1 Investigation of the Change in Orientation Observed for DSPS Lipids in 21PC/DSPS Bilayers at Low pH

In one of simulations of 21PC/DSPS at low pH, the DSPS lipids changed their orientation with respect to the bilayer from a vertical alignment with the z-axis to a horizontal alignment. This realignment was not seen in any of the neutral pH simulations of DPPC/DSPS or 21PC/DSPS or in any of the low pH simulations of DPPC/DSPS. This realignment of DSPS chains may have inflated the numbers of DSPS lipids with a DSPS nearest neighbor that we calculated for our 21PC/DSPS low pH system. We plan to investigate why the realignment of the DSPS lipids occurred. We will attempt to determine if our interaction energies at low pH are too large or if our square-well widths are too wide.

We will also run simulations to ensure that this realignment is not an artifact of the small system size that we are using to study the phase separation in bilayers.

6.2 Expanding LIME to Include Parameters for Cholesterol

LIME simulations of DPPC/DSPS bilayers showed similar levels of phase separation at both neutral and low pH. In addition, with the exception of the snapshot of the 21PC/DSPS system at low pH in which the DSPS lipids changed their orientation with respect to the bilayer, snapshots of 21PC/DSPS bilayers also showed approximately the same levels of phase separation at both neutral and low pH. This is similar to the experimental findings of Sofou and co-workers who found that liposomes composed of DOPC and DSPS without cholesterol had the same level of phase separation at pH values of 5.0, 6.0 and 7.0.[33] Their results also showed that liposomes composed of DOPC, DSPS and cholesterol had a higher level of phase separation at a pH of 5.0 than at a pH of 7.0. Therefore, it appears that cholesterol plays an important role in the phase separation process for PC and PS lipids. In the future we would like to use our multi-scale modeling approach to extract LIME parameters for cholesterol from atomistic simulations. This would allow us to run simulations to try to understand more about the interaction between cholesterol and lipids on a molecular level and how cholesterol affects the phase separation between PC and PS lipids at different pH values.

6.3 Expanding LIME to Include Coarse-Grained Parameters for Doxorubicin

In the future we would like to run atomistic simulations to obtain coarse-grained parameters for doxorubicin molecules. Currently, in our simulations of liposomes containing

doxorubicin we use an atomistic representation of doxorubicin to model the drug geometry. In this representation 67 sites are used to represent each doxorubicin molecule. These sites are connected in such a way that the molecule is relatively inflexible. All interactions between the lipids and doxorubicin molecules are modeled as hard sphere potentials. We would like to run atomistic simulations of doxorubicin/lipid systems so that we can determine square-well parameters for lipid/doxorubicin and doxorubicin/doxorubicin interactions. This will help us to more accurately model the behavior of these drug molecules in liposomal systems and make our simulations more realistic.

6.4 Applying LIME to Simulate the Behavior of a Specific Nanoparticle/Membrane System

We would like to perform nanoparticle/bilayer simulations that represent realistic systems. In this work we demonstrated how LIME/DMD simulations were capable of mimicking the way in which bilayers wrap hydrophilic nanoparticles and how hydrophobic nanoparticles directly penetrate membranes to embed themselves within the inner hydrophobic core. Now that we have verified the ability of our model to properly simulate nanoparticle behavior, we are ready to invest the time required to model larger systems. In the future we would like to simulate systems with much larger nanoparticles (10nm – 100nm). We would also like to study the merging of two membranes, including the process that occurs when a liposome reaches the surface of a membrane.

6.5 Further Refinement of DIME

In the multi-scale modeling procedure used in DIME, the coarse-grained parameters were extracted from a single atomistic simulation of a Dickerson-Drew dodecamer duplex in explicit solvent. While we were able to show that these parameters could be used to successfully model the hybridization of two single DNA strands, it would be better to calculate the DIME parameters from a larger number of atomistic simulations that contain different DNA sequences, different lengths of DNA strands, and multiple single strand systems as well as duplexes. This will help us to ensure that we are accurately parameterizing the DNA molecules. In addition to refining our DIME parameters based on a larger number of atomistic simulations, we would like to further explore the use of the second virial coefficient method to obtain square-well widths. We were able to show that square-well widths calculated using the second virial coefficient maintained the helical structure in a simulation started from an initial duplex conformation. However, the parameters did not result in the spontaneous formation of a double helix. We would like to investigate why simulations with the second virial coefficient square-well widths did not succeed in predicting bridization . We will try to determine if the parameters can be refined to better model the duplex formation process. We will also run simulations to ensure that the structure that we do see form is not a consequence of looking at a reduced temperature that is too high. .

6.6 References

1. Bandekar, A.; Sofou, S. *Langmuir*, 2012, *28*, 4113 – 4122.