Title
Cellular membrane dynamics and algorithms for studying their interactions with pharmaceutical compounds

Permalink
https://escholarship.org/uc/item/8jn062ps

Author
Madej, Benjamin D.

Publication Date
2015

Peer reviewed|Thesis/dissertation
Cellular membrane dynamics and algorithms for studying their interactions with pharmaceutical compounds

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Benjamin D. Madej

Committee in charge:

Professor J. Andrew McCammon, Co-Chair
Professor Ross C. Walker, Co-Chair
Professor Michael D. Burkart
Professor Michael K. Gilson
Professor Wei Wang

2015
The Dissertation of Benjamin D. Madej is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California, San Diego

2015
DEDICATION

To my friends who have helped me through graduate school.

To my family who have always been on my side.

To Mom and Dad for always looking out for me.

To Maria who makes me laugh and who makes me happy. Thank you for always having pizza in the freezer!
The universe is full of magical things patiently waiting for our wits to grow sharper.

*Eden Phillpotts*
# TABLE OF CONTENTS

Signature Page ......................................................... iii
Dedication ................................................................. iv
Epigraph ............................................................... v
Table of Contents .................................................... vi
List of Figures ........................................................... x
List of Tables ........................................................... xiii
Acknowledgements ...................................................... xiv
Vita ................................................................. xvi
Abstract of the Dissertation ........................................ xviii

1 Introduction .......................................................... 1
  1.1 Guide to this dissertation ........................................ 1
  1.2 Cellular membranes ............................................. 3
    1.2.1 Membrane developments ..................................... 4
    1.2.2 Membrane rafts .............................................. 6
  1.3 Membrane proteins .............................................. 8
    1.3.1 Protein lipid modifications ................................. 11
  1.4 Membrane transport and permeability ........................ 11
  1.5 Computational and theoretical studies of membranes .......... 12
    1.5.1 Membrane molecular dynamics .............................. 13

2 Molecular dynamics ............................................... 15
  2.1 Quantum mechanics and Newtonian dynamics .................. 15
  2.2 Molecular mechanics force fields .............................. 18
  2.3 Parameterization .............................................. 19
  2.4 Parameters ..................................................... 20
    2.4.1 Atom types ................................................ 21
    2.4.2 Bonds ..................................................... 21
    2.4.3 Angles ..................................................... 24
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.4 Dihedrals</td>
<td>25</td>
</tr>
<tr>
<td>2.4.5 Improper dihedrals</td>
<td>26</td>
</tr>
<tr>
<td>2.4.6 Cross terms</td>
<td>27</td>
</tr>
<tr>
<td>2.5 Non-bonded interactions</td>
<td>27</td>
</tr>
<tr>
<td>2.5.1 Electrostatics</td>
<td>28</td>
</tr>
<tr>
<td>2.5.2 RESP charge derivations</td>
<td>29</td>
</tr>
<tr>
<td>2.5.3 Van der Waals interactions</td>
<td>30</td>
</tr>
<tr>
<td>2.5.4 Many-body effects</td>
<td>32</td>
</tr>
<tr>
<td>2.6 Long range interactions</td>
<td>32</td>
</tr>
<tr>
<td>2.6.1 Periodic boundary conditions</td>
<td>32</td>
</tr>
<tr>
<td>2.6.2 Non-bonded cutoffs</td>
<td>33</td>
</tr>
<tr>
<td>2.6.3 Particle mesh Ewald</td>
<td>34</td>
</tr>
<tr>
<td>2.7 Solvent models</td>
<td>37</td>
</tr>
<tr>
<td>2.8 Energy surfaces in molecular mechanics</td>
<td>38</td>
</tr>
<tr>
<td>2.8.1 Energy minimization</td>
<td>39</td>
</tr>
<tr>
<td>2.9 Molecular dynamics algorithm</td>
<td>41</td>
</tr>
<tr>
<td>2.9.1 Constant temperature and pressure dynamics</td>
<td>43</td>
</tr>
<tr>
<td>3 Lipid11</td>
<td>45</td>
</tr>
<tr>
<td>3.1 A modular framework for phospholipids</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Initial parameters and atom types</td>
<td>49</td>
</tr>
<tr>
<td>3.3 Preliminary charge derivation</td>
<td>51</td>
</tr>
<tr>
<td>3.4 Lipid residue names and atom types</td>
<td>56</td>
</tr>
<tr>
<td>3.4.1 Phospholipid nomenclature</td>
<td>59</td>
</tr>
<tr>
<td>3.4.2 Other molecules</td>
<td>60</td>
</tr>
<tr>
<td>3.4.3 Lipid11 PDB file format</td>
<td>61</td>
</tr>
<tr>
<td>3.5 Lipid bilayer molecular dynamics</td>
<td>61</td>
</tr>
<tr>
<td>3.5.1 Lipid bilayer simulation methods</td>
<td>63</td>
</tr>
<tr>
<td>3.6 Lipid bilayer experimental properties</td>
<td>65</td>
</tr>
<tr>
<td>3.6.1 Area per lipid</td>
<td>65</td>
</tr>
<tr>
<td>3.6.2 Electron density profile and thickness</td>
<td>66</td>
</tr>
<tr>
<td>3.6.3 Deuterium order parameters</td>
<td>67</td>
</tr>
<tr>
<td>3.7 Lipid11 structural properties</td>
<td>68</td>
</tr>
<tr>
<td>3.8 Implications for further parameterization</td>
<td>74</td>
</tr>
<tr>
<td>3.9 Graphics processing units, lipid bilayers, and long-range interactions</td>
<td>75</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.1</td>
<td>Glycerophospholipid structure</td>
</tr>
<tr>
<td>1.2</td>
<td>Common membrane components</td>
</tr>
<tr>
<td>1.3</td>
<td>Rhodopsin protein structure in a membrane</td>
</tr>
<tr>
<td>2.1</td>
<td>Bond potentials for bond stretching</td>
</tr>
<tr>
<td>2.2</td>
<td>Dihedral potential for dihedral rotation</td>
</tr>
<tr>
<td>2.3</td>
<td>Van der Waals potential</td>
</tr>
<tr>
<td>2.4</td>
<td>Periodic boundary conditions</td>
</tr>
<tr>
<td>2.5</td>
<td>Ewald sum</td>
</tr>
<tr>
<td>3.1</td>
<td>A framework for modular phospholipids</td>
</tr>
<tr>
<td>3.2</td>
<td>POPE bilayer</td>
</tr>
<tr>
<td>3.3</td>
<td>Lipid group interchangeability</td>
</tr>
<tr>
<td>3.4</td>
<td>Preliminary charge derivation for full lipid molecules</td>
</tr>
<tr>
<td>3.5</td>
<td>Preliminary charge derivation for full lipid molecules: head and tails</td>
</tr>
<tr>
<td>3.6</td>
<td>Modular splitting of lipids</td>
</tr>
<tr>
<td>3.7</td>
<td>Representative Lipid11 residues</td>
</tr>
<tr>
<td>3.8</td>
<td>Area per lipid throughout lipid bilayer MD simulation</td>
</tr>
<tr>
<td>3.9</td>
<td>Electron density profile through lipid bilayers from MD simulation</td>
</tr>
<tr>
<td>3.10</td>
<td>Deuterium order parameters for sn-1 and sn-2 tails of phospholipids</td>
</tr>
<tr>
<td>4.1</td>
<td>A mixed Lipid14 bilayer</td>
</tr>
<tr>
<td>4.2</td>
<td>Lipid14 Bilayer</td>
</tr>
<tr>
<td>4.3</td>
<td>Pentadecane molecules with General Amber Force Field (GAFF) parameters</td>
</tr>
<tr>
<td>4.4</td>
<td>Structure and partial charges for Lipid11 and Lipid14 capping groups</td>
</tr>
<tr>
<td>4.5</td>
<td>A capped lauroyl molecule for fitting dihedrals</td>
</tr>
<tr>
<td>4.6</td>
<td>Energy profiles for torsions from a cis-5-decene molecule</td>
</tr>
<tr>
<td>4.7</td>
<td>(^{13}\text{C}) NMR (T_1) relaxation rates for selected alkanes</td>
</tr>
<tr>
<td>4.8</td>
<td>Energy profiles for selected torsions from lauroyl molecule</td>
</tr>
<tr>
<td>4.9</td>
<td>Six primary lipid types simulated in bilayers</td>
</tr>
</tbody>
</table>
Figure 4.10: Lipid bilayer tail order parameters for six lipid bilayer types with comparison to experiment ........................................ 101
Figure 4.11: Total and decomposed electron density for lipid bilayers for six bilayer types ..................................................... 104
Figure 4.12: Simulation X-ray scattering form factors for six lipid bilayer types ............................................................... 106
Figure 4.13: Plot of $\Delta D_{B-H}$ versus area per lipid $A_L$ for all-atom lipid force fields ......................................................... 107
Figure 4.14: Time averaged mean square displacement of the center of mass of the lipid molecules ........................................ 108
Figure 4.15: Lateral diffusion coefficients calculated using different time ranges of mean square displacement ........................................... 108

Figure 5.1: Cholesterol in a DMPC bilayer ........................................ 112
Figure 5.2: Partial charges for cholesterol ...................................... 115
Figure 5.3: Partial atomic charges for decalin .................................. 117
Figure 5.4: Average volume per molecule of cholesterol lipid bilayers ................................................................. 121
Figure 5.5: Average area per molecule of cholesterol lipid bilayers ........ 122
Figure 5.6: Bilayer head to head thickness $D_{HH}$ of cholesterol lipid bilayers ............................................................. 124
Figure 5.7: Deuterium order parameters $S_{CD}$ for DMPC and DOPC tails with cholesterol .............................................................. 125
Figure 5.8: Deuterium order parameters $S_{CD}$ for cholesterol in DMPC bilayer ............................................................. 126
Figure 5.9: Experimental X-ray form factors for cholesterol and DMPC bilayers ............................................................. 128
Figure 5.10: Experimental X-ray and neutron form factors for cholesterol and DOPC bilayers .............................................. 129

Figure 6.1: Cartoon of water permeation through a DOPC bilayer .......... 133
Figure 6.2: Constrained water in lipid bilayer .................................. 147
Figure 6.3: Constrained water in lipid bilayer over time ...................... 148
Figure 6.4: GPU memory architecture ........................................... 150
Figure 6.5: Parallel constraint implementation ................................ 154
Figure 7.1: Extended lipid force field .............................................. 157
Figure 7.2: Self assembly simulation of DOPS with Lipid14 parameters 159
Figure 7.3: Potassium channel structure ......................................... 160
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Lipid11 residue definitions</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Lipid11 atom types from the General Amber Force Field</td>
<td>58</td>
</tr>
<tr>
<td>3.3</td>
<td>Lipid11 atom types from the GLYCAM force field</td>
<td>59</td>
</tr>
<tr>
<td>3.4</td>
<td>Lipid11 molecular dynamics simulations</td>
<td>62</td>
</tr>
<tr>
<td>3.5</td>
<td>Lipid bilayer molecular dynamics structural properties compared with experimental values</td>
<td>68</td>
</tr>
<tr>
<td>4.1</td>
<td>Modification of LJ an Torsion Parameters of Alkane Chains</td>
<td>84</td>
</tr>
<tr>
<td>4.2</td>
<td>Thermodynamic properties of methyl acetate simulated with GAFF/Lipid11 and Lipid14 in comparison with experiment</td>
<td>87</td>
</tr>
<tr>
<td>4.3</td>
<td>Thermodynamic and dynamic properties of a selection of alkane chains simulated using the updated Lipid14 parameters and comparison to experiment</td>
<td>89</td>
</tr>
<tr>
<td>4.4</td>
<td>Thermodynamic properties of a selection of alkene chains simulated using the updated Lipid14 parameters and comparison to experiment</td>
<td>91</td>
</tr>
<tr>
<td>4.5</td>
<td>Average number of trans, gauche, end gauche (eg), double gauche (gg), and gtg’+gtg conformers per alkane molecule and comparison to experiment</td>
<td>92</td>
</tr>
<tr>
<td>4.6</td>
<td>Lipid14 lipid bilayer molecular dynamics</td>
<td>94</td>
</tr>
<tr>
<td>4.7</td>
<td>Average structural properties over five repeats of the six lipid systems simulated with Lipid14 and comparison to experiment</td>
<td>98</td>
</tr>
<tr>
<td>4.8</td>
<td>Analysis of rotamers and rotamer sequences in the acyl chains of the six lipid systems</td>
<td>103</td>
</tr>
<tr>
<td>4.9</td>
<td>Lipid lateral diffusion coefficients calculated from NPT runs, NVE runs, and experimental values</td>
<td>109</td>
</tr>
<tr>
<td>5.1</td>
<td>Decalin enthalpy of vaporization and density</td>
<td>117</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of cholesterol and lipid bilayer simulations at varying cholesterol concentrations</td>
<td>118</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to sincerely thank my advisors and teachers: To Dr. Ross Walker for supporting my work, for being passionate about research, and leading the group as it grows in reputation. It has been a great experience working in the lab. To Dr. J. A. McCammon for supporting my work and being a welcoming part of UCSD. To Dr. Knut Teigen for insightful discussion and support. To Dr. Ian Gould for enlightening discussion and my visit to Imperial College London. I would also like to acknowledge the rest of my committee including Dr. Michael Gilson, Dr. Wei Wang, and Dr. Michael Burkart for their feedback and review of my dissertation work.

I would like to thank several people and organizations who have financially supported my work: To Mark Berger and NVIDIA, Inc. for the NVIDIA Graduate Fellowship which has opened up the world of GPU programming for me. To Dr. Susan Taylor and Dr. J. A. McCammon for the Molecular Biophysics Training Grant fellowship, which enabled my graduate work.

I would like to send my appreciation to the members of my lab and collaborators: they all make the lab excellent. To Romelia Salomon-Ferrer and Andreas Götz for welcoming me into the lab and their help and support. To Robin Betz for talks about science and computers. To Åge Skjevik for excellent discussion in and out of lab. To Callum Dickson for good talks and hosting my visit to London. To all of the lab: Keshav Tadimete, Aditi Munshi, Perri Needham, Longhua Yang, Charlie Lin, and Dan Mermelstein for making the lab great. To the visitors to the lab, especially: Sarah Rosen, Davide Sabbadin, Jane Yin, Murtricia Charles, Chantel Ible, and Maria Ferraro for the contributions to the lab as well.

Chapter 3 includes material from “Lipid11: A modular framework for lipid simulations using Amber” Skjevik, Å. A., Madej, B. D., Walker, R. C., Teigen, K. Journal of Physical Chemistry B. 2012. The dissertation/thesis author was a primary investigator and author of this paper.

Chapter 4 includes material from “Lipid14: The Amber lipid force field” Madej, B. D., Dickson, C. J., Skjevik, Å. A., Betz, R. M., Teigen, K., Gould, I. R., Walker, R. C. Journal of Chemical Theory and Computation 2014 The dissertation/thesis author was a primary investigator and author of this paper.

Chapter 5 is currently being prepared for submission for publication of the material. Madej, B. D., Walker, R. C. The dissertation/thesis author was the primary investigator and author of this material.

Chapter 6 is currently being prepared for submission for publication of the
material. Madej, B. D. Walker, R. C. The dissertation/thesis author was the primary investigator and author of this material.
VITA

2015 Doctor of Philosophy, Chemistry and Biochemistry, University of California, San Diego

2012 Master of Science, Chemistry and Biochemistry, University of California, San Diego

2011-2015 Research Assistant, Chemistry and Biochemistry / San Diego Supercomputer Center, University of California, San Diego

2010-2011 Teaching Assistant, Chemistry and Biochemistry, University of California, San Diego

2010 Bachelor of Science, Biomedical Engineering, Washington University in St. Louis

2009 Intern, Late State Protein Purification, Genentech, Inc., South San Francisco

PUBLICATIONS


“Lipid14: The Amber lipid force field”

“Lipid11: A modular framework for lipid simulations using Amber”

“Structure prediction with the associative memory Hamiltonian.”

**FIELDS OF STUDY**

High performance computing: supercomputers, parallel architectures, and graphics processing units

Molecular dynamics algorithm and code development

Molecular dynamics force field development

Molecular dynamics simulations of cellular membrane systems

**FELLOWSHIPS AND AWARDS**


2013-2014 NVIDIA Graduate Fellowship Finalist. NVIDIA, Inc.

2012-2013 NVIDIA Graduate Fellowship. NVIDIA, Inc.

2010-2011 Urey Award Scholarship. Department of Chemistry and Biochemistry, University of California, San Diego.

2008 Dean’s List. School of Engineering and Applied Sciences, Washington University in St. Louis

2006-2010 Dean’s Honorary Scholarship. School of Engineering and Applied Sciences, Washington University in St. Louis
ABSTRACT OF THE DISSERTATION

Cellular membrane dynamics and algorithms for studying their interactions with pharmaceutical compounds

by

Benjamin D. Madej

Doctor of Philosophy in Chemistry

University of California, San Diego, 2015

Professor J. Andrew McCammon, Co-Chair
Professor Ross C. Walker, Co-Chair

Cellular membranes are incredibly complex structures composed of diverse biomolecules including lipids, proteins, and other molecules. Cellular membranes are important because they allow for the transfer of molecules and chemical signals in and out of cells. The structure of membranes and membrane proteins is difficult to determine through experiments. Membrane structure is also highly dynamic and depends on the mixture of molecular components. While the membrane is usually in the liquid-disordered phase, other local membrane assemblies have been observed. It
is challenging to predict the permeation of pharmaceutical compounds through the membrane.

Molecular dynamics (MD) is a computational method that allows for the study of membrane motions and drug interactions. Based on a chemical model of molecules and Newton’s equations of motion, it is possible to predict the dynamics of molecules on a computer. However, in order to simulate new molecules, it is necessary to refine an appropriate force field that models the chemical interactions of the molecules.

A new force field was developed for lipids, an essential membrane component in the Amber MD software package. This force field was parameterized with experimental data and quantum mechanical calculations on individual chemical components of the lipid molecule. Afterwards, parameters were validated against available membrane structural data. Parameters have been developed for a set of glycerophospholipids and cholesterol. Molecular dynamics simulations of lipid membranes with the new parameters have accurately predicted membrane structural properties.

With an accurate model of lipid membranes, it is now possible to examine complex membrane dynamics. Permeation of small molecules across the membrane is especially interesting in the pharmaceutical industry. Using the inhomogenous solubility-diffusion model it is possible to predict small-molecule permeability across a membrane from potential of mean force calculations. A constrained molecular dynamics algorithm was implemented in Amber for this task. The constraint implementation may be optimized to run on graphics processing units (GPU).

This dissertation marks the first expanded lipid force field in Amber for accurate membrane simulations. It also marks the implementation of a accelerated general constraint methods in Amber.
Chapter 1

Introduction

1.1 Guide to this dissertation

This thesis includes the research and development of new computational methods and theoretical models and methods for the study of lipid bilayers and cellular membranes. For the first time in the Amber molecular dynamics program it is possible to simulate complex membranes at atomic detail that reproduce membrane structural properties.\textsuperscript{1,2,3,4} Additionally, constraint methods for molecular dynamics have now been implemented in Amber that allow for the study of molecular diffusion and permeation across the membrane. In sum, this thesis examines and implements the components necessary to study membranes and membrane bound proteins in Amber. It details the development of the first modular lipid force field to support tensionless NPT simulations that give correct lipid bilayer behavior.

Due to the interdisciplinary nature of this work, there are several topics to be covered, and this thesis is thus organized in the following manner:

\begin{itemize}
  \item Chapter 1 introduces the underlying biophysical structure of membranes, as
well as reviews new hypotheses and theories for the structure of membranes. A major problem in the pharmaceutical industry is understanding how small molecules permeate across membranes. Furthermore, membrane dynamics have been studied with computational and theoretical approaches. The main hypotheses and aims of this work are defined at the end of this chapter.

- Chapter 2 briefly reviews the underlying theory, algorithms, and implementation of membrane molecular dynamics simulations. This includes a description of the challenges involved in force field development.

- Chapter 3 describes the development of the first ever modular framework for phospholipid force fields called Lipid11. After preliminary charge derivations, a modular framework was found to be feasible for high fidelity lipid bilayer simulations. This framework established the platform on which lipid parameters would be refined.

- Chapter 4 summarizes the parameterization of a robust molecular dynamics force field for phospholipids. Parameters for specific portions of the phospholipid were optimized in model molecules to fit quantum mechanical and experimental data. The full lipid parameter set was validated with MD simulations of six pure lipid bilayer types compared to structural bilayer data.

- Chapter 5 investigates the parameters for another crucial membrane component: cholesterol. A minimal parameterization for cholesterol was initially developed. Cholesterol in a range of concentrations in lipid bilayers was then simulated with this minimal parameterization set.

- Chapter 6 describes constraint molecular dynamics methods within the Amber software suite. Constrained molecular dynamics may be applied to potential of mean force calculations used in estimating the permeability and diffusion of
small molecules through membranes. This type of constraint was implemented in Amber, and strategies for optimizing performance with graphics processing units (GPUs) were explored.

- Chapter 7 summarizes the major results of this work, and the implications of these developments. Several other projects have stemmed from this dissertation work and are listed in this chapter. Future research directions are also discussed.

### 1.2 Cellular membranes

Cellular membranes serve as a stage for numerous molecular interactions and dynamics that control cellular physiology. The cell membrane hosts many different classes of biomolecules with diverse roles. For a long time, proteins have received the most attention as the primary executor of membrane functions. But all facets of membrane structure and function cannot be explained by membrane proteins alone. Other biomolecules in and near the membrane are crucial for a variety of cellular processes.

Membranes are generally composed of a diverse mix of glycerophospholipids, glycolipids, sterols, sphingomyelin, peptides, and proteins.Remarkably, cellular membranes not only separate and compartmentalize cellular environments, but also play a key role in cellular physiology. Cellular signalling pathways begin at the cell membrane and leads to a signal transduction cascade with a variety of cellular effects.\(^5\) Cellular constituents enter and leave through the membrane itself or through assistance with specific molecular channels.\(^6,7\) In certain cell types, the ionic concentration gradient across a membrane establishes a membrane potential important for cellular function.\(^5\) As a whole, membranes are dynamic, complex assemblies that execute a wide range of cellular functions. All of this occurs within or near a molecular structure.
Figure 1.1: Glycerophospholipid structure. Constituent chemical groups are listed in the figure.

approximately 30 Å thick.

1.2.1 Membrane developments

Originally presented in 1972 by Singer and Nicolson, the “fluid mosaic model” has been presented as a popular model of membrane structure. Lipids and proteins are mixed in a two-dimensional viscous fluid. The fluid mosaic model was developed from thermodynamic and structural data that indicated that membrane proteins and lipids diffused laterally across the surface of the membrane. The lateral diffusion of proteins across the membrane was observed to be stochastic in nature. In the fluid mosaic model, the membrane could be visualized as a dynamic mosaic, in which membrane patches were constantly moving. While this model has evolved, some of its original principles are still relevant.

On a basic level, lipids are the main constituent of membranes through which the other molecular components are distributed. The basic phospholipid structure is shown in figure 2.5. Other common lipid components are shown in figure 1.2. In membranes, lipids are generally found in a bilayer structure with the polar headgroup
Figure 1.2: Common membrane components. The chemical structure of palmitoyl-oleoyl-phosphatidylcholine (POPC), sphingomyelin, and cholesterol is shown.

oriented towards the cytosolic and extracellular solution and the tails aggregated in the center of the bilayer. The lipid tails are generally complex mixtures of saturated or unsaturated acyl chains with varying lengths. The lipid tails are generally found in the nonpolar interior of a membrane. Tails may be found in various configurations and orientations. Under certain conditions, some tails are found to be highly ordered and organized, while other tails are inherently disordered. Tail order also depends on the level of tail saturation, with carbon-carbon double bonds enforcing kinks within the tails which leads to increased disorder.

Membranes have often been modeled with pure lipid bilayer systems. With various approaches, it is possible to examine pure lipid bilayers in vesicles or other constructs such as lipid disks. However, the precise molecular structure of pure lipid bilayers is not known. Technological advances in experimental methods such as X-ray and neutron scattering experiments or $^2$H and $^{13}$C NMR have provided detailed structural information on lipid bilayers. Pure lipid bilayers by themselves exhibit complex behavior depending on the type of lipids included, the phase of the bilayer, as well as any additional bilayer components. In biological systems, most lipid bilayers are found in the fluid phase, but do exhibit other phases under other
conditions. The so-called gel phase is observed at lower temperatures when the lipid tails become highly ordered and lateral diffusion slows. On a larger scale, bilayers also exhibit macroscopic undulations and rippling across the bilayer. Water concentration also may influence the phase of the membrane. Lipid macromolecular structures such as micelles or vesicles may form under different water concentrations.

In vivo membranes tend to be very heterogeneous in nature. Different membranes within the cell contain different quantities of lipids, sterols, and proteins. More detailed descriptions of localized membrane components are becoming available through lipidomics. The diversity and specificity of membrane components in different membranes may be responsible for functional differentiation across membranes.

With detailed structural experiments, the model of membranes and lipid bilayers has become clearer.

1.2.2 Membrane rafts

A major recent development in the study of membrane structure has come from observations of so-called membrane “rafts.” Rafts were hypothesized to be composed of sphingolipids, cholesterol, and proteins that assemble for functional purposes. Furthermore, it was hypothesized that rafts were similar in structure to lipid bilayers in the liquid ordered phase. The liquid ordered phase is a special phase in lipid bilayers usually observed in the presence of cholesterol and sphingomyelin, characterized by increased order in the lipid tails and accompanied by rapid translation and lateral diffusion of the lipids. The increased packing and reorganization of the membrane around the protein could affect the protein conformation.

The membrane raft theory of membranes has been a highly controversial one. Limitations and incongruities in detergent solubilization and cyclodextrin treatment experiments have led to some debate over whether the raft phase is an actual
phenomenon in membranes.\textsuperscript{22,23} Some experimental results are conflicting and conflated, or show only indirect evidence of membrane rafts.\textsuperscript{24} However, after lengthy re-examination and review of experimental approaches, there is still evidence of membrane raft-like assemblies. Some of the advances in membrane raft experiments are described here:

There have been several advances in membrane raft experiments. First, advances in microscopy techniques such as stimulated emission depletion (STED) microscopy have improved microscopic observations of cellular membranes.\textsuperscript{25} With high resolution microscopy techniques, very small raft assemblies have been observed within the membrane itself. The raft assemblies on the membranes are highly dynamic, and many of the assemblies may cluster into raft platforms. Timescales of formation and disassembly vary by lipid type and protein components. What has still not been seen through microscopy is large (on the scale of hundreds of nanometers) domains of the membrane in the raft phase.\textsuperscript{26}

Secondly, recent significant advances in lipidomics and mass spectrometry have allowed for rapid, precise characterization of membrane components.\textsuperscript{18} Previously, membranes were described by only several major classes of molecular components. Now, it is possible to isolate the diversity of lipids and other components within a membrane. For example, it is possible to quantify not only the fraction of phosphatidylcholine lipids, but also the length and saturation of the phospholipid tails.\textsuperscript{27} Given the countless types of cells and cell membranes, this is a powerful technique to identify membrane components. In particular, raft phase components like cholesterol, sphingolipids, and proteins have also been identified in certain membranes.

Finally, while model membranes have been widely used to investigate the lipid raft phase, there has been much debate about the implications of these results. Model membrane studies are based on methods which isolate simple lipid bilayers composed
of raft lipids, cholesterol, and sphingomyelin. Model membranes with an appropriate mixture of components can enter the liquid ordered phase across the membrane. However, because evidence for large raft phase domains in actual membranes has still not been found, the transferability of such models has been called into question.

Regardless of whether the raft phase will be found in real cell membranes, there is already evidence of small, dynamic raft assemblies and platforms. This implies that the lipid environment around proteins is indeed dynamic and changing, and may contribute to membrane protein functionality. However, further investigation of the raft assemblies and membrane proteins will be necessary.

### 1.3 Membrane proteins

Membrane-bound proteins are an important component of cellular membranes with a variety of roles: structure, recognition, signaling, and transport. Some proteins are embedded in the membrane with a transmembrane domain (integral), while other proteins are only adjacent or transiently associated with the membrane (peripheral). It is known that many of these proteins regulate and implement many key cellular processes. Transmembrane proteins usually include regular secondary structure features like α-helices that span the membrane. The transmembrane portion of proteins often have specific orientations within the membrane.

G-protein coupled receptors (GPCR) are one class of essential membrane proteins. GPCRs are proteins with seven transmembrane α helices that span the membrane and may be oligomeric. Rhodopsin was one of the first GPCR crystal structures to be resolved. The structure of Rhodopsin with a membrane model around it is rendered in figure 1.3. Found in rod cells within the eye, rhodopsin is a type of GPCR that triggers and amplifies a signal cascade within the cell. More crystal structures of GPCRs have been resolved in membrane environments. As a whole, GPCRs
Figure 1.3: Rhodopsin protein structure in a membrane. PDB identification number: 1U19. Panel A shows a side view and panel B shows the top-down view. Lipids were packed around the protein. POPC and POPE are drawn in grey lines, cholesterol is marked in green spheres, and solvent is rendered in blue spheres.
represent one of the largest class of modern drug targets today.\textsuperscript{32}

Ion channels are crucial functional components of the membrane with incredible diversity and specificity.\textsuperscript{5} These receptors control the flux of ions across membranes. It is well known that the ionic concentration gradient across the membrane establishes a membrane potential. This membrane potential plays a crucial role in intracellular signaling in certain types of cells including cardiac myocytes and neurons.\textsuperscript{33} The crystal structure of the proteins has been resolved for several families of ion transporters. KCSA, gramicidin, and other structures are now known.\textsuperscript{7,6}

Some proteins are thought to be regulated by lipids as well. The Na\textsuperscript{+} and K\textsuperscript{-} ATPases are hypothesized to be regulated by cholesterol, a key membrane component.\textsuperscript{34} Ca\textsuperscript{2+} ATPases are hypothesized to be regulated by phosphatidylethanolamines in the sarcoplasmic reticulum.\textsuperscript{34} Other proteins are known to have lipids associate and surround the transmembrane portions of their structure in stable configurations.\textsuperscript{35} While the exact mechanism of interaction is still not fully understood for many of these systems, there is growing evidence that membrane components have regulatory or functional roles.

Even with these advances in studies of membrane proteins, it is still relatively difficult to determine membrane bound protein structures. The membrane environment is notoriously difficult for crystallographic and NMR structural studies. Crystallography is possible with special detergents, but it is not always certain this reproduces the membrane environment.\textsuperscript{36} Solid-state NMR is another approach for structural studies of membrane proteins.\textsuperscript{37} However, it is orders of magnitude more difficult to resolve the structure of membrane proteins than globular proteins. While the number of published protein structures in solution have increased exponentially, membrane protein structures have increased at a much slower rate.\textsuperscript{38}
1.3.1 Protein lipid modifications

Peripheral membrane proteins are an interesting class of proteins in that they may transiently associate with membranes. A subset of peripheral proteins actually contain lipid modifications to their structure. Common modifications include N-myristoylation, palmitoylation, and prenylation to the ends of the protein chain. Other proteins may include modification with glycosphingolipids and related moieties.

Remarkably, the lipid modifications may influence bilayer interactions and act as an anchor for membrane association. Some examples of lipid modified proteins include several serine/threonine kinases, tyrosine kinases, and other proteins in signal transduction cascades. Additionally, the Src family of protein kinases and the α subunit of heterotrimeric G proteins are both acylated and associate with membranes. The catalytic subunit of protein kinase A is N-myristoylated and thought to anchor to membranes.

It is hypothesized that acyl modifications may be significant in the localization of proteins to specific membrane types and locations. This would be an elegant mechanism, but still has not been conclusively shown. It is possible that lipid modifications may provide other functionality to the protein but the full range of modifications is still not fully understood.

1.4 Membrane transport and permeability

Given that the structure of membranes is heterogeneous, complex, and dynamic, an even more complicated phenomenon is molecular transport across the membrane. The membrane allows some molecules to pass through passively, and some molecules to pass through actively through various molecular mechanisms. Clever ex-
Experimental methods have isolated individual transporters in cells and have measured
the flux of molecules such as ions through the membrane. Other assays measure the
molecular partitioning in cells through the membrane.

However, even with these advanced approaches, it is difficult to explain the
exact mechanisms of passage through the membrane. It is even more difficult to
predict \textit{a priori} the molecular features that either allow or block molecular passage
through the membrane.

This is a major problem for pharmaceutical and biotech companies that de-
velop small molecule drugs with targets located in the interior of the cell. Without
a detailed model of membrane structure, it is extremely difficult to determine which
small molecules might cross the membrane. Furthermore, it would be desirable to
determine molecular features that facilitate transport across the membrane.

It has been hypothesized that lipid structural properties may influence perme-
ability across a membrane. However, there are weak correlations between the lipid
physical properties and molecular permeation. It is not understood why certain
small molecules pass through membranes.

\section{1.5 Computational and theoretical studies of membranes}

Cell membrane structures have long been studied through experimental ap-
proaches. However, with the rise of theoretical and computational chemistry, there
are now alternative methods to study membrane structure. Computational methods
seek to complement traditional experiments by building physical models of chemical
systems and predict the structure and dynamics of such systems. A variety of compu-
tational methods have been applied to lipids to simulate the structures and dynamics
of the molecules.

### 1.5.1 Membrane molecular dynamics

One widely used method is molecular dynamics (MD) and predicts molecular motions by integrating Newton’s equations of motion. The details of the molecular dynamics method are described in chapter 2. MD has been applied to many condensed-phase biological systems in order to study the structure and dynamics of biomolecules.

Molecular dynamics has been previously applied to lipid bilayer and membrane systems. Furthermore, there are already force fields for proteins, lipids, and other membrane components. However, many previous force fields suffered from issues that affected the accuracy of lipid membrane simulations.

One major molecular dynamics software suite is called Amber. Amber originally stood for Assisted Model Building and Energy Refinement. Amber includes software to simulate molecules with MD, as well as a set of force fields for condensed-phase biomolecular simulations. In 2010, Amber MD did not include a robust parameter set for lipid molecules. Furthermore, it did not include parameters for some other membrane components such as cholesterol and sphingomyelin.

There are many unanswered questions regarding membrane structure and dynamics, and molecular dynamics simulations were proposed to investigate several important questions about membranes. The following list defines the specific areas of focus for this dissertation, as well as the hypotheses explored:

- At the beginning of this work, given the lack of an accurate force field for lipids in Amber, a new lipid force field for simulation with the other Amber force fields was proposed. As the actual phospholipids for real cellular membranes are incredibly diverse, a modular approach for a lipid force field was developed
and tested. This would allow for any combination of lipid components. This was the first version of the force field developed but required constant surface tension.

- However, as real bilayers and membranes do not experience constant surface tension, it was hypothesized that the parameters could be refit to reproduce experimental bilayer properties in a tensionless ensemble. The initial parameters of the lipid force field were fit with model molecules representing portions of lipids. The final parameters were tested against a set of common bilayer types.

- Furthermore, given the importance of cholesterol in real membranes, cholesterol was parameterized for molecular dynamic. Cholesterol is found in liquid-ordered bilayer and possible raft assemblies, and was tested in bilayer simulations. A range of cholesterol concentrations was tested in lipid bilayers for eventual use in full membrane simulations.

- Finally, it was hypothesized that small molecule permeation depends significantly on membrane-ligand interactions. However, to test this hypothesis, constrained molecular dynamics methods were necessary. As Amber lacked this type of general constraint, it was implemented in its primary MD program. Constraint methods significantly impact MD performance, so performance optimizations were explored.

In sum, this work allows for accurate, fast lipid bilayer and membrane simulations in the Amber molecular dynamics suite. The simulation of liquid-ordered bilayers and their application in membrane raft assemblies was of particular interest in the parameterization approaches. An optimized method for the comparison of molecular permeability is also presented.
Chapter 2

Molecular dynamics

This chapter provides a summary of the underlying theory and algorithms that define classical molecular dynamics. Emphasis is placed on simulations of condensed-phase biological systems that will be described in detail in the following chapters.

While the theories and algorithms presented here are general and may be implemented in any molecular dynamics software package, emphasis is placed on methods currently implemented in the Amber molecular dynamics package, as a majority of the following work used Amber MD for development and simulation.\textsuperscript{1,55,56,57,58}

2.1 Quantum mechanics and Newtonian dynamics

Quantum mechanics calculations have provided a robust method for examining electronic molecular structures. There are various numerical approaches to solve the Schrödinger equation for molecular systems. However, the computational costs of such approaches are relatively high, scaling \( n^4 \) with the number of atoms \( n \) in the system. Given the high cost for evaluating the wavefunction of a molecule in a single
configuration, it is relatively expensive to predict the motions of large molecules (hundreds of atoms) with QM methods.

Thus, other less expensive approaches to calculating the motions of atomic nuclei and electrons have been used.\textsuperscript{59} Some current approaches rely on models including energy functions and parameters that approximate QM calculations for certain types of molecules. A molecular force field may be defined as an energy function representing the potential energy of molecular configurations along with parameters specific to certain molecular interactions. Parameters are fit for specific chemical features of the molecules.

One theoretical approximation is very important in the formulation of force-field approaches: the Born-Oppenheimer approximation. The approximation (originally described in Born and Oppenheimer’s 1927 paper) states that the wavefunction of a molecule may be decomposed into an electronic and nuclear component.\textsuperscript{60} This approximation allows for the Schrödinger equation to be solved as a combination of separate electronic and nuclear wavefunctions. In practical terms, this approximation is acceptable for quantum calculations largely because of the difference between nuclear and electronic masses in atoms.

Force-field methods do not explicitly include electronic contributions, but rather approximate atoms as points. The electron density of the molecule is reduced to points centered on the nucleus of the atom. The molecular dynamics model is structured around the motions of these points which represent the atoms of the molecule. Thus, all electronic effects are not explicitly calculated in a simulation.

Through such approximations and models, significant performance gains may be obtained in calculating molecular energies and forces.\textsuperscript{56,57} Parameterization of force fields attempts to reproduce quantum energies and forces through empirical calculations. However, these optimizations allow for large macromolecular systems including
proteins and nucleic acids with hundreds of thousands to millions of atoms.\textsuperscript{56,57}

Remarkably, with the Born-Oppenheimer approximation, molecular motions may be predicted with classical Newtonian dynamics of atoms. While it is difficult to solve $n$-body problems beyond $n = 3$ analytically, there are a variety of numerical approaches to solve Newton’s equations of motion for the larger systems seen in molecular dynamics.\textsuperscript{59} By coupling a molecular force field with integrations of the basic Newtonian equations of motion, it is possible to predict molecular positions and velocities over time. This method is called molecular dynamics (MD) and has been applied to numerous large biomolecular systems. MD is inherently iterative and depends on all previous steps to predict a position at time $t$.

At first inspection, it is difficult to imagine that Newtonian dynamics could possibly reproduce nuclear motions of molecules. One theoretical rationale for molecular dynamics comes from the equation for de Broglie thermal wavelength $\Lambda$ for an ideal gas in the nonrelativistic case:\textsuperscript{59}

\begin{align}
\Lambda &= \frac{h}{p} \quad (2.1a) \\
\Lambda &= \sqrt{\frac{2\pi h^2}{mk_B T}} \quad (2.1b)
\end{align}

where $E_K$ is the kinetic energy, $T$ is the temperature, and $m$ is the atomic mass. $h$ is Planck’s constant and $k_B$ is the Boltzmann constant.\textsuperscript{59} From classical statistical mechanics, if $\Lambda << \alpha$, where $\alpha$ is the average distance between atoms, the system will obey Maxwell-Boltzmann statistics. Therefore, for many condensed-phase biological systems at physiological temperatures, the classical model may obey Maxwell-Boltzmann statistics because $\alpha$ is less than the de Broglie thermal wavelength.

One major issue with this approach is that explicit electronic effects are not
included. Therefore systems with excited states or other electronic effects may not be simulated with only molecular mechanics methods.

For certain systems, this deficiency may be addressed by dividing the molecular system spatially into regions treated with molecular mechanics and regions treated with QM. These hybrid QM/MM approaches combine the performance of MD with the detail of QM for specific regions of chemical interest.\textsuperscript{61}

\section{2.2 Molecular mechanics force fields}

As previously described, Newtonian dynamics relies on an energy function with chemically specific parameters to evaluate the potential energy of molecular structures. In principle, this energy function can take any form that describes the underlying quantum chemistry; however, the functional form has been derived from parameterization of specific terms designed to model molecular interactions. Indeed, major MD software includes force fields with subtly different functional forms. Parameters are generally fit for specific classes of molecules. Often, parameters are derived from quantum calculations or empirically from experimental data. As such, parameter sets are not unique: for example, many different force fields exist for proteins yet are able to predict similar structural dynamics.

It is important to note that each force field employs a unique parameterization strategy that influences molecular dynamics. Therefore, each decision made in parameterization and approximation is information that manifests itself in the parameters that define the final MD simulation. In the following section, the Amber energy function is presented as a model with discussion of its functional form and parameterization.

Most parameters (but not all) have an underlying physical basis used in parameterization which is also discussed in the following sections. Terms and parameters
may be essentially divided between bonded and non-bonded interactions. Several of the widely-used Amber pair-wise additive force fields such as its latest generation of protein and nucleic acid force field *ff14SB*, General Amber Force Field *GAFF*, carbohydrate force field *Glycam*, and lipid force field *Lipid14* use a potential of the form.\textsuperscript{62,63,64,65,4}

\[ E_{\text{Amber}} = \sum_{i}^{n_{\text{bonds}}} b_i (r_i - r_{i,eq})^2 + \sum_{i}^{n_{\text{angles}}} a_i (\Theta_i - \Theta_{i,eq})^2 + \sum_{i}^{n_{\text{dihedral}}} \sum_{i}^{n_{i,\text{max}}} (V_{i,n}/2)[1 - \cos(n\phi_i - \gamma_{i,n})] + \sum_{i<j}^{n_{\text{atoms}}} \left( \frac{A_{ij}}{r_{12}^{ij}} - \frac{B_{ij}}{r_{6}^{ij}} \right) + \sum_{i<j}^{n_{\text{atoms}}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \]  \tag{2.2}

All terms and symbols will be discussed in detail in the following sections.

In general, pairwise additive force fields are designed such that each individual energy contribution may be calculated independently and then summed to calculate the total potential energy. In principle, the Amber force fields are designed to be combined in mixed molecular simulations.\textsuperscript{55} In the Amber force fields, each term of the energy function includes parameters that depend on the combinations of atom types as discussed in the following sections.

### 2.3 Parameterization

As seen in the previous section, each term in the energy function has several parameters that define the shape and magnitude of the associated potentials. Refinement of physically representative values for these functions is known as parameterization. To be clear, there is not a single correct way to develop parameters for molecular simulation.
Parameterization methods usually attempt to isolate related parameters and fit individual terms of the energy function. For each parameter, basic parameterization strategies will be discussed within the following sections. Decisions made at each stage of parameterization have subtle effects on the final parameter set and corresponding MD simulations. It is valuable to understand the errors for each term as well as the relative errors between energy terms.

A fundamental issue with molecular models is the balance between the fit of parameters and the predictive power of the model. While it is possible to perfectly fit parameters to all available theoretical and experimental data for molecule, there is a risk of overfitting parameters. Overfit parameters may lose their transferability in other conformations or chemical environments. Further, overfit parameters are not reusable for other systems, and are likely not able to predict dynamics for unknown systems.

Another fundamental issue is determining how to evaluate the correct behavior of the biomolecules. Parameter quality depends on criteria for each system: usually quantum or experimental data. One of the most important comparisons is through validation with experimental data.\textsuperscript{3,2,4} Ultimately, MD models are intended to complement experimental results through modeling systems, and explore and predict new chemical dynamics.

### 2.4 Parameters

The following sections examine the major energy terms for a molecular system. The functional form for each individual term is described and the related parameters are explained.
2.4.1 Atom types

The concept of an atom type arises from the fact that parameters are differentiated for atoms within different chemical contexts. With static parameter sets, it is difficult to capture molecular features with just one atom type per element. For example in the General Amber Force Field, there are several different types of carbon. A common carbon atom type is the $sp^3$ hybridized tetrahedral carbon. However, there are also $sp^2$ hybridized planar carbons for carbonyl groups. Several other variations of carbon exist for specific chemical cases.

The energy terms of the Amber energy function have parameters that depend on the atom types involved in that term. For example, each pair of bonded atoms have parameters for their combination of atom types, while each angle term depends on three sequentially bonded atom types.

Therefore, for each simulated molecule, atom types must be defined that correspond to the desired parameter set for MD. There are automated approaches to assigning atom types based on their bonding and chemical environment. However, new parameter sets must manually define new and unique atom types.

2.4.2 Bonds

For condensed-phase biomolecular systems, the most basic energy function term is the bond term between atoms. Within the Amber force fields, the bonded term is defined by a potential described by Hooke’s law.

$$E_{bond} = b_i (r_{ij} - r_{ij,eq})^2$$

(2.3)

$b_i$ is defined as the force constant and $r_{ij,eq}$ is the equilibrium distance between atoms. Therefore, this potential behaves as a harmonic oscillator around an equilibrium dis-
Figure 2.1: Bond potentials for bond stretching. Based on GAFF parameters for the c3-c3 (sp3) bond with a harmonic force constant of 303.1 kcal/(mol Å²). The minimum of the well depth is shifted to the origin; in the real parameter set, the minimum of the well depth is at 1.535 Å. In general the force constant $b_i$ is large, meaning that the energetic barriers beyond the minimum are high. This potential attempts to recreate the potential exhibited between bonded diatomic systems. However, the bond is defined by atom type, not element. Therefore, different types of bonds are represented with different atom types. Atom types are discussed in more detail in section 2.4.1.

There are several issues with such a potential form. The first is that as the distance between two atoms increases far beyond the equilibrium distance, the potential increases rapidly and approaches infinity. The problem with the Hooke potential is that it represents an infinite energetic barrier at distances far from equilibrium, which is not physically realistic. Additionally, because of this potential, bonds cannot be broken as the interaction distance increases from the equilibrium value. Therefore, during a plain MD simulation, bonds are defined by the initial molecular topology and bonds will not form or break between atoms.

The Hooke potential is derived from a Taylor expansion of the potential energy about an equilibrium distance of $r_{i,eq}$. Beginning with the second order Taylor
expansion between two atoms, $i$ and $j$

$$E_{bond} = E_{ij}(0) + \frac{dE}{dr}(r_{ij} - r_{ij,eq}) + \frac{1}{2} \frac{d^2E}{dr^2}(r_{ij} - r_{ij,eq})^2$$  \hspace{1cm} (2.4)

When the distance between atoms is near equilibrium, $\frac{dE}{dr} \rightarrow 0$ and will be small relative to the third term. Thus, while the bond distance is near equilibrium, the third term is the main contribution to the bond energy term in the form of Hookes law.

Higher order Taylor expansions could be applied towards the bond potential.\textsuperscript{59} These polynomials are more likely to capture the actual asymmetry of the potential. However, they suffer from the same issue as the Hooke potential in that they rapidly diverge away from the equilibrium and approach infinity at their limit.

A more realistic potential is the Morse potential:

$$E_{bond} = D(1 - e^{-\alpha(r_{ij} - r_{ij,eq})})^2$$  \hspace{1cm} (2.5)

$D$ and $\alpha$ are parameters that determine the well depth and size of the potential, respectively. This potential can be parameterized such that as the distance increases well beyond equilibrium it approaches the dissociation energy of the two atoms. During MD, unless the molecular structure is highly strained, it is unlikely for molecular bonds to be in the range beyond equilibrium. Further, the exponential function is a relatively expensive computation compared to the multiplication operation of the Hooke potential. For these reasons, it is more common to find the simple Hooke potential for the stretching of bonds between atoms in MD codes.
Figure 2.2: Dihedral potential for dihedral rotation. The dihedral potential of GAFF c3-c3-c3-c3 is plotted. The total dihedral is the black solid line, which is the sum of the other dashed lines terms with periodicity $n = 1, 2, 3$.

2.4.3 Angles

Three atoms bonded in sequence form an angle $\Theta$ which forms the basis of the next term in the Amber energy function. This interaction is also modeled with a Hooke potential defined as

$$E_{\text{angle}} = a_i (\Theta_i - \Theta_{i,eq})^2$$

Instead of an equilibrium distance, an equilibrium angle $\Theta_{i,eq}$ is parameterized for this potential. $a_i$ corresponds to an angular force constant. For Amber force fields, angular potentials generally have weaker potentials than bond distances. Figure 2.1 shows the Hooke potential; for angles, the angular force constant would create a shallower minimum. In general, Amber parameters allow flexibility of angles of $\pm 30^\circ$ from the equilibrium angle at 300 K. These parameters correspond to the geometries observed for different chemical groups. Again, atom types differentiate angle parameters and allow for multiple angle types.
2.4.4 Dihedrals

During molecular simulations with early force fields, it was hypothesized that force fields consisting of bond and angle parameters in combination with electrostatic and van der Waals terms could capture the subtle dynamics and ternary structures of proteins. Observations from simulations without dihedral terms showed it was difficult to capture this behavior with just these terms. To address this, an additional energy term was introduced for dihedrals to represent the torsional rotations around bonds. While this does not correspond directly to an actual chemical property, it does represent subtle quantum effects that are difficult to describe with just the other energy terms.

Dihedrals are defined by a sequence of four atoms bonded in order, with rotation around the central two atoms. This term is inherently periodic as torsions rotate around a bond. This type of potential can be represented with a Fourier series expansion of the periodic energy barrier. In general, torsions are described by the following sum:

\[
E_{\text{dihedral}} = \sum_i \sum_n (V_{i,n}/2)[1 + \cos(n\phi_i - \gamma_{i,n})]
\]  

(2.7)

The term \( V_{i,n} \) is the barrier height for the torsion, \( n \) is the periodicity, and \( \gamma_{i,n} \) is the phase of the torsion potential. In general, Amber dihedrals for nucleic acids and proteins are represented with either one to four different periodicity terms for the torsion potential. The sum of these multiple terms makes up the total dihedral potential. Additionally, due to the large number of potential atom type combinations in dihedrals, sometimes dihedrals are re-used for the same two central atom types.

Figure 2.2 shows the dihedral potential of the c3-c3-c3-c3 dihedral from the GAFF. The total potential is the sum of the three periodicity terms \( n = 1, 2, 3 \).
Clearly, the introduction of additional terms allow for complex periodic potentials and adjustment of the relative barrier height and phase allow for complex dihedral profiles.

Dihedrals are difficult to parameterize because there is no experimental data isolated for torsional rotations. The most common approach to parameterize dihedral parameters is through torsional scans about the dihedral coupled with geometry optimization during rotation. The energy of each structure is then evaluated through quantum mechanical calculations. The dihedral energy from the energy profile defined by equation 2.7 may then be fit to the quantum energy profile.

The Amber program `Paramfit` provides functionality to automatically fit bonds, angles, and dihedrals data to quantum mechanics calculations. Both a simplex and genetic algorithm are provided for fast parameter fitting. Multiple parameters and multiple structure may be fit simultaneously. Bonded terms in this dissertation were fit with paramfit.

Note that dihedral terms are coupled with the 1-4 electrostatics and 1-4 van der Waals terms. The first and fourth atom of each dihedral experience the additional non-bonded forces as described in sections 2.5.1 and 2.5.3. However, these terms are scaled by an empirical factor. For the Amber force fields, a common scaling is division by 1.2 for electrostatics and division by 2.0 for van der Waals interactions.

### 2.4.5 Improper dihedrals

Improper dihedrals are a special additional term that define the energy of the angle between four atoms that are not linearly bonded. The most common application of an improper dihedral term is to maintain the planarity of four atoms. For example, a $sp^2$ hybridized carbonyl group could include an improper dihedral to maintain the planarity of the three atoms connected to the central carbon. In the Amber force
fields, generally, these improper terms are small contributions to overall molecular geometry; however they may be important where normal dihedrals are not adequate to describe the geometry of certain chemical groups. Improper dihedrals have the same functional form as normal dihedrals (equation 2.7). The angle is defined by the four non-sequentially bonded atoms.

2.4.6 Cross terms

This chapter has focused on primarily independent energy contributions to the Amber energy function. This relies on the approximation that the molecular potentials and parameters are written in independent terms. This may not always be an appropriate approximation, as some degrees of freedom are correlated. For example, rotations of molecular groups around a bond are often coupled with a corresponding stretch of that bond. This type of cross-term interaction is not explicitly contained in the force field. Depending on the parameterization strategy, these cross terms could be refined into the parameters.

Other force fields include some cross-terms. One notable example is the Charmm CMAP cross-term for \( \phi, \psi \) dihedrals in the backbone of protein chains. This is included to fit backbone conformations of proteins. However, any additional energy terms requires additional code implemented in the MD. CMAP terms have been included in Amber and in other MD programs that simulate Charmm force fields. These terms are not used in the Amber force fields.

2.5 Non-bonded interactions

In addition to the bonded and dihedral terms that have previously been explained, additional terms represent intermolecular interactions in the energy function.
These can broadly be divided into two major categories: the electrostatics and van der Waals forces. This class of interactions is particularly computationally expensive due to the long-range nature of these interactions, meaning in principle that there are \( n^2 \) interactions that need to be calculated between every pair of atoms.

### 2.5.1 Electrostatics

Electrostatics interactions model electron density effects on molecules. Charges within nonpolarizable force fields may be approximated in several ways. One way to model the molecular electrostatics is through point charges. A simple model for this is to apply an average pairwise-additive partial charge to the nucleus of the atom.\(^{59}\)

The energy contribution of each point charge may then be calculated with the classic Coulomb potential

\[
E_{\text{electrostatic}} = \sum_{i<j}^{n_{\text{atoms}}} \frac{q_i q_j}{4 \pi \epsilon_0 \epsilon_r r_{ij}}
\]  

(2.8)

where \( n \) is the number of atoms, \( q_i \) and \( q_j \) are the partial atomic charges, \( \epsilon_0 \) is the permittivity in a vacuum, and \( \epsilon_r \) is the relative permittivity. The prime symbol \( \prime \) indicates that electrostatics contributions on atom \( i \) only come from atoms \( j \) which are not part of bond or angle interactions with atom \( i \).

It is not possible to obtain partial charges directly from experiments. Therefore partial charges must parameterized from some other source. Charges are usually obtained from QM electrostatic potential calculations that estimate the approximate electric field of the molecule.\(^{71}\)

Charges may also be modeled as a multipole expansion of the charge distribution. In this way, the electrostatic energy is modeled from the truncation of the charge and multipole series expansion. This approach is useful when the spherical Coulomb potential is inadequate. For example, the multipole expansion may capture
direction-dependent behavior of the charge distribution.

A recent trend in force fields has been the inclusion of polarizable charges that capture the change in local charge distribution due to molecular configuration.\textsuperscript{72,73} However, it is still unclear what molecular systems require simulations with this level of detail for fluctuating charges. Polarizable force fields come with a high computational cost, making the method orders of magnitude slower than partial charge methods.

Partial charge derivation approaches for molecular dynamics are difficult, as charges are dependent on the molecular conformation and local environment. Some force fields arbitrarily apply charges to atom types.\textsuperscript{74} The Amber ff99SB, GAFF, and Lipid14 force fields all use a similar approach for partial charge derivation. This method is known as RESP and is described briefly in the following section.

### 2.5.2 RESP charge derivations

The RESP method consists of an initial gas-phase electrostatic potential (ESP) calculation of the molecule at the HF/6-31G* level of theory and basis set.\textsuperscript{71,69} The electrostatic potential as a function of the charge density $\rho(r)$ is defined as

$$\phi(r) = \phi_n(r) + \phi_e(r) = \sum_{j=1}^{n} \frac{Z_j}{|r - R_j|} - \int \frac{dr' \rho(r')}{|r' - r|}$$

It is often discussed why some Amber force fields use the HF/6-31G* level of theory and basis set in gas phase for the ESP calculation.\textsuperscript{69} It has been observed that calculations with this level of theory and basis set overestimate the molecular dihedral moment. However, by a fortuitous cancellation of errors, this error is approximately the same as that which the dipole moment changes when the molecule is in solvent. ESP calculations perform well in condensed-phase simulations with explicit solvent.

The restrained ESP (RESP) method fits the charge density to discrete points
Figure 2.3: Van der Waals potential. Potential shown is for the GAFF c3 carbon type. Note that the repulsive term dominates this potential.

of the atom centers via a least-squares algorithm. During the first phase all atoms are simultaneously fit, and during the second phase the methyl and methylene hydrogens are equivalenced and refit. This method makes the the methyl and methylene hydrogen charges equivalent because in experiments these hydrogens are chemically indistinguishable.

2.5.3 Van der Waals interactions

The final non-bonded contribution to the energy function is from the van der Waals interactions. The name van der Waals refers to the researcher who first quantified these interactions from ideal gas experiments. In general, the van der Waals interactions are smaller in magnitude relative to the electrostatic potential for MD without unusual atomic overlaps. This contribution includes two distinct components: a repulsive and attractive term. The repulsive component is modeled after the Pauli exclusion electronic interactions. This term is very short range; namely, it increases rapidly as atoms become very close. The attractive term is from the dispersion forces first described by London. This type of dispersion force can partially be explained by the induced dipole attractive force between atoms within a certain range. The attractive term is not a very long range force and usually is relatively small relative to the repulsive term.
The most common form of the van der Waals interactions is described by a Lennard-Jones potential of the following form:

\[ E_{\text{vdw}} = \sum_{i<j}^n \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \]  

(2.10)

The coefficients \( A_{ij} \) and \( B_{ij} \) may be defined through \( A_{ij} = \epsilon r_m^{12} \) and \( B_{ij} = \epsilon r_m^6 \), where \( \epsilon \) is the well depth in kcal/mol and \( r_m \) is the well distance in Å. The repulsive interactions are modeled with the \( r_{ij}^{12} \) term, while the attractive term is represented in the \( r_{ij}^6 \) term. While the attractive \( r_{ij}^6 \) term has a physical basis, the repulsive \( r_{ij}^{12} \) term was mainly chosen for computational efficiency and because it decays rapidly.

Given \( t \) atom types, a MD simulation will have \((t^2 - t)/2\) van der Waals parameters. In order to simplify the parameterization of so many van der Waals terms, parameters are fit only for each individual atom type. After the van der Waals parameters are defined for each atom, the calculation of pairwise terms between different types is evaluated with the averages \( r_{AB} = (r_A^m + r_B^m)/2 \) and \( \epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \).

Van der Waals interactions decrease fairly rapidly with distance from the atom center. Therefore, at a certain distance, contributions from far-off particles are relatively small. Because it is so computationally slow to evaluate the van der Waals potential for all pairs of atoms, one approximation is to limit the number of pairwise potential calculations to those within a distance cutoff. Normal cutoff values for MD simulations range from 8 Å to 12 Å. Longer cutoffs may include more van der Waals contributions within the cutoff and are potentially more accurate, but come at a significant performance loss. Therefore, most MD programs use a cutoff based approach to limit the van der Waals calculations. The truncated interactions beyond the cutoff may then be approximated with a long-range correction term. This is the implementation for VDW that Amber MD software uses.
2.5.4 Many-body effects

The electrostatics and van der Waals potentials are calculated as the potential between pairs of atoms. However, this includes an assumption that other particles do not affect the pairwise potentials, and it explicitly omits the contributions from other particles on the pair of atoms. It is possible to include other contributions, but with a significant computational cost. Another approach to many-body effects is to generalize many-body effects for each potential and include them while parameterizing pairwise potentials. As a result, many-body effects may be included in the parameter set, but not explicitly in the functional form of the long-range interactions.

2.6 Long range interactions

Because the most intensive part of the force calculation for any MD simulation is calculating long range interactions, several clever methods have been developed that allow for efficient calculation of long range interactions.

2.6.1 Periodic boundary conditions

Condensed phase molecular dynamics with explicit solvent have boundaries at the edge of the system. Simulating this system alone, with just vacuum at the boundaries would be analogous to simulating a water drop in a vacuum. At the interfaces of the drop with the vacuum, water molecules would boil off into vacuum. It is possible to add other types of barrier beyond the system, but this approach may be susceptible to boundary artifacts.

A common approach to simulating biomolecules is to use periodic boundary conditions. In this case, molecules that translate beyond the box’s dimensions are treated as if they have translated into the box on the other side. Molecules at the
Figure 2.4: Periodic boundary conditions. 2D periodic boundary conditions are shown. The main system is in the middle panel. However molecules and forces may cross the periodic boundaries. Lx and Ly are the lengths of the 2D periodic box.

Periodic boundary experience forces from molecules on the other side of the box. As a result, the box is treated as an array of boxes. Figure 2.4 shows a 2D periodic boundary condition in effect. Periodic boundary conditions have been stably and reliably simulated in many condensed-phase MD simulations.

2.6.2 Non-bonded cutoffs

Periodic boundary conditions have practical implications in the evaluation of long-range interactions. Molecules at one edge of a periodic box will experience electrostatics and van der Waals forces across the periodic boundary. If treating these long range interactions with a cutoff, the cutoff must be less than half the shortest box dimension to ensure that the system only experiences forces from the first adjacent periodic image in each direction. As discussed in section 2.5.3, the van der Waals energy term uses a hard non-bonded cutoff coupled with a long-range correction to model the environment beyond the cutoff.
2.6.3 Particle mesh Ewald

One clever algorithm which takes advantage of the inherent periodicity of the partial charges across the periodic images is called the Ewald sum method.\textsuperscript{77} The Ewald sum method simplifies the calculation of the series of charges from the periodic images. The Ewald summation method has been applied to MD simulations for electrostatics calculations. Furthermore, performance was enhanced by using a grid based approach to represent partial charges called the particle mesh Ewald method.\textsuperscript{78} This algorithm reduces the algorithmic complexity of the problem to $O(n \log(n))$.

The electrostatic potential is replaced by the calculation of three different terms:

$$E_{elec} = E_{direct} + E_{recip} + E_{self}$$  \hspace{1cm} (2.11)

The summation of interatomic point charges is replaced by Gaussian charge distributions of the same magnitude but \textit{opposite} sign centered at the position of each particle. The potential takes the form of

$$E_{direct} = \frac{1}{2} \sum_{n}^{\text{atom}} \sum_{i,j} l_q i q_j \frac{\text{erfc}(\alpha r_{ij,n})}{r_{ij,n}}$$  \hspace{1cm} (2.12)

where $\text{erfc}()$ is the complementary error function, and $\alpha$ is the Ewald coefficient that determines how much the charge is shielded.

In order to offset the charge distributions introduced by the Gaussian functions, another Gaussian charge distribution is placed in the position of the charges with the \textit{same} magnitude and sign. This is called the reciprocal sum $E_{recip}$.

$$E_{recip} = \frac{1}{2\pi v} \sum \exp\left(-\left(\frac{\pi m/\alpha)^2}{m^2}\right) S(m)S(-m)$$  \hspace{1cm} (2.13)
where \( v \) is the volume of the cell, and \( m \) is a reciprocal-lattice vector, and \( S(m) \) is the structure factor. The structure factor may be defined as

\[
S(m) = \sum_{i=1}^{n_{\text{atom}}} t_q \exp(2\pi i m \times r_i)
\]  

(2.14)

Alternatively, the point charges may be represented as a grid of charges represented by \( Q(k_1, k_2, k_3) \) calculated from the interpolation of nearby point charges. Then the structure factor may be approximated with

\[
S(m) \approx \sum_{k_1, k_2, k_3} Q(k_1, k_2, k_3) \exp \left( 2\pi i \left( \frac{m_1 k_1}{K_1} + \frac{m_2 k_2}{K_2} + \frac{m_3 k_3}{K_3} \right) \right)
\]  

(2.15)

\[
= F(Q)(m_1, m_2, m_3)
\]  

(2.16)

\( F(Q) \) is the fast Fourier transform of \( Q \), for which there are algorithms to solve in \( n \log(n) \) time.

The self-interaction term is a correction term to remove the self-interactions included in the reciprocal term. It is of the form

\[
E_{\text{self}} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{n_{\text{atom}}} t_q^2
\]  

(2.17)

Combined with the particle mesh, the Ewald summation method reduces the complexity of the electrostatics calculation. Because this is the most computationally intensive portion of force calculations in MD, this change in algorithm leads to significant performance gains. Combined with graphics processing unit (GPU) acceleration, MD simulations can run much faster.\(^57\)
Figure 2.5: Ewald sum. Panel A shows the real space charge distribution of the Ewald sum with Gaussian distributions. Panel B shows the reciprocal space Gaussian distributions. Adapted from Allen.\textsuperscript{59}
2.7 Solvent models

With periodic boundary conditions, it is possible to simulate biomolecules with explicit solvent molecules. Solvent is represented by actual molecules in the simulation. In most biological systems, water and ions are the solvent of the system. However, simulation of water molecules is not trivial, with multiple water molecule models available.

The simplest water models are rigid models of oxygen bonded to two hydrogens. More complex models of water include flexibility of the bonds, or include polarization and many body effects. Simple water molecules include three particles that describe the water geometry. However, additional virtual atoms may be added to represent water molecule structure. The most common rigid models include TIP3P, TIP3P with an Ewald correction, TIP4P, TIP4P-Ew, and SPC-E. Three particles are included in these molecules centered on the oxygen and hydrogen nuclei. The TIP4P model also including an atom “lone pair” charge close to oxygen. Partial charges are included at the interaction sites and an artificial constraint is placed between the two hydrogen atoms. Hydrogens are assumed to have no van der Waals radius and instead only the water oxygen has van der Waals parameters.

Other water models include water flexibility, or possibly polarization effects. While these additional features may be important for modeling certain solvents, each additional feature increases computational cost relative to the simple rigid model. Unsurprisingly, simple models have been used mainly for their low computational cost. The number of parameters for each water molecule is minimized for efficient computations on a large number of molecules.

In addition to explicit water models, implicit solvent models are also available for MD. One common implicit solvent model is the Generalized Born (GB) model.
Instead of explicit water molecules, the solvent is simulated with a continuum elec-

trostatics equation. The general form of the GB correction to the solute energy may

be defined as

$$
\Delta E_{GB} = -\frac{1}{2} \sum_{i,j} \left( 1 - \frac{e^{-\kappa f_{ij}^{GB}}}{\epsilon_r} \right) \frac{q_i q_j}{4\pi\epsilon_0 f_{ij}^{GB}}
$$

(2.18)

where \( \kappa \) is the Debye-Hückel screening parameter. \( f_{ij}^{GB} \) is a function that defines the

interaction from the Born radius \( R \) at short distances to \( r_{ij} \) at long distances.

$$
f_{ij}^{GB} = \left[ r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j) \right]^{1/2}
$$

(2.19)

The physical meaning of the Born radius \( R_i \) can be described as how much a

charge buried in a low-dielectric medium like a solute. \( R_i \) is dependent on the intrinsic

radius \( \rho_i \) of the atom, and the radii and positions of other solute atoms.

Other non-polar solvents are possible for explicit solvation simulations as well. However, the most common MD solvent models are the explicit and implicit models

described above. Amber MD includes these solvent models by default.

## 2.8 Energy surfaces in molecular mechanics

Each biomolecular structure may sample an ensemble of conformations during

an MD simulation. If the molecule includes \( d \) degrees of freedom, this ensemble is a

d-dimensional surface with a structure uniquely defined by the position and velocities

of the molecule. Given the large number of degrees of freedom within a normal

MD simulation, these potential energy surfaces are large. For example, the potential

energy surface of a protein is very complex because of the number of internal degrees

do freedom of the protein structure.

For biomolecules like proteins, the energy landscape is inherently rugged with
many local minima and maxima corresponding to certain molecular conformations. During an MD simulation, the molecular system will sample a set of structures within a region of the energy surface. Because of the complexity of certain energetic barriers and the limited sampling time of MD, it is unlikely to explore the full energy surface for large molecules.

Many MD methods attempt to quantify and organize structures on energy surfaces. Clustering methods organize related structures on an energy surface. Often, it is important to examine transitions between different parts of the energy surface important for molecular functionality. Some methods, like the nudged elastic band algorithm, are designed to trace the lowest energy pathway across the energy surface.

Even with current computational power, it is still extremely difficult to sample the entire energy landscape through classical MD for all but the simplest systems. This has led to the rise of advanced sampling techniques. Some advanced sampling techniques include, but are not limited to, accelerated MD, replica exchange MD, and self-guided Langevin dynamics. Accelerated MD performance may be optimized with graphics processing units as well. These approaches enhance sampling beyond the local energy surfaces and are much more efficient than classical MD for studying the energy landscape of large biomolecules.

2.8.1 Energy minimization

Given a force field for molecular mechanics, it is possible to minimize the potential energy of a molecule by adjusting the coordinates of each atom for a lower total energy. This is commonly used to improve initial molecular structures before MD simulation by changing molecular structures so that unusual features such as atom overlap or bond lengths are brought closer to their optimum values.

The potential energy of a system is defined to be minimized when the first
derivative of the energy is zero and the second derivative of the energy is positive. However, given the complexity of the energy surface, it is highly unlikely to find the global minimum of a structure with just energy minimization. MD may sample structures on the potential energy surface around local minima.

Several optimization algorithms exist for energy minimization. Some minimization algorithms use derivatives of energy, while some do not. The order of the minimization algorithm refers to the highest order of the derivative used in the energy calculation.

The simplex algorithm is a zeroth order algorithm that does not use any energy derivatives for optimization. However, this approach is relatively slow to converge and unlikely to explore conformational space beyond a local minimum. Fortunately, due to the form of the Amber force field energy terms, the derivatives of the energy may be evaluated analytically. This allows for first order energy minimization algorithms using derivatives. Two commonly used minimization algorithms include the steepest descent and conjugate gradient methods.

The steepest descent energy minimization method uses the gradient of the energy to determine the direction of the search of structures. At each conformation, the gradient is calculated and the next conformation is chosen from the minimum along the gradient. This in turn is used as the starting point for the next search for a minimum along the gradient. For complex molecular systems, steepest descent does not converge well near local minima because the minimum search is always directly along the gradient.

The conjugate gradient approach is an adjustment to the steepest descent approach that can help converge conformations close to minima. Instead of searching along just the gradient for the current conformation, the algorithm uses the conjugate gradient. The conjugate gradient is a combination of the current gradient and the
gradient from the previous conformation. Thus, the search path contains information from previous conformations as well as the current conformation. When the conformation is close to the local minimum, this additional information guides the structure to a minimum faster.

It is common to use a combination of minimization approaches to relax an initial crystallographic structure before MD simulation. The steepest descent algorithm can initially quickly move the conformation away from any interactions with high energy. Following this, the conjugate-gradient algorithm allows for convergence to a local minimum. The Amber MD programs include multiple minimization algorithms.¹

2.9 Molecular dynamics algorithm

While energy minimization is useful in some contexts, the actual dynamics or change in time of the molecular system is more enlightening. The motion of biomolecules and their specific interactions with other molecules are particularly useful. The ergodic hypothesis states that the information from MD simulations under certain conditions can estimate statistical mechanics properties of the system.

Given that most molecular dynamics simulations have more than three particles, the classical dynamics of the system may not be solved analytically (n-body problem). Therefore, a number of numerical methods have been applied to molecular dynamics to solve Newton’s equations of motion for all atoms within the system. Time-propagation within MD is largely a matter of integration along discrete time steps within the simulation.

There are numerous numerical integration methods, with varying levels of complexity and accuracy. However, as noted previously, due to limits in computational power, performance must be considered as well. Given current computer hardware
capabilities, it is desirable to maximize sampling for the cost of simulations. Verlet’s algorithm and its variants for integration is probably one of the most widely used in MD.\textsuperscript{90} Other integrators like the Gear predictor integrator are more mathematically accurate, but come with significant computational cost.\textsuperscript{91} The Leap-frog version of the Verlet algorithm is a simple variant of the Verlet integrator that is widely used in MD.\textsuperscript{59}

**Algorithm 2.1** Leap-frog integrator

*Require:* positions \( r(t) \), velocities \( v(t - 1/2\Delta t) \)

Calculate forces \( F(t) \) from the atomic positions
\[
F_i(t) = -\nabla E_i(t)
\]
\[
a_i(t) = F_i(t)/m_i
\]

Calculate velocities from the previous velocities and current accelerations
\[
v_i(t + 1/2\Delta t) = v_i(t - 1/2\Delta t) + a_i(t)\Delta t
\]

Update previous positions with the current velocities
\[
r_i'(t + \Delta t) = r_i(t) + v_i(t + 1/2\Delta t)\Delta t
\]

The Leap-frog algorithm is summarized in algorithm 2.1. The Leap-frog integrator requires only the storage of the previous positions and velocities which are less than more advanced integration techniques. Other integrators use positions and velocities from multiple previous time steps. Note that the velocities are always a half-step out of sync with the positions. Therefore, to evaluate the kinetic energy at time \( t \) or \( t + \Delta t \), the velocities must be interpolated.

Time is measured discretely in a MD simulation and therefore has an associated time step duration. With classical MD involving no restraints, the upper limit for stable MD simulations is approximately 1 fs.\textsuperscript{1} This is fairly close to the inverse frequency of the most rapid motions within molecular systems. Usually the fastest vibrations of a molecular system include bond vibrations involving hydrogen. Time steps longer than 1 fs may result in instability of the bonds involving hydrogen.
However, there are several approximations available that may enable longer time steps. One approximation is to replace the vibrations of all bonds involving hydrogen with a distance constraint. Constrained molecular dynamics is possible by methods based on Lagrangian multipliers and was optimized with the SHAKE algorithm.\textsuperscript{92} It is now commonplace to apply SHAKE constraints to all bonds involving hydrogen to constrain the bond length. With this approach, the upper limit on timesteps increases to approximately 2 fs.\textsuperscript{1} The molecular dynamics timestep may potentially increase even further through methods such as hydrogen mass repartitioning.\textsuperscript{93} Hydrogen mass repartitioning replaces the mass of the hydrogen on other nearby atoms, approximating the dynamics of the attached hydrogen.

Integration of the molecular coordinates and velocities creates a trajectory as a function of time of the molecular system. The trajectories in turn may be used for a variety of analyses of the system’s dynamics.

### 2.9.1 Constant temperature and pressure dynamics

While basic molecular dynamics methods simulate a system with a constant number of molecules, volume, and energy, it may be desirable to simulate molecules in other thermodynamic ensembles. For example, the isothermal ensemble maintains constant temperature while the isothermal-isobaric ensemble maintains both temperature and pressure.

For isothermal ensembles, the temperature, which is a function of the kinetic energy of the system, must be kept constant. There are many methods called thermostats implemented in MD simulations that maintain constant temperature. These include the Andersen thermostat, the Berendsen thermostat, Nose-Hoover extended ensemble thermostat, and the Langevin thermostat.\textsuperscript{94,95,96,97,98} At a fundamental level, thermostats reset or scale the atomic velocities to maintain a specific system temper-
ature within a thermodynamic ensemble.

MD is also possible with constant pressure coupling. The algorithms that describe pressure scaling are known as barostats. Several common barostats include the Berendsen barostat, the Monte Carlo barostat, and the Parrinello-Rahman barostat. Barostats commonly however adjust the system pressure by scaling molecular coordinates, periodic box dimensions, and volumes.

Barostats may be classified by the coupling of pressure scaling dimensions. Pressure coupling in all dimensions is called isotropic and is the most common scheme for liquid simulations. Pressure coupling in two dimensions while the third is independent is called semi-isotropic. Pressure coupling independently in all dimensions is called anisotropic. Anisotropic pressure coupling is recommended for interfacial systems such as membranes as pressures along each axis may be independent.
Chapter 3

Lipid11

3.1 A modular framework for phospholipids

Phospholipids exhibit a wide variety of head group types and tail group types. Biological membranes are composed of complex mixtures of lipids, often classified by head group type. Glycerophospholipids have a common structure composed of a phosphate containing polar head group and two fatty acids esterified to a glycerol backbone. Differences mainly occur with the type of unit at the end of the nonpolar head group as well as with the length and saturation of the tail groups. Otherwise,

Figure 3.1: A framework for modular phospholipids. Two tail groups may be combined with any head group.
Figure 3.2: POPE bilayer. The phospholipids are rendered with van der Waals spheres. Solvent is shown with blue spheres.
the chemical structure of phospholipids are similar.

A common approach for molecular parameterization is to separate large molecules into smaller groups which are individually parameterized. In previous force fields, the aliphatic tails have been parameterized independently by fitting to alkane and alkene experimental data. Head groups have also been separated into small molecules for further parameterization. The reasons for parameterizing only portions of a molecule are varied. The size of the molecule may be an issue in parameterization, especially when fitting to quantum mechanics calculations. Furthermore, the size of a molecule may lead to noise in parameterization when attempting to fit individual parameters. Dividing the molecule into smaller parts has been effective in parameterization.

Smaller representative molecules are often used for parameterization of more complex molecules. Previous parameterizations have interchanged different tails attached to phosphatidylcholine head groups as well as swap the tails in the $sn-1$ and $sn-2$ position. Given that parameters have been transferred between lipids, interchangeability between tail and head group parameter sets has been attempted in lipid MD force fields. However, in previous work, parameters and topologies for the groups were combined and stored in full phospholipid units, not separate head and tail groups.

A new framework for lipid force field refinement was developed in 2011 called Lipid11 and released in Amber version 12 in 2012. For this work, a modular approach for parameterization and MD was proposed: each lipid molecule would be split into separate functional units. In principle, parameterization may be more efficient, if the chemical groups are shown to be interchangeable. Further, this approach ensures that parameter fitting occurs across an entire family of lipids, rather than individual lipids. Fitting across an entire family of lipids could potentially minimize overfitting...
of certain lipid types and allows for transferability within the family. This type of force field should also be smaller and more maintainable because data is stored separately in head and tail groups. Thus the force field enforces separation of information in units while allowing for any combination of head and tail groups.

Similar approaches have been used in all-atom force fields, most notably in protein force fields, where amino acids exist as reusable units from which to build enzymes.\textsuperscript{62,63,55} In the Amber ff99SB force field, protein residues are individual units that can be combined in any sequence. Based on atom naming and typing, some parameters are interchangeable between all residues. Certain parameters, for example in amino acid side chains, are unique.

One major issue with modular force fields is how to assign partial charges, because charges will depend on their environment and adjacent bonded groups and units. This becomes a major issue in non-polarizable force fields, where, in general, fixed partial charges are assigned to the center of all atoms. Clever partial charge derivation strategies have been used that address this problem. For example, in all-atom protein force fields, acceptable partial charge derivation approaches were developed for each amino acid independently of other amino acids. In the case of the Amber ff99SB force field, residues used the RESP method to derive charges with an additional capping group at each end of the residue.\textsuperscript{63,71} The capping group functions as an additional constraint in the restrained partial charge refinement. In some ways, the capping group may be thought of as modeling the electrostatic environment beyond the residue to be fit.

A prerequisite for a modular approach is to establish the independence of atomic partial charges between groups. In Lipid11, the interdependence of atomic partial charges of the glycerophospholipids between different lipid types was carefully evaluated. Using the same type of framework employed by the Amber protein force
fields ff94, ff99SB, and ff12SB, it is possible to separate lipids into three individual groups. For reasons that will be explained in the preliminary charge derivation section (section 3.3), these groups do not correspond directly to the traditional lipid head group and tail group. The lipid was split into interchangeable units that this text will refer to as “residues” because they were developed in a similar way to protein residues.

The new force field named Lipid11 is fully modular and defines residues, atom names, and atoms types consistent with the modular approach. Furthermore, a well-defined and reproducible charge derivation method was developed for the modular residues. This chapter examined the development of a modular all-atom lipid force field which forms the basis for all later lipid parameterization. Following the establishment of the Lipid11 framework, initial lipid bilayer structures were simulated with MD and compared with experimental data. After the framework was developed and tested, further parameterization was necessary and is fully described in the Lipid14 chapter and manuscript.

3.2 Initial parameters and atom types

Development of Lipid11 was done within the framework of the Amber molecular dynamics package. Ultimately, the force field parameterization was intended for use within the Amber suite of programs and for application to simulations with the other Amber force fields including the protein, nucleic acid, carbohydrate, and small molecule force fields.

Lipid11 used the General Amber Force Field from AmberTools 11 as a template for establishing a modular lipid framework. Most atom types of Lipid11 are derived from the GAFF and atom typing is fully explained in section 3.4.

In addition to the GAFF, a small subset of Glycam06 from AmberTools 11
parameters were included into the force field.\textsuperscript{65} The Lipid11 phosphoinositol head group includes an inositol group, the parameters for which are taken from Glycam. It is important to note that the 1-4 van der Waals and 1-4 electrostatic scaling factors are not scaled in Glycam06, while scaling does occur in GAFF. The GAFF and ff99SB force fields utilize 2.0 and 1.2 non-bonded scaling factors for 1-4 van der Waals and electrostatic interactions respectively. However, this is not an issue within the Amber programs because Amber 12 and beyond supports mixed scaling factors.

The GAFF has previously been applied to phospholipids with some success for bilayer molecular dynamics.\textsuperscript{48,103,104,105} However, in MD simulations the acyl chain tails were too ordered and did not reproduce the correct lipid bilayer phase.\textsuperscript{48} This incorrect phase could be alleviated somewhat with additional constraints such as constant area or constant surface tension constraints applied to the MD pressure coupling. Yet other force fields such as Charmm C36 for lipids were able to re-parameterize the lipids such that bilayers reproduce experimental structural features in a tensionless ensemble.\textsuperscript{49} The GAFFlipid force field used the GAFF as an initial parameter set with this in mind for re-parameterization of lipids for bilayer MD in a tensionless ensemble.\textsuperscript{2}

Lipid11 and further parameterization may be thought of as a fork of the GAFF and Glycam force fields.\textsuperscript{64,65} Parameters were chosen and refined from that point and now represent a fully independent force field included in Amber. Lipid11 and subsequent force fields do not depend on GAFF and Glycam parameters, so future updates to GAFF or Glycam would have to manually be modified in Lipid11. New atom types defined in Lipid11 correspond to an original GAFF or Glycam type allowing for further customization and parameterization independent of the original force fields. Atom types were derived from the GAFF and Glycam force fields and have evolved with later parameterization efforts.\textsuperscript{4}
3.3 Preliminary charge derivation

In order for a modular glycerophospholipid force field to be effective, the partial charges of each unit presumably should be independent of other units. Namely, the charges in the head unit should not influence the tail unit and the charges in the tail unit should not influence the head unit. If this holds true, the separation of charges would support a modular approach.

In order to investigate the dependence of each units’ partial charges, representative glycerophospholipid molecules were chosen in which the head group and the tail groups would be swapped. Then, the change in partial charges would evaluated for the change in head or tail group. For this task, dipalmitoylphosphatidylcholine (DPPC), palmitoyloleoylphosphatidylcholine (POPC), dipalmitoylphosphatidylserine (DPPS), and parmitoyloleoylphosphatidylcholine (POPE) were selected. Thus the difference in partial charges would be calculated between phosphatidylcholine and
phosphatidylserine for the head groups as well as palimitoyl and oleoyl for the sn-2 tail group. Figure 3.3 summarizes the charge derivations.

Charge derivations were conducted in the same method as the Amber RESP procedure described by Bayly and applied to biological molecules by Wang.\textsuperscript{71,106} This procedure has been applied to several of the other non-polarizable all-atom Amber force fields. It is well known that electrostatic potential (ESP) calculations depend on grid orientation. The program RED uses a standard protocol with reproducible ESP grid orientations for reproducible charge derivations.\textsuperscript{107} Charges were derived with RED version 3 with six ESP grid orientations for each molecule. RED calls Gaussian 03 for the molecular ESPs at HF/6-31G* level of theory and basis set.\textsuperscript{108} The Amber RESP method includes two stages of fitting with hyperbolic restraints of 0.0005 and 0.001 at each stage.

Figure 3.5 summarizes the change in charges for the representative lipid molecules. Change in the tail group type resulted in small (<0.05 e) change in the head partial charges and the glycerol and ester linkage region. Change in the head group led to small changes (<0.05 e) in the tail partial atomic charges. However, when the head group was changed from phosphatidylcholine to phosphatidylserine, the charges in the glycerol and ester linkage region were significant (>0.05 e). The charges in this region were dependent on the head group type. Therefore, when building parameter and topologies for lipids with Lipid11, the glycerol and ester linkages would depend on the head group. The solution taken by Lipid11 was to separate the glycerol and ester groups from the tail group and instead include them in the head group.

The glycerophospholipid was then split into three so-called “residues” for each lipid type: one head group containing the glycerol and ester linkages and two acyl chain tail groups. The residue is split between the carbonyl carbon of the ester group and the following carbon along the tail acyl chain. Figure 3.6 panel A shows
Figure 3.4: Preliminary charge derivation for full lipid molecules. The head group was interchanged between PC and PS. This change is highlighted with blue arrows and blue text. The sn-2 tail group was interchanged between palmitoyl and oleoyl. This change is highlighted in green text. With these combinations, partial atomic charges for four lipid types was derived: DPPC, DOPC, DPPS, and DOPS. The table in panel B summarizes the change in partial atomic charge \( \Delta q \) in units of electron charge for the glycerol group. Tables in panels B, C, D, and E summarize the change in partial atomic charge \( \Delta q \) in units of electron charge for the phosphatidyl-alcohol moiety, the glycerol group, the sn-2 chain and the sn-1 chain. \( \Delta q > 0.05 \, \text{e} \) are printed in red text in the tables. Note that atom naming in this charge derivation does not correspond to Lipid11 atom naming.
**Figure 3.5:** Preliminary charge derivation for full lipid molecules: head and tails. Tables in panels A, B, and C summarize the change in partial atomic charge $\Delta q$ in units of electron charge for the phosphatidyl-alcohol moiety, the sn-2 chain, and the sn-1 chain. Note that atom naming in this charge derivation does not correspond to Lipid11 atom naming.
Figure 3.6: Modular splitting of lipids. Panel A shows the splitting of glycerophospholipids into three residues: one head residue and two tail residues. The red wavy lines indicate the break point between residues. Panel B shows the charge derivation for a molecule which is the combination of the two capping groups. Two head caps are added to head group residues and one tail cap is added to the tail group residues which function as constraints on the following RESP charge derivation.

The splitting of the lipid into residues. In the following charge derivations using the RESP method, the ester linkage was thus included in the head group and was shown to significantly depend on head group type.

In addition to this splitting strategy, capping groups were included as constraints to the RESP charge derivation method. Head residues were capped with two methyl groups and tail residues were capped with an ester and methyl group. In an analogous way to capping amino acids in a protein force field, the methyl and ester group were combined into one molecule. Charges were derived for this molecule, in which each capping group charge was constrained to be zero using the previously described Amber RESP procedure.

Those charges are presented in figure 3.6 panel B with the tail cap in the blue box and the head caps in the red box. The charges of the caps were then constrained in the following charge derivation. As indicated in panel B, the charges
Figure 3.7: Representative Lipid11 residues. Residue names, atom names, atom type names, and partial charges are labeled. The PS residue consists of the phosphatidylserine head group and the ester linkage. Note that the term “head group residue” includes the ester group in addition to the normal head group. This is true for all Lipid11 residues. The OL residue is the aliphatic moiety of the oleoyl acyl chain. These drawings follow the naming and typing described in the text. Drawings of all residues are included in the supporting material of Lipid11.3

function as an additional constraint to the actual residue RESP charge derivation. This approach is analogous to the Amber ff99SB strategy that uses acetyl and N-methylamide groups for capping groups. Final charges for all Lipid11 residues are provided in the supporting material of Lipid11.3

3.4 Lipid residue names and atom types

Lipid11 includes parameters for a set of residues including several head group types, tail group types, and cholesterol. Residue names are listed in table 3.1.

Common phospholipid head groups from biological membranes are included and listed in appendix A. One head group residue, phosphatidic acid, is unique in
Table 3.1: Lipid11 residue definitions

<table>
<thead>
<tr>
<th>Lipid11 residue name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acyl chains</td>
<td></td>
</tr>
<tr>
<td>Palmitoyl (16:0)</td>
<td>PA</td>
</tr>
<tr>
<td>Stearoyl (18:0)</td>
<td>ST</td>
</tr>
<tr>
<td>Oleoyl (18:1 n-9)</td>
<td>OL</td>
</tr>
<tr>
<td>Linoleoyl (18:2 n-6)</td>
<td>LEO</td>
</tr>
<tr>
<td>Linolenoyl (18:3 n-3)</td>
<td>LEN</td>
</tr>
<tr>
<td>Arachidonoyl (20:4 n-6)</td>
<td>AR</td>
</tr>
<tr>
<td>Docosahexaenoyl (22:6 n-3)</td>
<td>DHA</td>
</tr>
<tr>
<td>Head groups</td>
<td></td>
</tr>
<tr>
<td>Phosphatidylcholine</td>
<td>PC</td>
</tr>
<tr>
<td>Phosphatidylethanolamine</td>
<td>PE</td>
</tr>
<tr>
<td>Phosphatidylserine</td>
<td>PS</td>
</tr>
<tr>
<td>Phosphatidic acid (HPO₄⁻)</td>
<td>PH-</td>
</tr>
<tr>
<td>Phosphatidic acid (PO₄²⁻)</td>
<td>P2-</td>
</tr>
<tr>
<td>R-phosphatidylglycerol</td>
<td>PGR</td>
</tr>
<tr>
<td>S-phosphatidylglycerol</td>
<td>PGS</td>
</tr>
<tr>
<td>Phosphatidylinositol</td>
<td>PI</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Cholesterol</td>
<td>CHL</td>
</tr>
</tbody>
</table>

that it has a second \( pK_a \) value within physiological range.\(^{109}\) The protonated and deprotonated forms of this residue are available in the Lipid11 framework. Phosphatidylglycerol is a head group with an additional chiral center in an additional glycerol in the head group. While the \( S \) enantiomer is the primary natural enantiomer, Lipid11 provides a topology for either chiral form.\(^{110}\) The \( R \) enantiomer is also available as it is sometimes used in experiments with synthetic PG and also may be present in some forms of cardiolipin.\(^{110}\)

Lipid11 includes common saturated and unsaturated fatty acids available for combination with the head groups. In addition to these standard head and tail groups, Lipid11 includes topology and preliminary parameters for other lipid components. Phosphoinositol is another important bilayer component that is indicated in several key membrane protein and kinase signaling pathways.\(^{111,112}\) It should be noted that the phosphoinositol residue included in Lipid11 does not include any additional phosphorylation states.

Because of the importance of sterols and specifically cholesterol in biological
Table 3.2: Lipid11 atom types from the General Amber Force Field\textsuperscript{64}

<table>
<thead>
<tr>
<th>GAFF</th>
<th>Lipid11</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>c3</td>
<td>cA</td>
<td>sp3 carbon</td>
</tr>
<tr>
<td>c2</td>
<td>cB</td>
<td>aliphatic sp2 carbon</td>
</tr>
<tr>
<td>c2</td>
<td>cC</td>
<td>carbonyl sp2 carbon</td>
</tr>
<tr>
<td>o</td>
<td>oC</td>
<td>carbonyl sp2 oxygen in ester group (i.e., C=O)</td>
</tr>
<tr>
<td>os</td>
<td>oS</td>
<td>sp3 oxygen in ester group</td>
</tr>
<tr>
<td>o</td>
<td>oO</td>
<td>sp2 oxygen in carboxyl group (i.e., COO)</td>
</tr>
<tr>
<td>o</td>
<td>oP</td>
<td>sp2 oxygen with one connected atom (phosphorus) in phosphate group</td>
</tr>
<tr>
<td>os</td>
<td>oT</td>
<td>sp3 oxygen in phosphate group</td>
</tr>
<tr>
<td>oH</td>
<td>oH</td>
<td>sp3 oxygen in hydroxyl group</td>
</tr>
<tr>
<td>n4</td>
<td>nA</td>
<td>sp3 nitrogen with four connected atoms</td>
</tr>
<tr>
<td>p5</td>
<td>pA</td>
<td>phosphorus with four connected atoms, such as in PO\textsubscript{4}\textsuperscript{2}</td>
</tr>
<tr>
<td>hc</td>
<td>hA</td>
<td>hydrogen bonded to aliphatic carbon without electron withdrawing group</td>
</tr>
<tr>
<td>h1</td>
<td>hE</td>
<td>hydrogen bonded to aliphatic carbon with one electron withdrawing group</td>
</tr>
<tr>
<td>hx</td>
<td>hX</td>
<td>hydrogen bonded to carbon next to positively charged group</td>
</tr>
<tr>
<td>ha</td>
<td>hB</td>
<td>hydrogen bonded to aromatic carbon</td>
</tr>
<tr>
<td>ln</td>
<td>hN</td>
<td>hydrogen bonded to nitrogen</td>
</tr>
<tr>
<td>ho</td>
<td>hO</td>
<td>hydrogen in hydroxyl group</td>
</tr>
</tbody>
</table>

membranes, preliminary cholesterol parameters were also included in Lipid11. Cholesterol, along with sphingolipids and glycolipids, plays a key role in bilayer structure and dynamics.\textsuperscript{113,114} A full description of parameterization and simulation of cholesterol lipid bilayers is included in chapter 5.

Lipid11 atom types are defined with one lower case and one upper case character. The full definitions of all atom types derived from the GAFF are included in table 3.2 and atom types derived from Glycam are included in table 3.3.\textsuperscript{64,65}

The default GAFF oxygens were separated in Lipid11 to differentiate carbonyl sp2 oxygens in ester groups, sp2 oxygens in carboxyl groups, and sp2 oxygens with
Table 3.3: Lipid11 atom types from the GLYCAM force field\textsuperscript{65}

<table>
<thead>
<tr>
<th>Glycam</th>
<th>LIPID11</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>cR</td>
<td>sp\textsuperscript{3} carbon in inositol ring</td>
</tr>
<tr>
<td>CP</td>
<td>cP</td>
<td>sp\textsuperscript{3} carbon bonded to an oxygen bonded to a phosphorus in inositol ring</td>
</tr>
<tr>
<td>OH</td>
<td>oR</td>
<td>oxygen in hydroxyl group in inositol ring</td>
</tr>
<tr>
<td>HO</td>
<td>hR</td>
<td>hydrogen in hydroxyl group in inositol ring</td>
</tr>
<tr>
<td>H1</td>
<td>hS</td>
<td>hydrogen bonded to aliphatic carbon with one electron withdrawing group in inositol ring</td>
</tr>
</tbody>
</table>

one connected atom in phosphate groups. Furthermore, the GAFF sp\textsuperscript{3} oxygen was differentiated for ester and phosphate groups. This splitting anticipated future independent refinement of parameters within these regions. Otherwise atom types mainly correspond to their original force field.

3.4.1 Phospholipid nomenclature

Currently, there is no widely accepted standard for atom names for lipids for molecular simulation. Consequently, a specific atom naming scheme was established for the lipids of Lipid11. The Amber program \textit{leap} is able to translate a properly formatted PDB with Lipid11 residue and atom names automatically. This program can assign the lipid parameters and topology with no additional user intervention because lipid residues were defined in the same way as protein residues and use the same type of residue connectivity. Amber 12 also includes scripts that can convert PDBs with other lipid naming schemes into a format consistent with Lipid11.

Phospholipid nomenclature for Lipid11 is based on the stereospecificity of glycerol derivatives as described by IUPAC.\textsuperscript{115} The glycerol group within the phospholipid contains three groups: C1 for the \textit{sn-1} carbon, C2 for the \textit{sn-2} carbon, and C3 for the \textit{sn-3} carbon. The atom name begins with the element type followed by a digit.
indicating whether the atom belongs to a tail (1 or 2) or a head group (3). The tail residues are interchangeable between the sn-1 and sn-2 positions, and therefore have the same nomenclature in either position. An sn-1 and sn-2 tail is differentiated by the residue order in the input structure file. This is possible because the Amber program leap establishes the connectivity of lipid residues based on the order of the residues which is described in the Lipid11 PDB file format section.

Hydrogens are numbered using the corresponding heavy atom number. If bonded to a nitrogen or oxygen, an “N” or “O” is included. Multiple hydrogens are distinguished by the letters “A”, “B”, and “C” in the head group or “R”, “S”, and “T” in the tail groups.

### 3.4.2 Other molecules

In addition to the phospholipids, an inositol head group and cholesterol are included with Lipid11. Inositol nomenclature follows IUPAC recommendations for cyclitol numbering. The first carbon is defined to be the carbon bonded to the phosphate group and is named C31. When shown in a chair configuration with C31 in the bottom left, carbon numbers continue in a counter-clockwise fashion from C31 to C36. Oxygens in the hydroxyl substituents are numbered in a counter-clockwise order from O35 to O39. All other atoms follow the same rules defined for the other head group residues.

Cholesterol is a common sterol found in membranes. Cholesterol carbon names conform to the IUPAC naming standard for steroids with carbons numbered from C1 to C27. Hydrogens bonded to carbons are assigned the same number as the carbon combined with the unique hydrogen number (1 to 3). The hydroxyl oxygen is named O1 and the hydroxyl hydrogen is named HO1.


3.4.3 Lipid11 PDB file format

The modular nature of Lipid11 allows for many different combinations of three head groups and tail groups. Within the Amber programs, leap is responsible for reading input structures, establishing molecular topology, and assigning parameters. In order for leap to accomplish this with Lipid11, a lipid in a PDB file must be formatted such that the residues follow a specific order. The three residues must be placed in the following order: \textit{sn-1} tail, head, and \textit{sn-2} tail. Because each lipid molecule is formatted like a protein chain, leap will automatically connect the three residues and build the full lipid molecule. Each set of three residues must be followed by a \texttt{TER} line. With this format, leap can read the PDB with lipids, build a lipid topology and assign parameters to the molecules.

There are many approaches to building and modeling initial membrane structures.\textsuperscript{118,119} Regardless of the strategy used, many structure builders have residue and atom names dependent on the force field used. Because there is no standard for lipid residue and atom naming, the residue and atom names must be converted.

One common naming convention is that used by the Charmm C36 force field.\textsuperscript{49} Structures generated with the CHARMM-GUI use this atom naming.\textsuperscript{118} Lipid11 includes a simple shell script named \texttt{charmmlipid2amber.x} to convert PDB structures with the Charmm C36 nomenclature to a format compatible with Lipid11 and leap.

3.5 Lipid bilayer molecular dynamics

Following the initial charge derivation, the modular force field was applied to sample lipid bilayers that were then simulated with molecular dynamics.

Molecular dynamics simulations of lipid bilayers with early parameter sets at constant temperature and pressure often resulted in incorrect bilayer phases.\textsuperscript{54} In
Table 3.4: Lipid11 molecular dynamics simulations. Structure and simulation parameters for each bilayer type are listed here

<table>
<thead>
<tr>
<th>Lipid</th>
<th>$N_L$</th>
<th>Time (ns)</th>
<th>$T$ (K)</th>
<th>$\gamma$ (dyn/cm/interface)</th>
<th>$N_W$</th>
<th>Exp. $N_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOPC</td>
<td>128</td>
<td>100</td>
<td>300</td>
<td>10</td>
<td>37.4</td>
<td>32.8$^{15}$</td>
</tr>
<tr>
<td>POPE</td>
<td>128</td>
<td>100</td>
<td>310</td>
<td>26</td>
<td>31.6</td>
<td>13.5$^{20}$</td>
</tr>
<tr>
<td>POPC</td>
<td>128</td>
<td>100</td>
<td>300</td>
<td>17, 20</td>
<td>31.6</td>
<td>31$^{15}$</td>
</tr>
</tbody>
</table>

those early simulations, the bilayer lipid tails would enter a gel-like phase in which tails were too ordered.

One potential correction for this problem is to employ alternative pressure coupling schemes to influence bilayer structure. The use of barostats in molecular dynamics controls system pressure through various algorithms, which usually include scaling of system dimensions. Pressure coupling may be independent along each axis (anisotropic), coupled along two axes (semi-isotropic), or all coupled (isotropic). It has been argued that anisotropic pressure coupling is most appropriate for interfacial systems as pressure changes along each axis should be independent.$^{121}$ Semi-isotropic pressure coupling combines pressure coupling in the two dimensions of the bilayer independent of the dimension normal to the bilayer. In effect, a semi-isotropic barostat links and controls the relative lateral bilayer dimensions together.

Since the area per lipid of a lipid bilayer can be determined with relatively high precision, one proposed correction to the lipid bilayer phase is to constrain the system dimensions within the bilayer plane to a set area per lipid.$^{122}$ This can be accomplished by establishing initial planar dimensions and then allowing only the system dimensions normal to the bilayer to vary. Constant surface area allows for system pressure and density to equilibrate in the direction normal to the bilayer. Simulations thus allow for an area per lipid to be set for each bilayer type that directly matches experimental values.

Another variation of these pressure control methods is constant surface tension
Constant surface tension allows system dimensions to fluctuate in all dimensions, however with an additional constraint. The surface tension $\gamma$ is defined as

$$\gamma = \int_{-\infty}^{\infty} dz[P_n - P_t(z)]$$  \hspace{1cm} (3.1)

The axis normal to the bilayer is defined to be $z$, $P_n$ is the pressure normal to the vector, and $P_t(z)$ is the tangential pressure as a function of $z$. During an MD simulation, a surface tension and target pressure $P_0$ is set. At a given surface tension and periodic box dimensions, the target pressure in each dimension is adjusted to conform to equation 3.1.\textsuperscript{123,101}

Previous lipid bilayers simulated with GAFF parameters resulted in bilayers in the incorrect gel phase, which needed further correction.\textsuperscript{2} With the initial charge derivation and GAFF parameters, Lipid11 suffered from similar issues. In order to compare with other first generation lipid force fields, it was necessary to use alternate thermodynamic ensembles like constant surface tension. Semi-isotropic pressure coupling and constant surface tension were implemented in Amber 12 and used for lipid bilayer simulations with Lipid11.\textsuperscript{101}

### 3.5.1 Lipid bilayer simulation methods

Several pure glycerophospholipid bilayers were selected for initial testing including a set of available head and tail groups: dioleoylphosphatidylcholine (DOPC), palmitoyloleoylphosphatidylethanolamine (POPE), and palmitoyloleoylphosphatidylcholine (POPC). Initial structures were generated with the CHARMM-GUI membrane builder, as well as obtained from previous simulations done by Siu available in the Lipidbook database.\textsuperscript{118,103,124} Structures were solvated with TIP3P water molecules greater than the full hydration solvation number $n_w$ specific to that type of
lipid. The Lipid11 framework and parameters were used for the lipids. Lipid bilayers were simulated with the Amber 12 software package using the GPU-accelerated version of pmemd called pmemd.cuda. Bilayers were minimized for 10000 steps; the first and second halves were minimized with the conjugate gradient and steepest descent algorithms, respectively. Lipid bilayers were then heated from 0 K to 100 K within 50 ps under constant volume simulations with the Langevin thermostat. Then, bilayers were further heated to production temperature over 1 ns with the box size allowed to vary via pressure control with a Berendsen barostat. Lipids were restrained with a harmonic potential for the heating phases of the simulation with a force constant of 10 kcal/mol Å^2.

Langevin dynamics was used for temperature control with a decay constant of 1.0 ps^{-1}. The Berendsen barostat was also used with a relaxation time of 1.0 ps and target pressure of 1 atm. Constant surface tension and semi-isotropc pressure coupling were enabled for the barostat. Constant surface tension values differed for each bilayer type. Bonds that included a hydrogen were constrained with the SHAKE algorithm. The pmemd leap-frog integrator used a 2.0 fs timestep throughout all simulations. Electrostatic interactions were calculated with the particle mesh Ewald (PME) summation method. Van der Waals interactions were truncated at 8.0 Å and the direct space cutoff for PME was set to the same value.

Table 3.4 summarizes the system parameters for the lipid bilayers chosen for simulation. Production temperature was set above the phase transition temperature of the lipid bilayer type. Multiple simulations at a range of surface tensions were chosen for each bilayer ranging from 0 to 30 dyne/cm. The full range of surface tension values simulated is plotted in figure 3.8.

Simulation convergence was determined by monitoring lipid bilayer experimental properties such as area per lipid and density throughout the course of the
simulation. The equilibrated portion of trajectories were analyzed to calculate other bilayer structural features. Simulations that deviated greatly from the experimental area per lipid of the bilayer were stopped, while simulations close to the experimental area per lipid were extended to at least 100 ns.

3.6 Lipid bilayer experimental properties

The initial molecular dynamics simulations of the lipid bilayers were compared with available experimental bilayer structural properties. Lipid bilayer structural data is available in several common forms. This data and analysis is summarized in the following sections.

3.6.1 Area per lipid

Area per lipid is measurement of the dimensions of the lateral area of the bilayer divided by the number of lipids within the bilayer per interface.

\[ A_L = \frac{A_{XY}}{N_L/N_{leaflet}} \]  \hspace{1cm} (3.2)

In equation 3.2 assume the bilayer is oriented in an orthorhombic box in the \( XY \) plane and the dimension \( Z \) is normal to the bilayer. The area per lipid \( A_L \) is a function of the lateral area of bilayer calculated from the periodic box dimensions \( A_{XY} \). The bilayer area is divided by the total number of lipids in the system \( N_L \) and the number of leaflets \( N_{leaflet} \). Given the box dimensions throughout a trajectory, it is relatively straightforward to calculate this value.

This value may be obtained experimentally with relatively high precision, however, some bilayer types have multiple published values available.\(^{15,127}\) Relative area per lipid values may give an indication of the phase of the bilayer. Usually a lower
area per lipid indicates that the lipids are packed together tighter and the tails may be more ordered. While area per lipid is an important experimental bilayer structural property, it is one of many bilayer features.

3.6.2 Electron density profile and thickness

Electron densities may be modeled from experimental X-ray and neutron scattering profiles through the bilayer. Scattering profiles are fit to molecular models to obtain an electron density profile.\textsuperscript{16,127}

The electron density profile (EDP) shows the electron density of the bilayer and/or its components through the bilayer. The center of mass of the bilayer is defined as the origin and the plot shows the electron density passing through the bilayer.

Structural features of bilayers such as the high density of the phosphate group and the relative low density of the alkyl tails can often be resolved in the electron density. Often EDPs from MD simulations resolve the density profiles of component groups of the phospholipid.

Several features from the electron density may be calculated, including the thickness of the bilayer. The bilayer thickness may be defined in several ways including head-to-head thickness $D_{HH}$ and Luzzatti thickness $D_B$. The head-to-head thickness $D_{HH}$ is defined as the distance between the two leaflet density peaks. The Luzzatti thickness is defined in section 4.4.4 and was not calculated for Lipid11 bilayers.

Electron density may be evaluated from MD simulations in a straightforward manner. The system box is divided in 0.1 Å slices through the bilayer. The number of atom centers which fall within each slice is the number density, from which the electron density may be calculated trivially.
3.6.3 Deuterium order parameters

Solid state NMR experiments with $^2\text{H}$ and $^1\text{H}-^{13}\text{C}$ are able to reveal the orientation and configuration of lipid tails. The deuterium order parameter $S_{CD}$ is defined to be

$$S_{CD} = \frac{1}{2} \langle 3\cos^2\theta - 1 \rangle$$  \hspace{1cm} (3.3)

where $\theta$ is the angle between the carbon-deuterium vector relative to the bilayer normal vector. In the NMR experiments, order parameters must be carefully assigned to each carbon in the chain, and it is sometimes not possible to resolve the order for all carbons. From MD simulations, order parameters are calculated based on the average of the C-D vectors throughout a trajectory. In general, a higher $S_{CD}$ value corresponds to a segment of the hydrocarbon chain that is more highly ordered. Order profiles are specific to each individual lipid type and degree of saturation.

The structure and dynamics of phospholipid acyl chain tails are closely related to the phase of lipid bilayers. Depending on the bilayer phase, the tails may be found in various configurations, ranging from highly disordered, fluid-like aggregates to highly ordered, liquid-crystalline states. Order parameters represent the ensemble average of order of all $sn-1$ or $sn-2$ tails.

Because lipids are generally oriented in the same direction in a bilayer structure, order may correlate to the $trans$-$gauche$ sequences and ratios of the bilayer. Highly ordered tails are more likely to be found in an all $trans$ configuration, while disorder is associated with more flexibility and kinks in the acyl chain tail.
Table 3.5: Lipid bilayer molecular dynamics structural properties compared with experimental values

<table>
<thead>
<tr>
<th></th>
<th>DOPC</th>
<th>POPE</th>
<th>POPC</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_L$ (Å)</td>
<td>$D_{HH}$ (Å)</td>
<td>$A_L$ (Å²)</td>
<td>$D_{HH}$ (Å)</td>
<td>$\gamma$ (dyn/cm/ int.)</td>
</tr>
<tr>
<td>Lipid11</td>
<td>69.95 ± 0.75</td>
<td>36.8</td>
<td>62.24 ± 1.18</td>
<td>37.9</td>
<td>17</td>
</tr>
<tr>
<td>Exp.</td>
<td>67.4 ± 1.0$^{16}$</td>
<td>36.7$^{16}$</td>
<td>61.0$^{120}$</td>
<td>39.0$^{120}$</td>
<td>64.3 ± 1.3$^{127}$</td>
</tr>
</tbody>
</table>

3.7 Lipid11 structural properties

Lipid bilayers were simulated with the initial Lipid11 charge derivation and initial GAFF parameters. Bilayers were simulated at a range of surface tensions and structural properties were compared directly with experimental structural data. Results from each surface tension were compared to determine the conditions in which the bilayers resembled experimental results most closely. Trajectories at surface tensions that fit the experimental area per lipid best were extended and analyzed in more detail and only the equilibrated portion of the bilayer trajectory was used for further analysis.

Figure 3.8 plots area per lipid throughout the trajectories at different surface tension values. Surface tensions ranged from 0 dyne/cm (equivalent to an NTP simulation) to 30 dyne/cm. Depending on surface tension, bilayers would equilibrate into a range of box dimensions and corresponding area per lipids. Simulations with no surface tension equilibrated to area per lipid values well below experiment. Increasing surface tension values, in general, corresponded to higher area per lipid bilayers.

Figure 3.8 shows the area per lipid of DOPC bilayers at various surface tension values. The area per lipid of 10 dyne/cm equilibrated to an area per lipid that most closely resembled the published experimental value.$^{16}$ Figure 3.9 plots the DOPC bilayer electron density profile in comparison with the experimental electron density.
Figure 3.8: Area per lipid throughout lipid bilayer MD simulation. Surface tension values are reported as dyne/cm per interface. No surface tension was used in NPT simulations. Panel A shows area per lipid for DOPC. Experimental value are from Kucerka. Pane B shows area per lipid for POPE. Experimental value taken from Rappolt. Panel C shows the area per lipid for POPC. Experimental values are from Kucerka.
Figure 3.9: Electron density profile through lipid bilayers from MD simulation. Distance from bilayer center is along the axis normal to the bilayer plane. Surface tension values are reported as dyne/cm per interface. Panel A shows the electron density profile for DOPC. Experimental profile is from Kucerka. Panel B shows the electron density profile for POPE. Panel C shows the electron density profile for POPC at 17 and 20 dyne/cm per interface. Experimental density profile is from Kucerka.
Figure 3.10: Deuterium order parameters for \textit{sn-1} and \textit{sn-2} tails of phospholipids. Surface tension values are reported as dyne/cm per interface. Panel A shows the order parameters for DOPC. Experimental order parameters are from Warschawski. Panel B shows order parameters for POPE. Experimental order parameters for \textit{sn-1} are from Lafleur and \textit{sn-2} are from Perly. Panel C shows order parameters for POPC. Experimental order parameters for \textit{sn-1} are from Seelig and \textit{sn-2} are from Seelig and Perly.
The DOPC electron density profile was in good agreement with the experimental profile. The corresponding bilayer thickness in table 3.5 matches the experimental thickness as well. Figure 3.10 plots the order parameters for the sn-1 and sn-2 oleoyl tails. Deuterium order parameters had a similar profile relative to experimental values; in particular, the characteristic dip in order for carbons 10 and 11 in the double bond was captured. However, order parameters were overestimated for carbons 15-18. This may, in part, be due to the temperature difference between experiment and simulation with temperatures of 310 K and 300 K, respectively. However, increased order in the saturated portion of lipid tails has been observed with just GAFF parameters previously.

Given the limitations of the GAFF, the Lipid11 force field exhibited similar limitations. However, the modular approach and charge model were at least capable of reproducing previous lipid MD with GAFF results, and thus was potentially a good platform for further parameterization. Further re-parameterization of the van der Waals and dihedral terms within the tails were necessary to reproduce experimental order parameters.

Figure 3.8 displays the area per lipid of POPE at a range of surface tension values. A high surface tension of 26 dyne/cm was required to reproduce the area per lipid of the POPE bilayer. A surface tension value this high seemed unusual, but was consistent with previous bilayer simulations with GAFF parameters. System structural properties were analyzed at the surface tension of 26 dyne/cm. The electron density profile was calculated from the equilibrated trajectory and is plotted in figure 3.9. Unfortunately, no absolute experimental electron density profile was available for comparison. Nevertheless, this profile was compared to a previously published simulation relative density profile and was found to have a similar profile (not shown in figure). The calculated thickness from the profile resembled published experi-
Figure 3.10 shows the order parameters for the palmitoyl and oleoyl tails in POPE. Calculated deuterium order parameters are relatively close to the limited experimental data for the unsaturated oleoyl tail. The palmitoyl tail, however, is too ordered relative to experimental values.

Figure 3.8 shows area per lipid of POPC bilayer simulations. Two surface tensions produced POPC structures with experimental area per lipids close to published values: 17 dyne/cm and 20 dyne/cm. Because there are two published values of experimental area per lipid for POPC, both trajectories were analyzed at each surface tension value. The electron density profile of POPC at 17 dyne/cm and 20 dyne/cm are plotted and compared to an experimental electron density profile in figure 3.9. The electron density profile of each trajectory resembled the experimental density profile of POPC. The characteristic decrease in bilayer density at the center of the bilayer was not reproduced with the GAFF parameters. Because of this difference, it is likely that the POPC tails were too closely intertwined, with terminal methyl groups overlapping. Even with the differences in the bilayer interior, the calculated thickness of the bilayer was close to published experimental values. Order parameters for POPC are plotted in figure 3.10 and had mixed results. The oleoyl tail order parameters resembled the experimental order parameters and reproduced the double bond order parameters. However, the palmitoyl tail was much too highly ordered in both the 17 and 20 dyne/cm surface tension simulations. This was more evidence of serious issues with the GAFF saturated tail parameters. The high order parameters suggested that the palmitoyl tails were in a trans configuration that rigidly extend through the bilayer.

From the Lipid11 simulations, it was clear that the bilayer exhibited significant structural differences based on small adjustments to the surface tension. This exemplifies the sensitivity of bilayers to pressure coupling and constant surface ten-
sion. Given that structural issues persist with a constant surface tension term, it was more likely that the bilayer parameters were at issue rather than the constant surface tension simulation. Indeed further re-parameterization was able to reproduce bilayer structural properties in a tensionless ensemble.²

3.8 Implications for further parameterization

Lipid11 presents a modular framework for parameterization of glycerophospholipid bilayers.³ Residues, atom names, and atom types are clearly assigned for the set of molecules in Lipid11. The charge model is well-defined for the lipid residues and allows for many different combinations of lipid head and tail groups. This framework was established and released in the Amber 12 software package and allows for relatively simple system building and parameter assignment.¹⁰¹

Lipid11 includes a novel charge derivation and initial parameters from the GAFF parameter set. MD simulations of lipid bilayers with Lipid11 produced similar results to previous work with GAFF applied to lipids. As seen in preliminary GAFF lipid simulations, Lipid11 MD could not reproduce certain features of lipid bilayer structure. First, a constant surface tension term was necessary to maintain the box dimensions and area per lipid of the bilayer. Second, even with the constant surface tension constraint, the POPC electron density profile did not fully match an experimentally calculated density. Third, the order parameters of saturated lipid tails were too highly ordered, which significantly influences the bilayer structure and phase. These factors indicated that even with constant surface tension, the phase of the bilayer was not entirely correct.

Even with constant surface tension, basic lipid force field parameters based on GAFF were not appropriate for membrane simulations. At the time of Lipid11, other force fields were being developed for simulation in a tensionless ensemble.⁴⁹,²
Inspired by success in correcting lipid force fields for tensionless simulations, the Lipid11 framework would be used for further parameterization. In particular, the high tail order parameters suggested that the van der Waals and dihedral parameters for the saturated tails were not correct. Furthermore, given the importance of the ester region in controlling the flexibility and orientation of the tails, it was likely that further re-parameterization was necessary in that region. Chapter 4 explores the full re-parameterization of the lipid residues for tensionless bilayer simulations.4

3.9 Graphics processing units, lipid bilayers, and long-range interactions

Lipid11 and GAFFlipid lipid bilayers were simulated with the Amber MD program pmemd.cuda, which is GPU-accelerated. The GPU-accelerated MD software increases performance for lipid bilayer systems by orders of magnitude.1,55,57,56,58 One aspect of GPU acceleration had unintended side-effects for lipid bilayer simulations. This unintended side-effect and its correction are briefly described here. One of the optimizations that the accelerated MD engine takes is to spatially decompose the system into smaller “cells” that include nearby atoms for long-range interactions. Cells are defined by dividing up the system into an equally spaced 3-D array of cells. Each cell maintains a list of potential atomic interactions that includes atoms within the cell and within a certain distance of the cell. The additional distance beyond the cell is defined a “skin” that is used to determine the neighbor-list of pairwise interactions.

During a normal MD simulation with a barostat, the box dimensions of the system may change. In a normal MD simulation the cells would also change size. However, in the GPU-accelerated code, the cell size is calculated at the beginning
of a simulation, and assumed not to change significantly during the course of the
dynamics.

Initial lipid bilayers structures are often constructed \textit{de novo} from individual
lipid conformations. Because the bilayer structure is not well packed or equilibrated,
there is usually a large decrease in volume and increase in density of the bilayer
during the initial molecular dynamics. Sometimes after the initial volume decrease,
lipid simulations with Lipid11 and GAFFlipid on the GPU-accelerated version of
Amber would equilibrate to unusual structures.

When the system volume decreased beyond a certain amount, the GPU-
accelerated code began to skip atoms in the neighbor list. Missing atoms in the
neighbor list led to missing contributions to the electrostatics and van der Waals
forces. Effectively, the electrostatics and van der Waals forces were truncated. This
accumulated throughout the simulation and resulted in an incorrect molecular struc-
ture.

To correct this, the GPU accelerated MD code now monitors changes in box
dimensions. When the box dimensions have changed beyond the allowed value, the
MD code halts. The simulation may be restarted and the cells will be recalculated
for the new volume which prevents any missing interaction in the long-range force
calculations. This issue is an example of how subtle implementation decisions can
affect actual MD simulations. By the time Lipid14 was being developed, this issue
had been fixed, and it was verified that the GPU implementation matched the CPU
implementation during large volume changes.

\section{Acknowledgement}

Chapter 3 includes material from “Lipid11: A modular framework for lipid
simulations using Amber” Skjevik, Å. A., Madej, B. D., Walker, R. C., Teigen, K.
Journal of Physical Chemistry B. 2012. The dissertation/thesis author was a primary investigator and author of this paper.
Chapter 4

Lipid14

While Lipid11 established the framework for further lipid parameterization for an Amber lipid force field, the initial GAFF parameters were not sufficient to reproduce some bilayer structural features.\textsuperscript{48} Parallel to the work that established the modular framework for lipids, another force field was being developed for MD in Amber that would allow for tensionless bilayer simulations: GAFFlipid.\textsuperscript{2} GAFFlipid re-examined the van der Waals parameters for the lipid acyl chains and parameterized these terms independently of lipid bilayers. Then, bilayers were simulated with MD and compared to experimental structural data. The combination of the Lipid11 modular framework and GAFFlipid re-parameterization strategy provided a basis for the Lipid14 force field. Lipid14 improved upon the work of Lipid11 and GAFFlipid to create a robust, modular lipid force field for tensionless bilayer simulations.\textsuperscript{4}
**Figure 4.1:** A mixed Lipid14 bilayer. A top-down view of a membrane with DOPC and cholesterol with Lipid14 parameters. Each lipid molecule is colored differently.
4.1 Parameterization Strategy

4.1.1 Generation of Lipid14 parameters

Many parameters may contribute to overall lipid bilayer structure and properties. It was expected that lipid bond and angle terms would not need further parameterization because of the previous parameterization in GAFF. Previous lipid bilayer MD simulations suggest that the van der Waals and dihedral parameter terms are important for determining bilayer structure and dynamics.\textsuperscript{49} It was observed that the acyl tails with default GAFF parameters were too ordered for lipid simulations.\textsuperscript{2} In other force fields like Charmm C36 and Slipids, these parameters were refit.\textsuperscript{49,50}

Simulation of a box of 144 pentadecane molecules with the default GAFF van der Waals and dihedral terms at 298 K and 1 atm converges to a solid crystalline
phase with MD. The system density and enthalpy of vaporization are higher than experimental values and the diffusion of the molecules is lower.\textsuperscript{132} This suggests that the parameters developed for short acyl chains are not transferable to longer molecules. This is consistent with previous lipid bilayer MD in which the tails were too highly ordered.\textsuperscript{133} The van der Waals and dihedral terms are target parameters for fitting to density and enthalpy and vaporization. The van der Waals repulsive force largely determines the density of a solution. The enthalpy of vaporization is largely determined by nonbonded effects such as van der Waals and electrostatics. Changes in nonbonded parameters warrant a reinvestigation of the dihedrals because of the 1-4 nonbonded terms. Because the initial parameters deviated from experimental physical and thermodynamic values, the van der Waals and dihedral parameters were refit simultaneously.

The ester linkage parameters were re-examined as well. This region is important because it is where the tails are linked to the head group and can influence orientation and flexibility of the tails.\textsuperscript{2} As a model for this region, pure methyl acetate was simulated. The density and enthalpy of vaporization of methyl acetate was compared to experimental values.\textsuperscript{132} Then the van der Waals term was varied until it fit the experimental density and enthalpy of vaporization values.

Partial charges were re-derived for the molecules in a method based on Lipid11’s charge derivation.\textsuperscript{3} Lipid molecules were divided into three “residues” and charges were derived for each residue. Capping groups from Lipid11 were added to the residues and constrained during a RESP charge derivation in the same method as the Amber protein force fields.\textsuperscript{71,69}

It has been shown in previous parameterizations, however, that charge derivations in the method of RESP may be dependent on conformation.\textsuperscript{107} Some charge derivation strategies have used the average of multiple conformations for RESP charge
At a temperature above the melting point, the pure system was frozen. This indicates an issue with the long acyl chain parameters.\textsuperscript{64} Residue conformations were extracted from previous MD simulations of lipid bilayers and were therefore implicitly Boltzmann weighted. Otherwise, the charge derivation followed the same approach as the Lipid11 charge derivation.\textsuperscript{3}

With new van der Waals and partial charges for the ester linkage region, the dihedral angles of that region were re-examined. Changing the non-bonded parameters may influence related dihedrals. Therefore, two dihedral angles were re-parameterized to fit QM scans of the related dihedrals on a capped lauroyl molecule.

4.1.2 Tail parameters

The hydrocarbon tails are highly dependent on the van der Waals, dihedral, and partial charge parameters. The procedure for fitting these parameters is described in the following section.
Beginning with GAFF parameters as an initial starting point, first the basic CH₂-CH₂-CH₂-CH₂ torsion van der Waals and dihedral terms were simultaneously fit to experimental data. Model molecules including pentane, hexane, heptane, octane, decane, tridecane, and pentadecane were used for parameterization.⁶⁴

Van der Waals terms were adjusted in an incremental fashion while monitoring several quantitative system properties. Dihedral scans were performed on hexane and octane molecules and refit. Quantum energies were calculated by using an estimation of the CCSD(T)/cc-pVQZ level of theory with the HM-IE relation:¹³⁵

\[
E_{\text{conf}}^{\text{f}}[\text{CCSD(T)}/\text{LBS}] = E_{\text{conf}}^{\text{f}}[\text{CCSD(T)}/\text{SBS}] + (E_{\text{conf}}^{\text{f}}[\text{CCSD(T)}/\text{LBS}] - E_{\text{conf}}^{\text{f}}[\text{CCSD(T)}/\text{SBS}])
\]

\[
\approx E_{\text{conf}}^{\text{f}}[\text{CCSD(T)}/\text{SBS}] + (E_{\text{conf}}^{\text{f}}[\text{MP2}/\text{LBS}] - E_{\text{conf}}^{\text{f}}[\text{MP2}/\text{SBS}])
\]

\[
\equiv E_{\text{conf}}^{\text{f}}[\text{MP2:CC}] \quad (4.1)
\]

The small basis set (SBS) was cc-pVDZ and the large basis set was (LBS) was cc-pVQZ. Before the single-point energy calculations were performed with these basis sets, structures were optimized at the MP2/cc-pVDZ level. The paramfit program included in AmberTools 13 was used for a multiple molecule weighted torsion fit.¹ The tgt local minima of hexane and tgttt local minima were weighted at 10.0, and other local minima weighted at 4.0. Cis conformers were given a weighting of 0.1 and all other structures were weighted at 1. The weights used here have been previously used for alkane parameter fitting.¹³⁶ A genetic algorithm was used for dihedral parameter fitting.⁶⁸

Parameters were adjusted until MD simulations matched densities and heats
Table 4.1: Modification of LJ and Torsion Parameters of Alkane Chains

<table>
<thead>
<tr>
<th>Atom type</th>
<th>Radius (Å)</th>
<th>Well-depth (kcal/mol)</th>
<th>Force constant (kcal/mol)</th>
<th>Periodicity</th>
<th>Phase (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipid11</td>
<td>cA</td>
<td>1.9080</td>
<td>0.1094</td>
<td>0.20</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>hA</td>
<td>1.4870</td>
<td>0.0157</td>
<td>0.25</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Lipid14</td>
<td>cD</td>
<td>1.9080</td>
<td>0.1094</td>
<td>0.3112</td>
</tr>
<tr>
<td></td>
<td>hL</td>
<td>1.4600</td>
<td>0.0100</td>
<td>-0.1233</td>
<td>0.1149</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.2199</td>
<td>0.2170</td>
</tr>
</tbody>
</table>

Densities and enthalpies of vaporization are listed in table 4.1.

Partial charges were derived using the Amber RESP method. Amber force fields have used an optimization of the molecule and calculation of a restrained ESP with the HF/6-31G* level of theory in gas phase. The new charges, van der Waals, and dihedral terms were then used in MD simulations of liquid hydrocarbons. The system was kept at 298 K for 10 ns and structures were randomly selected from the trajectory. The ESP was calculated for each extracted structure at the HF/6-31G* level of theory and partial charges were estimated by the RESP method.

144 and 288 hydrocarbons in the NPT ensemble were simulated with MD. The temperature was controlled with the Langevin thermostat at 298 K and a collision frequency of 5 ps⁻¹. Pressure was maintained with an isotropic Berendsen barostat, and a pressure relaxation time of 1 ps. Long range electrostatics were calculated with the particle mesh Ewald method using a direct space cutoff of 10 Å which was also the cutoff for van der Waals. Bonds with hydrogen were constrained with the SHAKE algorithm. The system was heated from 0 to 298 K over 20 ps with a weak harmonic restraint on all molecules. Three more equilibrations were simulated with a decreasing restraint. The restraint started at 20.0 kcal/mol/Å² and was then
decreased to 10.0, 5.0, and 1.0 for each following simulation segment.

The heat of vaporization may be estimated from MD simulations as

\[ \Delta H_{vap} = E_{pot}(g) - E_{pot}(l) + RT \]  

where \( E_{pot} \) is the potential energy, \( g \) refers to gas phase, \( l \) refers to liquid phase and \( RT \) is the gas constant and temperature. In order to calculate \( E_{pot} \) for both phases, the molecules were simulated once in each phase. The van der Waals radius and well-depth were decreased and the torsions were refit such that MD matched experimental density and enthalpy of vaporization.

In order to examine the nature of unsaturated lipid tails, the dihedrals of a \( cis \)-5-decene molecule were re-examined. The dihedral was scanned and refit using the estimation of MP2:CC theory previously described in equation 4.1.\textsuperscript{135} The van der Waals parameters of the alkene hydrogen atom type \( hB \) in model molecules \( cis \)-2-hexene, \( cis \)-5-decene, and \( cis \)-7-pentadecene were refit. Adjusting the well depth and radius of this atom type resulted in MD simulations of these molecules that reproduced experimental densities. The heat of vaporization was adjusted as well, but just a change in van der Waals parameters did not yield complete agreement with experiment. This has been observed in other lipid force fields as well.\textsuperscript{51,50}

The diffusion of the alkanes may be calculated from a trajectory from the slope of the mean squared displacement (MSD) plot versus time of the centers of mass. Diffusion MD was simulated in the NVE ensemble for 100 ns. In order to sufficiently minimize energy and temperature drift it was necessary to remove the system center of mass movement every 500 time steps, make the SHAKE and Ewald direct sum tolerances smaller, and decrease the time step to 1 fs. Particle mesh Ewald was used with a direct space cutoff of 10 Å.\textsuperscript{78} The slope of the alkane MSD was used for the diffusion coefficient calculation. The diffusion coefficient calculated
with periodic boundary conditions is defined as

\[ D_{PBC} = \lim_{t \to \infty} \frac{\langle (\Delta r(t))^2 \rangle}{2n_f t} \]  

(4.3)

where \( n_f \) is the number of dimensions of diffusion (three in this case), and \( \Delta r(t) \) is the distance traveled in time \( t \). The diffusion with periodic boundary conditions may be corrected with an additional term proposed by Yeh.\(^{137}\)

\[ D_{corr} = D_{PBC} + \frac{k_B T \sigma}{2n_f \pi \eta L} \]  

(4.4)

where \( k_B \) is Boltzmann’s constant, \( T \) is the temperature, \( \sigma = 2.837297 \) (the dominant contribution comes from the compensating background term in the Stokes equation), \( \eta \) is the viscosity, and \( L \) is the simulation box dimension. Experimental viscosity values were used for calculation of the corrected diffusion value.

The population of \textit{trans}, end \textit{gauche}, double \textit{gauche}, and kinked \textit{gtg’+gtg} conformers were calculated from the previous NPT simulations. Dihedral angles were classified geometrically: \textit{gauche} plus as 0-120°, \textit{trans} as 120-240°, and \textit{gauche} minus as 240-360°.

Experimental \(^{13}\)C NMR \( T_1 \) relaxation times are available for selected alkane chains.\(^{138}\) The reorientation correlation of the CH bond vectors can be used to calculate the dipolar relaxation from MD simulations with the following equation:

\[ \frac{1}{NT_1} = (1.855 \times 10^{10} \text{s}^{-1}) \int_0^\infty \langle P_2(\hat{\mu}(0)\hat{\mu}(t)) \rangle dt \]  

(4.5)

The dipolar relaxation calculation assumes that the C-H bond length is 1.117 Å which is attributed to a decrease in the corresponding dipolar couplings resulting from fast angular fluctuations. \( N \) is the number of protons bonded to the carbon and \( \hat{\mu} \) is the CH vector.\(^{139}\) Heptane, decane, tridecane, and pentadecane were simulated
Table 4.2: Thermodynamic properties of methyl acetate simulated with GAFF/Lipid11 and Lipid14 in comparison with experiment. All values at 298.15 K.

<table>
<thead>
<tr>
<th></th>
<th>Atom type</th>
<th>Radius (Å)</th>
<th>Well (kcal/mol)</th>
<th>ΔH_{vap} (kJ/mol)</th>
<th>ρ (kg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipid11</td>
<td>oC</td>
<td>1.6612</td>
<td>0.210</td>
<td>39.11 ± 0.04</td>
<td>928.38 ± 0.09</td>
</tr>
<tr>
<td></td>
<td>oS</td>
<td>1.6837</td>
<td>0.170</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cC</td>
<td>1.9080</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>oC</td>
<td>1.6500</td>
<td>0.140</td>
<td>33.0 ± 0.07</td>
<td>925.8 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>oS</td>
<td>1.6500</td>
<td>0.120</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cC</td>
<td>1.9080</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expt</td>
<td></td>
<td></td>
<td></td>
<td>32.29¹³²</td>
<td>934.2¹³²</td>
</tr>
</tbody>
</table>

for 200 ns in the NPT ensemble.

4.1.3 Head parameters

The ester linkage on the head group that connects to the tail group was re-examined. The enthalpy of vaporization of a model molecule, methyl acetate, with the default GAFF parameters of the ester linkage, methyl acetate was simulated and compared with experimental values.¹³² The van der Waals parameters for the carbonyl oxygen oC, carbonyl carbon cC, and ester oxygen oS were adjusted so that simulations more closely matched the enthalpy of vaporization. Fit parameters are listed in table 4.2.

The enthalpy of vaporization of methyl acetate was calculated in a similar way to the tail enthalpy of vaporization. A box of 288 methyl acetate molecules in the liquid phase was used for 60 ns of simulation.

4.1.4 Partial charges

The partial charge derivation was very similar to the Lipid11 charge derivation.³ Charges were derived using the RESP procedure on capped lipid residues. The major difference between the Lipid11 and that used for Lipid14 charge derivation is
Figure 4.4: Structure and partial charges for Lipid11 and Lipid14 capping groups.\textsuperscript{3,4} The total charge of the molecule was constrained to 0.0 e and the total charge of each boxed unit was constrained to 0.0 e.

![Diagram of Lipid11 and Lipid14 capping groups]

Figure 4.5: A capped lauroyl molecule for fitting dihedrals. The two dihedrals indicated in the structure were refit.

The current charge derivation uses the average partial charges from multiple conformations.\textsuperscript{4}

Twenty-five conformations of each head group and fifty conformations of each tail group were averaged for the charge derivation. Conformations were extracted from previous molecular dynamics simulations and were capped with the Lipid11 capping group. Each conformation’s ESP was calculated using Gaussian 2009.\textsuperscript{140} With the average across multiple conformations, the charges were implicitly Boltzmann weighted. Charges were derived for the head groups phosphatidylcholine (PC), and phosphatidylethanolamine (PE). Charges were also derived for the tail groups lauroyl (LA), myristoyl (MY), palmitoyl (PA), and oleoyl (OL). Charges are included in the supporting material of Lipid14.\textsuperscript{4}
Table 4.3: Thermodynamic and dynamic properties of a selection of alkane chaines simulated using the updated Lipid14 parameters and comparison to experiment. All values at 298.15 K.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta H_{\text{vap}}$ (kJ/mol)</th>
<th>$\rho$ (kg/m$^3$)</th>
<th>$D_{\text{PBC}}$ ($10^{-5}$ cm$^2$/s)</th>
<th>$D_{\text{corr}}$ ($10^{-5}$ cm$^2$/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentane</td>
<td>Lipid14: 23.03 ± 0.16</td>
<td>592.45 ± 0.16</td>
<td>6.45 ± 0.56</td>
<td>7.1 ± 0.56</td>
</tr>
<tr>
<td></td>
<td>Expt: 26.43$^{132}$</td>
<td>626.2$^{132}$</td>
<td></td>
<td>5.45$^{141}$</td>
</tr>
<tr>
<td>Hexane</td>
<td>Lipid14: 28.54 ± 0.1</td>
<td>636.3 ± 0.09</td>
<td>4.55 ± 0.29</td>
<td>5.02 ± 0.29</td>
</tr>
<tr>
<td></td>
<td>Expt: 31.56$^{132}$</td>
<td>656.1$^{132}$</td>
<td></td>
<td>4.21$^{141}$</td>
</tr>
<tr>
<td>Heptane</td>
<td>Lipid14: 33.37 ± 0.11</td>
<td>667.31 ± 0.14</td>
<td>3.47 ± 0.23</td>
<td>3.85 ± 0.23</td>
</tr>
<tr>
<td></td>
<td>Expt: 36.57$^{132}$</td>
<td>679.5$^{132}$</td>
<td></td>
<td>3.12$^{141}$</td>
</tr>
<tr>
<td>Octane</td>
<td>Lipid14: 38.67 ± 0.31</td>
<td>690.96 ± 0.10</td>
<td>2.11 ± 0.15</td>
<td>2.46 ± 0.15</td>
</tr>
<tr>
<td></td>
<td>Expt: 41.49$^{132}$</td>
<td>698.6$^{132}$</td>
<td></td>
<td>2.354$^{143}$</td>
</tr>
<tr>
<td>Decane</td>
<td>Lipid14: 49.34 ± 0.30</td>
<td>724.47 ± 0.07</td>
<td>1.44 ± 0.15</td>
<td>1.65 ± 0.15</td>
</tr>
<tr>
<td></td>
<td>Expt: 51.42$^{132}$</td>
<td>726.6$^{132}$</td>
<td></td>
<td>1.39$^{143}$</td>
</tr>
<tr>
<td>Tridecane</td>
<td>Lipid14: 64.62 ± 0.27</td>
<td>756.19 ± 0.24</td>
<td>0.48 ± 0.04</td>
<td>0.57 ± 0.04</td>
</tr>
<tr>
<td></td>
<td>Expt: 66.68$^{132}$</td>
<td>756.4$^{132}$</td>
<td></td>
<td>0.712$^{143}$</td>
</tr>
<tr>
<td>Pentadecane</td>
<td>Lipid14: 74.99 ± 0.39</td>
<td>770.67 ± 0.25</td>
<td>0.30 ± 0.02</td>
<td>0.36 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>Expt: 76.77$^{132}$</td>
<td>768.5$^{132}$</td>
<td></td>
<td>0.461$^{143}$</td>
</tr>
</tbody>
</table>

4.1.5 Head group dihedral fits

Two dihedrals related to the ester group in the head group were refit: oS-cC-cD-cD and oC-cC-cD-cD. A capped lauroyl group was used for refinement of the dihedrals. The dihedrals were scanned in the same way as described for the tail dihedrals with the HM-IE relation.$^{133}$ Five term dihedrals were fit with Paramfit using the genetic algorithm.$^{68}$
Figure 4.6: Energy profiles for torsions from a *cis*-5-decene molecule. Dotted lines represent torsion energies before fitting, triangles represent quantum mechanical energies, and solid lines represent final fit torsions.
Table 4.4: Thermodynamic properties of a selection of alkene chains simulated using the updated Lipid14 parameters and comparison to experiment

<table>
<thead>
<tr>
<th></th>
<th>$\Delta H_{\text{vap}}$ (kJ/mol)</th>
<th>$\rho$ (kg/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cis-2-Hexene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>$26.17 \pm 0.21$</td>
<td>$656.23 \pm 0.13$</td>
</tr>
<tr>
<td>Expt</td>
<td>$32.19^{132}$</td>
<td>$683^{142}$</td>
</tr>
<tr>
<td>cis-5-Decene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>$45.27 \pm 0.22$</td>
<td>$739.19 \pm 0.16$</td>
</tr>
<tr>
<td>Expt</td>
<td>$42.9^{132}$</td>
<td>$744.5^{132}$</td>
</tr>
<tr>
<td>cis-7-Pentadecene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>$69.5 \pm 0.22$</td>
<td>$781.44 \pm 0.24$</td>
</tr>
<tr>
<td>Expt</td>
<td></td>
<td>$775^{142}$</td>
</tr>
</tbody>
</table>

4.2 Parameterization

4.2.1 Hydrocarbon parameters

Thermodynamic and dynamic properties of the different length alkanes with the new parameters are shown in table 4.3. Experimental enthalpy of vaporization increases with length of the simulated hydrocarbon.$^{132}$ This trend was captured with the new parameters and heats of vaporization reproduced experimental values with an RMS error of 7.67%. The densities also increase with the length of the hydrocarbon chain. MD simulations match the experimental values with an RMS error of 2.60%. Diffusion of the alkanes decreases with increasing alkane chain length.$^{141}$ This trend was observed in the MD, but with a higher RMS error of 20.92%. Better agreement with experimental diffusion is found with longer alkane chains, while shorter chains show less agreement. This difference may be acceptable in lipid simulations, as longer chains are generally found in glycerophospholipids.

MD simulations of pure alkenes are compared with experiment in table 4.4.$^{132,142}$ New parameters do not perfectly match the experimental enthalpy of vaporization, but are similar with a RMS error of 13.80%. This may be in part due to the inability of just van der Waals parameterization to reproduce these values as the double bonds
Table 4.5: Average number of trans, gauche, end gauche (eg), double gauche (gg), and gtg'+gtg conformers per alkane molecule and comparison to experiment

<table>
<thead>
<tr>
<th></th>
<th>trans</th>
<th>gauche</th>
<th>t/g ratio</th>
<th>eq</th>
<th>gg</th>
<th>gtg'+gtg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentane</td>
<td>Lipid14</td>
<td>1.20</td>
<td>0.80</td>
<td>1.49</td>
<td>0.80</td>
<td>0.13</td>
</tr>
<tr>
<td>Hexane</td>
<td>Lipid14</td>
<td>1.83</td>
<td>1.17</td>
<td>1.57</td>
<td>0.83</td>
<td>0.22</td>
</tr>
<tr>
<td>Heptane</td>
<td>Lipid14</td>
<td>2.49</td>
<td>1.51</td>
<td>1.65</td>
<td>0.83</td>
<td>0.31</td>
</tr>
<tr>
<td>Octane</td>
<td>Lipid14</td>
<td>3.15</td>
<td>1.85</td>
<td>1.71</td>
<td>0.81</td>
<td>0.39</td>
</tr>
<tr>
<td>Decane</td>
<td>Lipid14</td>
<td>4.47</td>
<td>2.53</td>
<td>1.77</td>
<td>0.81</td>
<td>0.57</td>
</tr>
<tr>
<td>Tridecane</td>
<td>Lipid14</td>
<td>6.47</td>
<td>3.53</td>
<td>1.84</td>
<td>0.81</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Expt$^{144}$</td>
<td>6.50</td>
<td>3.50</td>
<td>1.86</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>Pentadecane</td>
<td>Lipid14</td>
<td>7.80</td>
<td>4.20</td>
<td>1.86</td>
<td>0.81</td>
<td>1.00</td>
</tr>
</tbody>
</table>

in the alkenes are highly charged. The densities were close to experiment with an RMS error of 2.35%.

The populations of trans, gauche, end gauche, double gauche, and kinked gtg'+gtg conformers was calculated from the MD simulations. Simulation populations are printed in table 4.5. Limited data on these populations is available, including only FTIR values for tridecane.$^{144}$ The new parameters match the experimental conformation populations well. For tridecane and the other alkanes, these values suggest that the problems with the overabundance of trans conformations with GAFF parameters were alleviated with this refit.$^4$

The $^{13}$C NMR $T_1$ relaxation times were calculated for four of the simulated alkanes and compared to experiment. The $T_1$ relaxation time is a dynamic property related to the rotation of CH vectors.$^{139}$ Relaxation times calculated from MD of the four alkanes compare well with experimental values, with only some deviation in pentadecane.$^{138}$
4.2.2 Head group parameters

For the head group, the change in van der Waals and dihedrals in the head groups led to better agreement with experimental properties.

For van der Waals refitting, the radius and well depth was decreased, and consequently the enthalpy of vaporization decreased to values closer to experiment. At the same time, the system density remained approximately the same and relatively close to experiment.

Dihedrals were refit due to overly large rotational energetic barriers. The fit dihedrals match quantum energies much closer than the previous parameter set and removed the large energetic barriers for rotation.

4.3 Lipid bilayer simulations

4.3.1 Generation of structures

Initial structures were created using the Charmm-GUI web site. Structures were solvated at a full hydration level and included 150 mM KCl ionic concentration.
Figure 4.8: Energy profiles for selected torsions from lauroyl molecule. Energy from initial parameters are plotted with dotted lines, quantum mechanics energies are marked with filled triangles, and final parameters plotted with a solid line.

Table 4.6: Lipid14 lipid bilayer molecular dynamics. System, size $n_{lipid}$, hydration $n_w$, temperature $T$, and simulation time for lipid bilayer systems

<table>
<thead>
<tr>
<th>Lipid</th>
<th>$n_{lipid}$</th>
<th>Sim. Time (ns)</th>
<th>$T$ (K)</th>
<th>$n_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLPC</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>303</td>
<td>31.3</td>
</tr>
<tr>
<td>DMPC</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DPPC</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>323</td>
<td>30.1</td>
</tr>
<tr>
<td>DOPC</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>POPC</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>303</td>
<td>31</td>
</tr>
<tr>
<td>POPE</td>
<td>128</td>
<td>$5 \times 125$</td>
<td>310</td>
<td>32</td>
</tr>
</tbody>
</table>
Figure 4.9: Six primary lipid types simulated in bilayers. Phospholipid types are defined in the text. This represents a combination of two head groups with many different tails.

The TIP3P water model was used for water and compatible ionic parameters were applied. MD simulation parameters are listed in table 4.6. Dilauroylphosphatidylcholine (DLPC), dimyristoylphosphatidylcholine (DMPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylcholine (DOPC), palmitoyloleoylphosphatidylcholine (POPC), and palmitoyloleoylphosphatidylethanolamine (POPE) bilayers were simulated. The structure of each phospholipid is shown in figure 4.9.

### 4.3.2 MD protocol

The combined system was minimized for 10000 steps, with 5000 steps of steepest descent and 5000 steps of conjugate gradient.

The system was then heated for 5 ps from 0 K to 100 K. The Langevin thermostat and constant volume was used for MD simulations with a collision frequency of 1.0 ps$^{-1}$. Lipid molecules were restrained with a force constant of 10 kcal mol$^{-1}$ Å$^{-2}$. 
The system was then heated to a temperature above each lipid’s phase transition temperature. In general, temperatures were chosen to match available experimental structural data. The Langevin thermostat was used for MD now with an anisotropic Berendsen barostat at 1 atm with a time constant of 2 ps.\textsuperscript{98,94} The second heating phase was simulated for 100 ps with the same restraint on the lipids. Because the lipid bilayer box dimensions often decrease significantly at the beginning of a MD simulation, it was necessary to both increase the non-bonded skin distance and restart the simulation several times at the beginning of MD production.

Lipid bilayers were simulated in the constant pressure and temperature ensemble using the Amber 12 MD software package.\textsuperscript{101} The GPU-accelerated MD program \textit{pmemd.cuda} was used for simulations.\textsuperscript{1,55,56,57,58} Bonds involving hydrogen were constrained using the SHAKE algorithm with a 2 fs timestep.\textsuperscript{92} PME was used for electrostatic calculations with a direct space cutoff of 10 Å.\textsuperscript{78} The same cutoff was applied to van der Waals interactions with a long range correction applied beyond the cutoff.\textsuperscript{59} Temperature was controlled with the Langevin thermostat at a system-dependent temperature and a collision frequency of 1.0 ps\textsuperscript{-1}.\textsuperscript{98} Pressure coupling was controlled via a Berendsen barostat at 1 atm and a pressure relaxation time of 1.0 ps.\textsuperscript{94}

Each initial structure was repeated five times with randomized initial velocities. Each repeat was run for at least 125 ns, with the equilibrated trajectory analyzed. In total, 3 μs of MD were sampled. Most analysis was conducted with \textit{ptraj} or \textit{cpptraj} and modifications made by Loeffler.\textsuperscript{145,146} Convergence was examined by monitoring system density and lipid bilayer area per lipid as a function of time.

Stability was tested by extending simulations to 250 ns without significant change. Further, a set of GPU runs were replicated on CPUs and all bilayers were consistent with CPU simulations.\textsuperscript{4}
4.4 Validation

4.4.1 Bilayer structural properties

The lipid bilayer area per lipid is described in section 3.6.1. The area per lipid is simple to calculate and may give an indication of the bilayer phase. Area per lipid was calculated in a consistent way with Lipid11 (section 3.6.1). All average bilayer structural properties from simulation and experiment are printed in table 4.7. Area per lipid for all bilayers was within 3% of the experimental values, indicating that the bilayers are in the $L_\alpha$ liquid disordered phase. The POPE area per lipid is closer to the published value of 56.5 Å$^2$ than another published value of 59-60 Å$^2$.

4.4.2 Volume per lipid

The volume per lipid $V_L$ is more precisely measured than area per lipid in experiments and is a better comparison to MD. Volume per lipid may be calculated from MD via

\[ V_L = \frac{L_x L_y L_z}{N_L} - n_{\text{wat}} \langle V_{\text{wat}} \rangle \]  

(4.6)

where $L_x$, $L_y$, $L_z$ are the periodic box dimensions, $N_L$ is the number of lipids, $n_{\text{wat}}$ is the number of waters, and $\langle V_{\text{wat}} \rangle$ is the average water volume. The average volume of water was estimated by simulating a box of 1936 neat TIP3P water molecules under the same protocol as the lipid bilayers at the corresponding temperature.

Volume per lipid values are reported and underestimate the experimental values overall. However, the simulation values as a whole are within 5% of the experimental values.\textsuperscript{15,12,157,120}

Because the van der Waals of the tail groups have been carefully fit to exper-
Table 4.7: Average structural properties over five repeats of the six lipid systems simulated with Lipid14 and comparison to experiment. Area per lipid $A_L$, volume per lipid $V_L$, isothermal area compressibility modules $K_A$, bilayer thickness $D_{HH}$, bilayer Luzzati thickness $D_B$, $\Delta D_B - H$ (see text), ratio $r$ of terminal methyl to methylene.

<table>
<thead>
<tr>
<th>Lipid System</th>
<th>$A_L$ ($\text{Å}^2$)</th>
<th>$V_L$ ($\text{Å}^3$)</th>
<th>$K_A$ (mN/m)</th>
<th>$D_{HH}$ (Å)</th>
<th>$D_B$ (Å)</th>
<th>$\Delta D_B - H$ (Å)</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>63.0 ± 0.2</td>
<td>948.9 ± 0.3</td>
<td>381 ± 37</td>
<td>30.4 ± 0.4</td>
<td>30.2 ± 0.1</td>
<td>-0.1 ± 0.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Expt</td>
<td>63.2,15 60.8,127</td>
<td>991,15</td>
<td>-</td>
<td>30.8,15</td>
<td>31.4,15</td>
<td>0.8,147</td>
<td>1.8-2.1,148</td>
</tr>
<tr>
<td>DMPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>59.7 ± 0.7</td>
<td>1050.2 ± 1.5</td>
<td>264 ± 90</td>
<td>34.7 ± 0.6</td>
<td>35.2 ± 0.4</td>
<td>0.3 ± 0.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Expt</td>
<td>60.6,15 59.9,127</td>
<td>1101,12,15</td>
<td>234,149</td>
<td>34.4,150 35.5,100</td>
<td>36.3,15 36.7,127</td>
<td>36.9,150</td>
<td>0.8,147</td>
</tr>
<tr>
<td>DPPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>62.0 ± 0.3</td>
<td>1177.3 ± 0.4</td>
<td>244 ± 50</td>
<td>37.9 ± 0.5</td>
<td>38.0 ± 0.2</td>
<td>0.1 ± 0.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Expt</td>
<td>63.1,127 64.3,151</td>
<td>1232,12</td>
<td>321,12</td>
<td>38.16 38.3,12</td>
<td>39.0,127,16</td>
<td>0.8,147</td>
<td>1.8-2.1,148</td>
</tr>
<tr>
<td>DOPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>69.0 ± 0.3</td>
<td>1249.6 ± 0.2</td>
<td>338 ± 31</td>
<td>37.0 ± 0.2</td>
<td>36.2 ± 0.2</td>
<td>-0.4 ± 0.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Expt</td>
<td>67.4,16 72.5,12</td>
<td>1303,12</td>
<td>338,149 300,152</td>
<td>35.3,153 36.7,16,154</td>
<td>35.9,12 36.1,153,155</td>
<td>38.7,16</td>
<td>1.0-1.7,148</td>
</tr>
<tr>
<td>POPC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>65.6 ± 0.5</td>
<td>1205.4 ± 0.4</td>
<td>257 ± 47</td>
<td>36.9 ± 0.6</td>
<td>36.8 ± 0.3</td>
<td>-0.1 ± 0.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Expt</td>
<td>64.3,127 68.3,157</td>
<td>1256,157</td>
<td>180-330,158</td>
<td>37,157</td>
<td>36.8,157 39.1,127</td>
<td>0.8,147</td>
<td>1.8-2.1,148</td>
</tr>
<tr>
<td>POPE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipid14</td>
<td>55.5 ± 0.2</td>
<td>1138.7 ± 0.3</td>
<td>350 ± 81</td>
<td>42.4 ± 0.2</td>
<td>41.0 ± 0.1</td>
<td>-0.7 ± 0.1</td>
<td>2</td>
</tr>
<tr>
<td>Expt</td>
<td>56.6,159 59-60,120</td>
<td>1180,120</td>
<td>233,159</td>
<td>39.5,120</td>
<td>-</td>
<td>-</td>
<td>1.8-2.1,148</td>
</tr>
</tbody>
</table>
imental densities, it is possible that the head group van der Waals may be fit further to correct the volume per lipid value. The program SIMtoEXP allows for calculation of volumes of components in the bilayer and was used to compare head group and tail groups. The head group volume for DOPC was 305.41 Å³ while the experimental estimate is 319-331 Å³ and the tail group volume was found to be 965.88 Å³ while experimental estimates are 972-984 Å³. Further, SIMtoEXP was used to calculate the ratio of terminal methylene to methylene volume \( r \). The reported bilayers here have a ratio \( r \) of 1.85-2.17 close to the experimental range of 1.8-2.1.

### 4.4.3 Isothermal area compressibility modulus

Isothermal area compressibility modulus \( K_A \) is related to the change in the area per lipid and can be calculated as

\[
K_A = \frac{2k_B T \langle A_L \rangle}{n_{lipid} \sigma_A^2}
\]

where \( k_B \) is Boltzmann’s constant, \( T \) is the temperature, \( A_L \) is the area per lipid, \( n_{lipid} \) is the number of lipids, and \( \sigma_A^2 \) is variance in the area per lipid. Overall, area compressibility modulus values are close to the reported experimental values. Experimental values fall within one standard deviation of the MD simulations, but the standard deviations are large. The DOPC isothermal compressibility area was 338 ± 31 from MD while one experimental value was reported to be 300. Through a personal communication with E. Evans, it was revealed that a new \( K_A \) value for DOPC was 318 mN/m. Note that DOPC was simulated and analyzed before the new experimental DOPC \( K_A \) was published.

The \( K_A \) values depend on the fluctuations of system dimensions and therefore the barostat. Amber 12 MD only supported the Berendsen barostat, which has been suggested to be not ideal for MD where volume fluctuations are significant.
Therefore, $K_A$ values may depend on the Berendsen barostat. The Monte Carlo barostat has been implemented in Amber, and other barostats may be added, but at the time of writing, the Berendsen barostat was the only option for pressure coupling. Other factors that may influence $K_A$ include system size and simulation time. Larger systems and longer simulations may result in more converged $K_A$ values.

4.4.4 Bilayer thickness

Bilayer thickness was calculated from simulation density values. $D_{HH}$ is defined as the peak to peak distance from the electron density corresponding to the distance between phosphate groups in the two leaflets. Good agreement was achieved for all bilayer types, except for POPE, which was thicker than experiment.

The Luzzati thickness $D_B$ is another reported bilayer thickness that may be calculated from the integral of the water density. $D_B$ values from MD are close to experimental values. For saturated lipid bilayers, the thickness is lower than experiment, indicating that water may be penetrating too far into the bilayer.

The Luzzatti thickness is an alternate thickness calculation defined as:

$$D_B = 2V_L/A_L$$

(4.8)

$V_L$ is defined as the molecular volume of the lipid. Equivalently,

$$D_B = L_Z - D_W$$

(4.9)

where $D_W$ is the water thickness and $L_Z$ is the periodic box dimension normal to the bilayer (In this case, the height of the system). In order to evaluate Luzzatti thickness from a trajectory, it is convenient to use the probability density $\rho_W$ of water through
Figure 4.10: Lipid bilayer tail order parameters for six lipid bilayer types with comparison to experiment.$^{131,163,128,164,165,166,167}$

The deuterium order parameter has been previously described in section 3.6.3. These geometric measurements of the CD vectors in the lipid acyl chains may be determined from $^2$H NMR or $^1$H-$^{13}$C NMR.$^{13}$ $S_{CD}$ is a function of the angle $\theta$ of CD vectors relative the bilayer normal.

Simulations of lipid bilayers with Lipid14 parameters were analyzed to calculate the order parameters for each tail. All the lipid bilayer types follow the experimental order parameter trends. The first carbon of the sn-1 and sn-2 tails are

$$D_B = L_Z - \int_{-L_Z/2}^{L_Z/2} \rho_W(z)dz$$  \hspace{1cm} (4.10)

4.4.5 Lipid acyl chain order parameters

The deuterium order parameter has been previously described in section 3.6.3. These geometric measurements of the CD vectors in the lipid acyl chains may be determined from $^2$H NMR or $^1$H-$^{13}$C NMR.$^{13}$ $S_{CD}$ is a function of the angle $\theta$ of CD vectors relative the bilayer normal.

Simulations of lipid bilayers with Lipid14 parameters were analyzed to calculate the order parameters for each tail. All the lipid bilayer types follow the experimental order parameter trends. The first carbon of the sn-1 and sn-2 tails are
markedly different due to the difference in orientation of these two groups. The $S_{CD}$ value for the first carbons in the $sn$-1 chains are in general higher than the $sn$-2 tails.

Overall, each tail residue fits the experimental deuterium order parameters well. However, the palmitoyl tail in POPE is slightly too ordered, which correlates with the larger $D_{HH}$ value. For oleoyl tails, the 9 and 10 carbon display the characteristic drop in order for double bonds. Tails in POPC and POPE are clearly differentiated, with the palmitoyl showing higher order than the oleoyl.

Furthermore, the acyl chain conformation can be classified by rotamer and compared with experimental Fourier transform infrared spectroscopy (FTIR). FTIR has been used to count the number of trans ($t$) and gauche ($g$) conformers as well as rotameric sequences. Dihedral angles $\phi$ of the acyl chain were calculated and classified as $t$ for $\phi < -150^\circ$ or $\phi > 150^\circ$, $g-$ for $-90^\circ < \phi < -30^\circ$ or $g+$ for $30^\circ < \phi < 90^\circ$. The rotameric sequences $gtg$ corresponds to $g^+tg^+$ or $g^-tg^-$ while the kink sequence $gtg'$ corresponds to $g^+tg^-$ or $g^-tg^+$.

Rotamer fraction analysis is presented in table 4.8. In general, MD simulations reproduce experimental rotamer values. The differences for DLPC and DPPC $gtg'$ may be in part from ambiguity in assigning $gtg$ and $gtg'$ wagging modes. These rotamer populations seem to indicate that the bilayers are in the correct phase: an increase in the number of gauche rotamers and kink rotamer sequences is associated with the transition to liquid phase. The populations for POPE and POPC end gauche ($eg$) and gauche-gauche ($gg$) are similar. Senak reports a significant increase in $eg$, $gg$, and $gtg'+gtg$ values from DPPE to DPPC, which contradicts the populations found in POPE and POPC simulations.
Table 4.8: Analysis of rotamers and rotamer sequences in the acyl chains of the six lipid systems. End gauche (eq), double gauche (gg), kinks (g'g), g'g’+g, and number of gauche (ng). *The g'g’ sequence may be ascribed to a g’g’+g sequence.\textsuperscript{168}

<table>
<thead>
<tr>
<th>Lipid system</th>
<th>eg</th>
<th>gg</th>
<th>g'g’</th>
<th>g’g’+g</th>
<th>ng</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLPC</td>
<td>0.35</td>
<td>0.44</td>
<td>0.28</td>
<td>0.52</td>
<td>2.50</td>
</tr>
<tr>
<td>Lipid14</td>
<td>0.45</td>
<td>0.32</td>
<td>0.88*</td>
<td>-</td>
<td>2.85</td>
</tr>
<tr>
<td>Expt\textsuperscript{169}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMPC</td>
<td>0.34</td>
<td>0.48</td>
<td>0.35</td>
<td>0.62</td>
<td>2.82</td>
</tr>
<tr>
<td>Lipid14</td>
<td>0.38</td>
<td>0.67</td>
<td>-</td>
<td>0.44</td>
<td>2.60</td>
</tr>
<tr>
<td>Expt\textsuperscript{170}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DPPC</td>
<td>0.36</td>
<td>0.66</td>
<td>0.47</td>
<td>0.83</td>
<td>3.58</td>
</tr>
<tr>
<td>Lipid14</td>
<td>0.38, \textsuperscript{170} 0.4, \textsuperscript{171}, 0.54\textsuperscript{169} 0.4, \textsuperscript{171,169} 0.57\textsuperscript{170} 1.19\textsuperscript{169} 0.46, \textsuperscript{170} 1.0\textsuperscript{171} 2.44, \textsuperscript{170} 3.6-4.2, \textsuperscript{172} 3.7, \textsuperscript{169} 3.8\textsuperscript{165}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expt</td>
<td>0.38, \textsuperscript{170} 0.4, \textsuperscript{171}, 0.54\textsuperscript{169} 0.4, \textsuperscript{171,169} 0.57\textsuperscript{170} 1.19\textsuperscript{169} 0.46, \textsuperscript{170} 1.0\textsuperscript{171} 2.44, \textsuperscript{170} 3.6-4.2, \textsuperscript{172} 3.7, \textsuperscript{169} 3.8\textsuperscript{165}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOPC</td>
<td>0.36</td>
<td>0.75</td>
<td>0.37</td>
<td>0.70</td>
<td>3.93</td>
</tr>
<tr>
<td>POPC</td>
<td>0.36</td>
<td>0.69</td>
<td>0.41</td>
<td>0.75</td>
<td>3.73</td>
</tr>
<tr>
<td>POPE</td>
<td>0.35</td>
<td>0.60</td>
<td>0.42</td>
<td>0.73</td>
<td>3.50</td>
</tr>
<tr>
<td>Lipid14</td>
<td>0.05</td>
<td>0.20</td>
<td>-</td>
<td>0.80</td>
<td>-</td>
</tr>
<tr>
<td>Expt\textsuperscript{171}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{168} Reference number.
Figure 4.11: Total and decomposed electron density for lipid bilayers for six bilayer types. Each group is defined in the text.
4.4.6 Electron density profiles

Electron densities are plotted and shown in figure 4.11. The electron density has been described previously in section 3.6.2. Here the total electron density has been decomposed into its component groups including: choline (CHOL), amine (NH$_3^+$), phosphate (PO$_4$), glycerol (GLY), carboxyl (COO), methylene (CH$_2$), unsaturated bonds (C=C), terminal methyl (CH$_3$), and water (Water). The electron density shows the penetration of water to the interior of the bilayer, but not to the center of the bilayer, consistent with previous studies.

4.4.7 Scattering form factors

Number densities calculated from MD simulation trajectories were loaded into the program SIMtoEXP where scattering form factors were calculated. Scattering form factors allow for direct comparison of experimental and simulation data.\textsuperscript{160} Scattering profiles shown here are for only tensionless bilayer simulations.

Calculated scattering form factors match experimental form factors remarkably well, indicating that bilayer structures are correct. The simulation scattering form factors reproduce experimental minima and the relative form factor nodes are of the correct magnitude.

X-ray scattering data has been indicated to be dependent on the thickness $D_{HH}$, while neutron scattering data has shown to be dependent on the thickness $D_B$. One proposed quality measurement for scattering profiles is the value $\Delta D_{B-H} = (D_B - D_{HH})/2$. Bilayer simulations with the GROMOS united-atom force field has been shown to match experimental values.\textsuperscript{148} However, other all-atom force fields do not show such correlation. $\Delta D_{B-H}$ is plotted as a function of area per lipid in figure 4.13 for Lipid14, Charmm C36, and Slipids force fields. All three all-atom force
Figure 4.12: Simulation X-ray scattering form factors for six lipid bilayer types. Profiles from simulations are drawn with solid lines. Experimental data is represented as unfilled circles. Cyan circles are from experimental X-ray scattering profiles\textsuperscript{15,127,16,154,157} and other circles are from experimental neutron scattering form factors.\textsuperscript{127,173}
fields do not follow the experimental trend and in fact as area per lipid increases, the deviation from experiment increases. Results for Lipid14 are similar to Charmm C36 and Slipids force fields. Regardless, Lipid14 scattering profiles show satisfactory agreement with actual experimental scattering profiles.

4.4.8 Lipid lateral diffusion

Lipid lateral diffusion was evaluated using the Einstein relation in two dimensions defined in equation 4.3. Diffusion was calculated for each lipid by finding the mean squared displacement (MSD) of the lipids over 20 ns windows through the trajectories. The slope of the MSD curves corresponds to the diffusion coefficient and the linear portion of the curve was used to evaluate the slope. Lipid coordinates were processed to remove artificial center of mass drift of the bilayer.

Lateral diffusion results are on the same order of magnitude as experimental values but overall underestimated. There is no correction term in the diffusion constant for the collective motion because this cannot be sampled in a periodic box system. As a result, the size of the system may influence the diffusion calculation.
Figure 4.14: Time averaged mean square displacement of the center of mass of the lipid molecules. Lipid types indicated by colors.

Figure 4.15: Lateral diffusion coefficients calculated using different time ranges of mean square displacement.
Table 4.9: Lipid lateral diffusion coefficients calculated from NPT runs, NVE runs, and experimental values.

<table>
<thead>
<tr>
<th>Lipid system</th>
<th>Calc NPT $D_{xy}$ ($10^{-8}$ cm$^2$/s)</th>
<th>Calc NVE $D_{xy}$ ($10^{-8}$ cm$^2$/s)</th>
<th>T (K)</th>
<th>Exptl $D_{xy}$ ($10^{-8}$ cm$^2$/s)</th>
<th>Exptl T (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLPC</td>
<td>7.65</td>
<td>7.78</td>
<td>303</td>
<td>8.5, 174</td>
<td>298</td>
</tr>
<tr>
<td>DMPC</td>
<td>5.05</td>
<td>6.32</td>
<td>303</td>
<td>5.95, 21, 9, 175, 176</td>
<td>303, 303</td>
</tr>
<tr>
<td>DPPC</td>
<td>9.21</td>
<td>11.94</td>
<td>323</td>
<td>12.5, 177, 15.2, 178</td>
<td>323, 323</td>
</tr>
<tr>
<td>DOPC</td>
<td>6.45</td>
<td>9.49</td>
<td>303</td>
<td>11.5, 176, 179</td>
<td>303, 308</td>
</tr>
<tr>
<td>POPC</td>
<td>5.74</td>
<td>6.54</td>
<td>303</td>
<td>10.7, 176</td>
<td>303</td>
</tr>
<tr>
<td>POPE</td>
<td>4.67</td>
<td>4.85</td>
<td>310</td>
<td>5.2, 180</td>
<td>305</td>
</tr>
</tbody>
</table>

The diffusion calculation is, however, consistent with other bilayer simulations.

Initial diffusion coefficients were calculated with a Langevin thermostat (NPT) which randomizes particle velocities and may influence diffusion calculations. Therefore the simulations were repeated in the microcanonical ensemble (NVE) for 100 ns with the same settings as the hydrocarbon diffusion simulations. The calculated diffusion coefficients are reported in table 4.9. Overall values are similar to NPT diffusion values, though diffusion from NVE is slightly higher.

Lateral diffusion of lipids has two major modes including “rattling” of the lipid in its local solvation as well as long range diffusion. The MSD versus time plot captures these two modes of movement. Correspondingly, the diffusion coefficient calculated for these two modes is different. The initial rattling mode has a much higher diffusion coefficient, while the long distance lateral diffusion is much smaller.

4.5 Additional lipid parameters

Initial parameterization of this subset of lipids has provided a robust collection of lipid parameters for MD simulations. The modular approach of Lipid14 allows for many combinations of lipids. The re-parameterization allows for tensionless anisotropic lipid bilayer simulations.

Even with this set of phospholipids, there are thousands of biologically relevant
lipids found in membranes. It was especially important to begin to add other phospholipid head groups and tails, other classes of lipids, as well as important membrane sterols to the lipid force field. This will allow the MD simulation of diverse membranes and membrane bound proteins. The following chapter explores lipid membrane simulations with cholesterol, a crucial membrane component.

4.6 Acknowledgement

Chapter 4 includes material from “Lipid14: The Amber lipid force field” Madej, B. D., Dickson, C. J., Skjevik, Å. A., Betz, R. M., Teigen, K., Gould, I. R., Walker, R. C. *Journal of Chemical Theory and Computation* 2014 The dissertation/thesis author was a primary investigator and author of this paper.
Chapter 5

Cholesterol

5.1 Introduction

Cholesterol is an important and ubiquitous component of cell membranes. Cholesterol and ergosterol may be commonly found in many different mammalian membranes in molar fractions up to 0.50. Cholesterol has been indicated as important in not only the structure of membranes but also in functional interactions with membrane proteins.

Cholesterol significantly affects the order and phase of lipid bilayers. Cholesterol’s effects are highly dependent on bilayer type. In general, though, cholesterol has been shown to have a condensing effect on lipid bilayers at higher concentrations. With sufficient cholesterol concentration, the bilayer enters a liquid ordered phase in which the acyl tails are highly ordered and tightly packed, while lipids overall rapidly move and diffuse.

It has been hypothesized that the liquid-ordered phase is related to the raft assemblies seen in biological membranes. However, large raft assemblies have still not been directly observed in membranes. It would be valuable to simulate these types
Figure 5.1: Cholesterol in a DMPC bilayer. Cholesterol is drawn with a licorice representation, while phospholipids are rendered with van der Waals spheres in grey. Water is rendered with blue spheres.
of systems to investigate the differences between their dynamics.

Cholesterol in membranes has been simulated with MD in previous simulations. Several force fields have included cholesterol parameters: Cholesterol has been included in united-atom and coarse grain simulations.\textsuperscript{52,51} Cournia presented an all-atom cholesterol parameter set for molecular dynamics, yet still required constant surface tension.\textsuperscript{185} The Charmm force fields have included cholesterol in their lipid force fields.\textsuperscript{186} The Slipids force field has also been expanded to include cholesterol.\textsuperscript{187} Nevertheless cholesterol parameters have not been available for simulation with other Amber force fields.

As the collection of lipids has expanded to include a variety of lipid types in tensionless bilayer simulations, it is apparent that other key components should be included in the lipid force field. Accurate cholesterol parameters are needed since it is a very common membrane component and may have functional roles.\textsuperscript{188}

The parameterization of cholesterol presented here is a minimal one. The parameterization strategies are based on those previously used.\textsuperscript{2,3,4} Simulations with this basic parameter set are presented as a baseline for possible further parameterization. Figure 5.1 shows a snapshot of cholesterol in a lipid bilayer. Even with the minimal parameter set it was possible to compare MD simulations with available experimental data. The combination of Lipid14 and cholesterol in MD simulations is fully explored in this chapter.

### 5.2 Parameterization

Previous force field parameterization for cholesterol has taken several varying approaches. One common denominator for cholesterol parameterization is the development of partial charges for the sterol. In general, charges have been refined using a variant of the RESP method with some differences.\textsuperscript{71,69} Jambeck used a RESP charge
refinement with a polarizable continuum model for the ESP calculation.\textsuperscript{50} The addition of a polarizable continuum model may change the ESP such that it mimics a different dielectric environment, such as the interior of a bilayer.

Other parameterization strategies have fit similar sterol van der Waals parameters in order to match experimental enthalpies of vaporization.\textsuperscript{186} Some other approaches have examined the tail dihedral parameters.\textsuperscript{186} Another parameterization strategy has included examining the molecular frequencies in fitting.\textsuperscript{185} Cholesterol partitioning between polar environments have also be tested on the parameters.\textsuperscript{51}

Previously, the Lipid11 framework and initial force field included cholesterol with default GAFF parameters and a simple charge derivation. However, with the re-parameterization of the six bilayer types in Lipid14 for a tensionless ensemble, an investigation of cholesterol parameters with the Lipid14 parameter set was warranted. This chapter presents a force field for cholesterol using the Lipid14 parameterization strategies as a basis for further refinement. The Lipid14 parameter set is expanded to include cholesterol for future lipid simulations. Following parameterization, the parameters were tested and simulated in two bilayer types with published structural data.

Initial parameters were drawn from the GAFF parameter set. The Amber program \textit{antechamber} was used to assign atom types to the cholesterol molecule, as well as assign basic bond, angle, and dihedral terms.\textsuperscript{1,66} Additionally for the first pass, van der Waals parameters from GAFF were used for the initial molecule.\textsuperscript{64} Thus the only missing parameters were the partial charges. Partial charges were derived for cholesterol in the method described in Lipid14.

In addition to the partial atomic charges for cholesterol, the GAFF van der Waals parameters were tested for the sterol portion of the molecule. A small molecule called decalin was chosen to investigate the sterol carbon and hydrogen van der Waals
Figure 5.2: Partial charges for cholesterol. Charge were derived using the same method as molecules in Lipid14. Carbon charges are printed normally and hydrogen charges are listed in parentheses.

5.2.1 Partial charge derivation

Preliminary partial charges were needed for initial MD of cholesterol. Charges were refined using a method compatible with the Amber RESP procedure. The ESP of cholesterol was calculated on a single optimized cholesterol molecule at the HF/6-31G* level with Gaussian 2009. The HF/6-31G* level of theory and basis set was used for reasons explained previously in section 4.1.4. Besides the normal Amber RESP criteria, no additional caps or constraints were used for cholesterol.

Preliminary charges were used for initial MD simulations of cholesterol in two representative bilayers: palmitoyloleoylphosphatidylcholine (POPC) and palmitoyloleoylphosphatidylethanolamine (POPE). POPC and POPE were chosen because...
they cover the two main head groups included in Lipid14 as well as include a saturated and unsaturated tail group. Structures were built that contained 128 bilayer molecules with 0.25 molar fraction cholesterol in the bilayer. Systems were solvated with TIP3P water molecules with full lipid solvation of 32 waters per bilayer molecule. Ionic concentration of 150 mM K\(^+\) and Cl\(^-\) was included with Joung’s ion parameters in the systems. The cholesterol and lipid bilayer systems were simulated for 125 ns. POPC was simulated at 303 K and POPE was simulated at 310 K.

From the equilibrated MD trajectories, fifty random cholesterol structures were extracted from the POPC simulation and fifty random structures from the POPE simulation were extracted. RESP charges for each cholesterol conformation were calculated without optimization. The mean charges of all cholesterol conformations was used for the final charges. Because the structures are extracted from previous MD, the mean is assumed to be implicitly Boltzmann-weighted.\(^4\)

Figure 5.2 shows the molecular structure and atomic partial charges for cholesterol. Through this charge derivation method, no additional modifications to the derivation were necessary. No additional constraints were imposed and a polarizable continuum model was not used in the ESP calculation.

### 5.2.2 Sterol van der Waals

A majority of the cholesterol molecular structure is based on the basic sterol unit so the parameters of the ring portion of the molecule were re-examined. During a normal MD simulation, it is likely that the sterol portion of the molecule will mainly be inflexible due to the sterol rigidity. However, van der Waals parameters in this region may influence bilayer structural properties such as head group packing as well as bilayer volume.

The atom types assigned to the sterol carbons are basic GAFF \(sp^3\) carbons
Figure 5.3: Partial atomic charges for decalin. Hydrogen charges are written in parentheses. Using a RESP method, the partial atomic charges for decalin were derived for MD simulations of density and heat of vaporization.

Table 5.1: Decalin enthalpy of vaporization and density

<table>
<thead>
<tr>
<th>T (K)</th>
<th>$\Delta H_{vap}$ (kcal/mol)</th>
<th>Density (g/cm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>trans-decalin Sim. 298.15</td>
<td>13.77</td>
<td>0.8641 ± 0.0047</td>
</tr>
<tr>
<td>Exp. 298.15</td>
<td>9.61$^{132}$</td>
<td>0.8659$^{132}$</td>
</tr>
</tbody>
</table>

and hydrogens are assumed to be without an electron withdrawing group. The initial carbon and hydrogen van der Waals parameters were taken from the GAFF.$^{64}$

A small model molecule called decahydronaphtalene (decalin) in the $trans$ configuration was constructed with matching atom types and parameters. Charges were derived on an optimized decalin structure with the Amber RESP approach.

In order to evaluate the van der Waals parameters, decalin was simulated with MD. A SHAKE constraint for the bonds with hydrogen was used.$^{92}$ MD was run for 10 ns with a 0.002 ps timestep. Gas-phase decalin was simulated with constant volume with the Langevin thermostat at 295.15 K and 1 ps$^{-1}$ collision frequency.$^{98}$ Liquid-phase decalin with 160 molecules was simulated with the Berendsen barostat and a pressure relaxation time of 1 ps.$^{94}$ The particle mesh Ewald method was used for electrostatics with a 10 Å cutoff.$^{78}$

The average bulk density was obtained from liquid MD simulations and the enthalpy of vaporization was calculated as previously described in section 4.1.2. Ta-
Table 5.2: Summary of cholesterol and lipid bilayer simulations at varying cholesterol concentrations. $X_c$ is the molar fraction of cholesterol in the bilayer; $N_{lip}$ and $N_{chl}$ are the number of lipids and cholesterol in each bilayer, respectively; $n_w$ is the number of water molecules per bilayer molecule.

<table>
<thead>
<tr>
<th></th>
<th>$X_c$</th>
<th>$N_{lip}$</th>
<th>$N_{chl}$</th>
<th>Sim. Time</th>
<th>$T$</th>
<th>$n_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMPC</td>
<td>0</td>
<td>128</td>
<td>0</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DMPC</td>
<td>0.1</td>
<td>116</td>
<td>12</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DMPC</td>
<td>0.2</td>
<td>102</td>
<td>26</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DMPC</td>
<td>0.3</td>
<td>90</td>
<td>38</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DMPC</td>
<td>0.4</td>
<td>76</td>
<td>52</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DMPC</td>
<td>0.5</td>
<td>64</td>
<td>64</td>
<td>3 x 200</td>
<td>303</td>
<td>25.6</td>
</tr>
<tr>
<td>DOPC</td>
<td>0</td>
<td>128</td>
<td>0</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>DOPC</td>
<td>0.1</td>
<td>116</td>
<td>12</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>DOPC</td>
<td>0.2</td>
<td>102</td>
<td>26</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>DOPC</td>
<td>0.3</td>
<td>90</td>
<td>38</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>DOPC</td>
<td>0.4</td>
<td>76</td>
<td>52</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
<tr>
<td>DOPC</td>
<td>0.5</td>
<td>64</td>
<td>64</td>
<td>3 x 200</td>
<td>303</td>
<td>32.8</td>
</tr>
</tbody>
</table>

Table 5.1 lists the simulation and experimental parameters. The density of liquid decalin matches experimental densities at the same temperature. However, the enthalpy of vaporization is significantly overestimated. Previous work has suggested that it may be difficult to fit enthalpy of vaporizations through only van der Waals parameters. For this study, the GAFF parameters were unchanged in the MD simulations. This minimal parameterization is examined in detail in lipid bilayer MD presented in the following sections.

### 5.3 Cholesterol molecular dynamics

For testing of the cholesterol parameter set, lipid bilayers with varying concentrations of cholesterol were simulated with MD. Dimyristoyl-phosphatidylcholine (DMPC) and dioleoyl-phosphatidylcholine (DOPC) were selected for validation because there is experimental structural data for these bilayer types with cholesterol.

Bilayers with 128 molecules were constructed with cholesterol molar fractions $X_c$ from 0-0.5, which is below the known cholesterol saturation concentration in
bilayers. The number of lipids $N_{lip}$ and the number of cholesterols $N_{chl}$ is listed in table 5.2. The number of water molecules per bilayer molecule, including lipid and cholesterol, is also listed. Each bilayer concentration was simulated in triplicate for at least 200 ns resulting in sampling for these systems of more than 7 µs.

Initial structures were generated with the Charmm-GUI web site. Cholesterol were evenly distributed between the two bilayer leaflets. Initial structures formatted with the Charmm C36 atom naming were converted to the Lipid14 format with the charmm lipid2amber.py script in AmberTools 14. Formatted structure files were loaded into leap and parameters and topology assigned. The Lipid14 parameters were used for glycerophospholipids. Each bilayer was fully solvated with water molecules and 150 mM K$^+$ and Cl$^-$ ions were included. The TIP3P water model was used for water molecules and Joung’s ion parameters were used for the monovalent ions.

Simulation of cholesterol and lipid bilayers was carefully designed to allow for gradual system equilibration. The protocol overall is similar to that used for Lipid14 simulations. Initial structures are minimized for 10000 steps consisting of 5000 steps of steepest descent followed by 5000 steps of conjugate gradient minimization. Two stages of heating the bilayer follow minimization. In the first stage the bilayer was heated from 0 K to 100 K over 5 ps with constant volume. The Langevin thermostat was used throughout heating and production with a collision frequency of 1 ps$^{-1}$. A weak restraint was placed on the bilayer atoms with a force constant of 10 kcal mol$^{-1}$ Å$^{-2}$. In the second stage of heating, the system was heated from 100 K to production temperature over 1 ns, with the Langevin thermostat. The Berendsen barostat with a 1 ps relaxation time was used and equilibrated system volume and density.

Production MD was simulated in the NPT ensemble with the Langevin thermostat and the Berendsen barostat. Production temperature was specific to bi-
layer type and a Langevin collision frequency of a ps$^{-1}$ was set. The SHAKE constraint algorithm was used for bonds involving hydrogen. Simulations were run for 200 ns with a 0.002 ps timestep. Anisotropic Berendsen barostat was set with a reference pressure of 1 atm and a pressure relaxation time of 1.0 ps. The particle mesh Ewald summation method was applied to electrostatic interactions with a real-space cutoff of 10 Å. The same cutoff was used for van der Waals interactions.

The membranes were simulated with the Amber MD software packages using the GPU-accelerated MD engine pmemd.cuda. Each membrane simulation was run the same initial configuration with randomized velocities and run in triplicate. All analysis was completed with scripts written for ptraj and cpptraj.

5.4 Parameter validation

Following the MD simulations, trajectories were analyzed and compared with available experimental data. Along with estimates of area per lipid and volume per lipid of the lipid bilayer, there are several important datasets available for comparison with experiment. X-ray and neutron scattering profiles provide an important source of structural information that may be compared directly to simulation results. Furthermore, features of the electron density profile like bilayer thickness are comparable with experiment. Acyl chain order parameters allow for another important comparison of the effect of cholesterol on lipid bilayers.

5.4.1 Volume per molecule

Volume per molecule measurements differ slightly from volume per lipid measurements because they are defined as the total volume of the bilayer divided by the number of lipids and cholesterol molecules. This may be calculated in a similar man-
Figure 5.4: Average volume per molecule of cholesterol lipid bilayers. Boxes are mean simulation area per molecule across three simulations. Circles are experimental volumes per average molecule at 30 C from Greenwod, and triangles are values at 35 C from Hodzic.

Figure 5.4 shows the change in average volume per lipid as a function of cholesterol concentration in lipid bilayers. Averages for all simulation properties are calculated as the mean value across all three 200 ns simulations. As the fraction of cholesterol in the lipid bilayer increases, the volume per molecule in the bilayer decreases. This is expected as the increase in cholesterol concentration facilitates tighter packing between lipids. Overall, the volume per molecule in the bilayer is systematically underestimated. Results are consistent with Lipid14 simulations in that the volume per lipid was also underestimated. However, the MD simulations follow the same trend as the experimental volume per lipid. Note that there is some
Figure 5.5: Average area per molecule of cholesterol lipid bilayers. Bars indicate the mean standard deviation of simulation volume across simulations. Circles are the experimental area per average molecule at 30 C from Pan.\textsuperscript{184} variability between experimental values at higher cholesterol concentrations.

5.4.2 Area per molecule

Similar to area per lipid described in section 3.6.1, area per molecule is a measurement of the average lateral area for lipid and cholesterol molecules. This value is usually derived from volume and thickness measurements.\textsuperscript{184} Several methods have been proposed for calculating the area per molecule for each component and have been applied to binary bilayers.\textsuperscript{191} The average area per molecule presented here simply includes both cholesterol and lipid components. Thus the area per molecule $A_M$ may be defined as $A_M = \frac{A_{XY}}{(N_l + N_{chl})}$.

Figure 5.5 shows the area per molecule as a function of cholesterol concentration. The area per molecule is shown to decrease with increasing cholesterol concentration. The area per molecule calculated from MD simulations follows this same
trend, but it always underestimates the area in DOPC. Simulations with other cholesterol force fields report low areas per molecule: Jambeck lists an area of 57 Å² and 43 Å² for 20% and 50% cholesterol, respectively.\textsuperscript{187}

It is also notable that cholesterol has an observed saturation point in lipid bilayers. Beyond saturation, cholesterol precipitates from the bilayer as cholesterol monohydrate. Huang reports a cholesterol saturation point in phosphatidylcholine bilayers to be 66%.\textsuperscript{192} Hung reports a lower saturation point: approximately 44% in DMPC bilayers and 40% in DOPC bilayers.\textsuperscript{193} It is possible that the cholesterol bilayer simulations presented here approach the saturation points of these bilayers. During cholesterol simulations, precipitation was not observed. This is because it unlikely on the current time scales of simulations to observe cholesterol precipitation.

### 5.4.3 Bilayer thickness

Thickness may be calculated in several ways. The head-to-head thickness $D_{HH}$ and Luzatti thickness $D_B$ are described previously in section 4.4.4. Bilayer thicknesses were calculated in the same way as in the Lipid14 simulations.

Figure 5.6 shows the dependence of bilayer head-to-head thickness on cholesterol concentration. Increasing cholesterol molar fraction results in an increase in bilayer thickness. This is consistent with the condensing effects of cholesterol, resulting in more ordered lipids.\textsuperscript{183,19} Head to head thicknesses overall followed the experimental trend in bilayer thicknesses.\textsuperscript{184,194}

Simulation thicknesses may also be compared with other all-atom cholesterol force field simulations. The Charmm C36c force field from Lim reports electron density profiles in comparable bilayer simulations from which the peak-to-peak distance may be estimated.\textsuperscript{186} Charmm C36c reports bilayer thicknesses of approximately 38 Å and 42 Å for DMPC with 10% and 30% fraction cholesterol, respectively. Lipid14
cholsterol bilayer thicknesses for the same systems were 37.5 Å and 44 Å. Charmm C36c DOPC bilayers with 10% fraction cholesterol had a thickness of approximately 40 Å compared with Lipid14’s thickness of 37.5 Å.

Slipids by Jambeck simulated the same systems at a range of cholesterol concentrations. For both DOPC and DMPC, Lipid14 follows the same trend of bilayer thicknesses. For DMPC, Lipid14 bilayer thicknesses are slightly higher overall than the Slipids values by about 2 Å. For DOPC, Lipid14 follows the trend of Slipids thickness within 1 Å. Lipid14 cholesterol parameters resulted in bilayers with thicknesses that fall within the Charmm and Slipids values.
Figure 5.7: Deuterium order parameters $S_{CD}$ for DMPC and DOPC tails with cholesterol. Simulation profiles were calculated directly from the trajectory C-D vectors and averaged across the three simulations. DMPC/CHL 0.3 order parameters shown are for the sn-2 tail. DMPC/CHL 0.3: Circles are experimental order parameters at 25 C from Douliez,\textsuperscript{165} upward triangles are at 25 C from Urbina,\textsuperscript{196} orange downward triangles are with 0.33 molar fraction cholesterol at 30 C from Vermeer.\textsuperscript{197} DMPC/CHL 0.5: Upward triangles and downward triangles are $sn$-1 and $sn$-2 perdeuterated order parameters at 40 C from Trouard,\textsuperscript{198} 39 DOPC/CHL 0.3 Upward triangles and downward triangles are $sn$-1 and $sn$-2 order parameters at 37 C from Warschawski.\textsuperscript{128}
Figure 5.8: Deuterium order parameters $S_{CD}$ for cholesterol in DMPC bilayer. 0.3 molar fraction cholesterol is present in the bilayer. Boxes and circles are axial and equatorial simulation order parameters respectively. Upward triangles are axial and downward triangles are equatorial order parameters at 30 C from Vermeer.\textsuperscript{197}

### 5.4.4 Cholesterol order

NMR order parameters provide important information about the lipid tail orientation. The deuterium order parameter is described in detail in section 3.6.3. Vermeer reports order parameters for axial and equatorial C-D bonds in cholesterol.\textsuperscript{197}

Figure 5.7 plots the deuterium order parameter $S_{CD}$ as a function of each carbon atom along the acyl chain. Order parameters were calculated for each lipid tail at every simulated molar fraction of cholesterol. Deuterium order parameters were calculated from the DMPC and 0.3 molar fraction cholesterol bilayer simulation. The myristoyl order parameters presented here fall within the range of values in experiments under similar conditions.

Other force fields examine the order parameters of the phospholipid under the same conditions. The Charmm C36c parameters have a similar myristoyl order profile, except with a lower order plateau (approximately 0.38).\textsuperscript{186} The Slipids order profile of the $sn$-2 8-13 carbons is similar, except that the Lipid14 profile is slightly more ordered.\textsuperscript{187} The order parameters presented here lie within the range of experimental order parameters at 25 and 30 C and follow a similar trend as the Charmm C36c and Slipids force fields.
Order parameters for myristoyl with 0.5 molar fraction cholesterol are also presented. Trouard reports experimental order parameters from di-perdeuterated DMPC experiments.\textsuperscript{198} It is important to note that perdeuterated order parameters are usually reported as monotonically decreasing, and may not correspond to the actual carbon sequence.\textsuperscript{199} Simulation order parameters fall within the range of experimental values and have a similar plateau value. Slipids reports a similar trend for myristoyl order parameters though with lower order in the plateau region for the sn-1 tail.

DOPC order parameters with 0.3 molar fraction cholesterol was reported by Warschawski.\textsuperscript{128} Simulations done under the same conditions revealed that the order parameters for carbons 9-11 and 15-18 resemble but slightly overestimate the experimental values.

Slipids order parameters for the oleoyl tail are overall higher than the experimental values presented by Warschawski.\textsuperscript{187} Lipid14 more closely fits the experimental values in this regard.

The addition of cholesterol significantly influences the dynamics of the acyl tails of phospholipids. The order parameters of the tails increase significantly. This increase is suggestive of the liquid ordered phase of the bilayer, in which the tails are highly ordered, but the lipids still diffuse and rotate.

Interestingly, Vermeer also reports cholesterol order parameters.\textsuperscript{197} Axial and equatorial C-D bonds were reported for cholesterol in actual bilayers. The DMPC and 0.3 molar fraction cholesterol bilayer was analyzed to calculate the order parameters of C-D vectors near the hydroxyl group. Order parameters from the simulation are in excellent agreement with Vermeer’s experimental values. Based on the deuterium order parameters, on average, the cholesterol in this bilayer are oriented in a similar manner to actual cholesterol and lipid bilayers. It is encouraging that the cholesterol
Figure 5.9: Experimental X-ray form factors for cholesterol and DMPC bilayers. Each panel is a different cholesterol concentration. Solid lines are simulation form factors. Boxes are oriented (ORI) and ULV form factors at 30 C from Pan et al., circles are form factors at 35 C from Hodzic et al., and triangles are ULV form factors at 30 C from Kucerka et al. has the same order parameters and is oriented in the same way as in Vermeer’s experimental results.

5.4.5 Scattering form factors

Scattering form factors are important direct comparisons of experimental structural properties with simulation. With the form factor and molecular modeling it is possible to estimate an electron density profile that goes through the bilayer normal. Scattering analysis is described in section 4.4.7.

However, this method is still dependent on some model parameters to determine the EDP. A more direct comparison is to calculate the scattering form factors from the electron density profile via a Fourier transform operation. The program SIMtoEXP converts a trajectory number density or electron density to a scattering curve. For the cholesterol and lipid bilayers, the number density was calculated
through modifications to *ptraj*. The densities were loaded into *SIMtoEXP* and processed into scattering profiles with the default settings. Form factors were scaled to minimize the mean squared deviation compared to the simulation form factors.

Experimental scattering form factors are available for some bilayers that include cholesterol. Figure 5.9 shows the X-ray scattering profiles for DMPC and cholesterol bilayers in direct comparison with experimental profiles. DMPC bilayers with no cholesterol have a form factor that is consistent with both experimental form factors and previous Lipid14 simulations.\(^{127,188}\) Form factor maxima and minima are reproduced in the simulation results. At higher molar fractions of cholesterol the minimum of the scattering profile is shifted. The relative peak heights are similar to experimental profiles. Nodes at higher \(q\) values are also predicted. Throughout a range of cholesterol concentrations, the simulation scattering profile matches experimental scattering profiles.

Other force fields compare scattering profiles to experimental scattering data as well. The Slipids form factors resemble the Lipid14 scattering profiles, however with minima shifted to smaller \(q\) values.\(^{187}\) The Charmm C36c compares scattering form
factors of DMPC with 0.1 molar fraction cholesterol. At high $q$ values, the Charmm C36 scattering profile does not reproduce the node, while the Lipid14 scattering profile does capture this node.

Figure 5.10 shows X-ray scattering profiles for DOPC and cholesterol bilayers in comparison with experiment. Limited X-ray experimental scattering data exist for DOPC bilayers. At 0.3 molar fraction cholesterol, simulation form factors match experimental form factors. Neutron form factors were available in 100% D$_2$O at a range of cholesterol concentrations. Simulation neutron scattering factors match experimental profiles.

Slipids also includes DOPC X-ray form factors. At 0.3 cholesterol, the Slipids profile resembles the Lipid14 profile closely.

## 5.5 Membrane molecular dynamics

With the parameters for phospholipids and cholesterol, it is now possible to accurately simulate a wide variety of membrane types. There are countless potential applications for the force fields presented here. For example, it is now possible to simulate complex membrane environments around membrane proteins.

One application of growing interest in the field of drug development is MD simulations of small molecules passing through membranes. The next chapter describes methods in Amber to calculate small-molecule permeability across a membrane. With accurate membrane simulations, it may be possible to compare relative drug permeability across the membrane.
5.6 Acknowledgement

Chapter 5 is currently being prepared for submission for publication of the material. Madej, B. D., Gould, I. R., Walker, R. C. The dissertation/thesis author was the primary investigator and author of this material.
Chapter 6

Accelerated molecular permeability simulations

There are several approaches to estimate diffusion constants and permeability from molecular dynamics simulations that can be applied to the transport of small molecules across the membrane. Given that the calculation of potential of mean force through constraint methods is relatively slow, an accelerated implementation would be beneficial to making these approaches efficient. In this chapter an implementation of constrained molecular dynamics accelerated by graphics processing units (GPUs) is presented.

Figure 6.1 shows a cartoon of a water molecule passing through a lipid membrane. With classical MD it is difficult to sample enough small-molecule permeation events on current time-scales of simulations. This is especially true for molecules that have high energetic barriers to permeation across the membrane. Other MD methods, such as constrained MD, allow for sampling at specific depths within the membrane. The inhomogenous solubility-diffusion mechanism provides a theoretical framework for the calculation relative permeabilities of small-molecules with an
appropriate membrane force field. With an accelerated constraint MD method this approach could be applied to larger libraries of small-molecules.

6.1 Inhomogeneous solubility-diffusion mechanism

Many algorithms for examining the permeation of molecules across a lipid bilayer are based on a model of inhomogeneous solubility-diffusion originally described by Marrink and Berendsen. This is a modification of the homogeneous solubility-diffusion model applied to membranes. The derivation of the model is reproduced here as a basis for the eventual calculation of experimental membrane permeability.

Through the inhomogeneous solubility-diffusion mechanism, the permeability of the bilayer may be calculated directly from MD simulations and also compared to experimental properties. Thus the solubility-diffusion mechanism provides the theoretical framework for several approaches examining the movement of molecules across the membrane.

The derivation begins with particles of $i$ species. The particles are assumed
to be at the diffusional limit.

\[ u_i = -\frac{1}{\xi_i} \nabla \mu_i \]  (6.1)

The average velocity \( u_i = \langle \nu \rangle \) is the negative gradient of the thermodynamic potential \( \mu_i \) and is the frictional coefficient \( \xi_i \) of the particles.

The flux of the particles \( J_i \) can then be expressed by

\[ J_i = c_i u_i = -\frac{c_i}{\xi_i} \nabla \mu_i \]  (6.2)

where \( c_i \) is the concentration of the \( i \)th component. When a concentration gradient of an ideal solution is considered for which \( \mu_i = \mu_i^0 + RT \ln c_i \), equation 6.2 becomes Fick’s law:

\[ J_i = -D_i \nabla c_i \]  (6.3)

where \( D_i \) is the diffusion constant. The Einstein-Smoluchowski relation describes the diffusion constant in terms of the the frictional coefficient

\[ D_i = \frac{RT}{\xi_i} \]  (6.4)

where \( R \) is the Boltzmann constant and \( T \) is the temperature. When the material properties depends on just one dimension, the flux may be rewritten

\[ J_i(z) = -\frac{c_i(z)D_i(z)}{RT} \frac{d\mu_i(z)}{dz} \]  (6.5)

The flux is now written as a function of the driving force. By limiting this formulation to one dimension, cross terms are not included in the flux and driving forces. Based
on the conservation relationship, the sum of derivatives of flux and concentration are

$$\frac{dJ_i(z)}{dz} + \frac{dc_i(z)}{dt} = 0 \quad (6.6)$$

Equation 6.5 and the conservation equation 6.6 define the change in the density distribution. At steady-state flux not far from equilibrium, assume $J_i$ is not a function of $z$. Equation 6.5 may be rearranged and integrated with respect to $z$ across the membrane:

$$\Delta \mu_i = \mu_i(z_2) - \mu_i(z_1) = -J_i RT \int_{z_1}^{z_2} \frac{dz}{c_i(z)D_i(z)} \quad (6.7)$$

$c_i(z)$ is the concentration in the presence of the imposed gradient. If small gradients are assumed, the concentration may be replaced with an equilibrium concentration $c_i^{eq}$ without an imposed gradient. The permeation resistance $R_p^i$ may then be defined as

$$R_p^i = c_i^* \int_{z_1}^{z_2} \frac{dz}{c_i^{eq}(z)D_i(z)} \quad (6.8)$$

$c_i^*$ is the concentration in the bulk solution on each side of the membrane.

Several assumptions are made in the derivation of permeation resistance. One important assumption should be discussed, and that is the local diffusion model must be valid. Namely, the thermodynamic gradient must be constant within particle correlation distances. This may be an issue in actual membranes because the concentration gradients in the membrane may not always be small. This is an important assumption for simulations with this approach.

With the permeation resistance from 6.8, the linear response equation 6.7 is rearranged to form

$$J_i = -\frac{c_i^* \Delta \mu_i}{R_p^i RT} \quad (6.9)$$
6.1.1 Permeation resistance and permeability

Marrink and Berendsen further derived the relation between permeation resistance and permeability for the inhomogeneous solubility diffusion mechanism.\textsuperscript{201} The permeability coefficient may be determined experimentally for molecular permeation through membranes.

In this model, the main force for permeation is divided among hydrostatic pressure differences, osmotic pressure differences, and concentration difference. Marrink and Berendsen then derive the flux for an isotope of water and solute (derivation not shown).\textsuperscript{201} From equation 6.9 it is possible to derive expressions for the flux of an isotope of water $J_{is}$ and flux of the solute $J_s$:

$$J_{is} = -\frac{1}{R_{pi}} \Delta c_{is}$$

$$J_s = -\frac{1}{R_{ps}} \Delta c_s$$  \hspace{1cm} (6.10)

The permeability coefficient may be defined as the ratio between flux and concentration difference. As shown in equation 6.10, the permeability $P_i$ is equivalent to the inverse of the permeation resistance $R_{pi}$

$$R_{pi} = 1/P_i$$  \hspace{1cm} (6.11)

6.2 Calculation of a potential of mean force

In order to calculate the permeability coefficient, first the local concentration as ratio to the bulk concentration $c_i^{eq}(z)/c_i^*$ in equation 6.8 must be evaluated from an MD simulation.\textsuperscript{201} The local concentration is proportional to the probability that
a structure with the permeant in the interval \((z, z + dz)\) is in the total phase space of the structure. This can be defined with the constrained partition function \(Q'\).

\[
c_i^{eq}(z) \sim Q'(z) = a \int dr_1 \ldots dr_N \delta(z_0 - z) \exp\{-V(r_1 \ldots r_N)/kT\}
\]

(6.12)

\(a\) is a constant in this partition function. Therefore, the ratio \(c_i^{eq}(z)/c_i^*\) is the ratio of \(Q'(z)\) and \(Q'(z_1)\) at position \(z_1\). This in turn can be written as a function of the potential of mean force \(\Delta G\).

\[
\Delta G_i(z) = -RT \ln \frac{Q'(z)}{Q'(z_1)} = -RT \ln \frac{c_i^{eq}(z)}{c_i^*}
\]

(6.13)

Finally, the permeation resistance from equation 6.8 can be written in terms of the potential of mean force and diffusion coefficient:

\[
R_i^p = \int_{z_1}^{z_2} \frac{\exp(\Delta G_i(z)/RT)}{D_i(z)} \, dz
\]

(6.14)

Several approaches are available to estimate the potential of mean force.\(^{201,204}\) These PMF methods include the analysis of local density, particle insertion methods, and average force on a constrained molecule. Each has its own advantages and limitations.\(^{201,204}\)

The analysis of local density is based on equation 6.13 and involves sampling the frequency that the permeant is present at a certain depth. However, an inherent problem with the local density analysis is that sampling is a major issue. Given classical MD and slow permeation rates, it is unlikely to sample many permeation events during the course of the simulation with local density analysis. Therefore, it is unreliable to calculate a PMF through a lipid bilayer without proper sampling of the permeation event.
Another method is the particle insertion method described by Widom. The method relies on the insertion of particles that do not disturb the system. The energy of insertion of the artificial particle into the system is calculated in this method. This method may be applied to membrane permeation. However, a major issue with this approach is that there needs to be enough sampling in the distribution of states across the maximum energy and transition of states.

Another method is to use the average force on a constrained molecule which is included in the following section.

6.2.1 Average force on constrained particles

The derivative of the potential of mean force exerted on a constrained molecule can be calculated from an MD simulation. First, the derivative of the free energy in equation 6.13 is written as

$$\frac{d\Delta G(z)}{dz} = -RT \frac{dQ'(z)}{dz}$$  \hspace{1cm} (6.15)

Then the derivative of energy equation 6.12 may be evaluated using the method of partial integration.

$$\frac{dQ'(z)}{dz} = -\frac{a}{KT} \int dr_1 \ldots dr_N \delta(z_0 - z) \frac{\partial V}{\partial z_0} \exp\{-V/KT\}$$  \hspace{1cm} (6.16)

Therefore

$$\frac{d\Delta G(z)}{dz} = -N_{Av} \left\langle \frac{\partial V(r_1 \ldots r_N)}{\partial z_0} \right\rangle = -N_{Av} \langle F_z(z_0) \rangle$$  \hspace{1cm} (6.17)

$\langle F_z(z_0) \rangle$ is the mean force from the constraint averaged over the simulation. This can be calculated directly from the MD simulation. There are various algorithms available to constrain molecules. However, it is also necessary to calculate of the
corresponding constraint force for each constraint applied.

6.3 Computation of local diffusion constants

The only missing component to calculate the permeability of a molecule through a membrane is the calculation of the local diffusion $D(z)$ constant within the membrane.

One common method to evaluate the diffusion constant is from the derivative of the mean-square displacement of the molecule.\textsuperscript{201} During normal diffusion of the molecule, the diffusion coefficient $D(z)$ is calculated from

$$D(z) = \lim_{t \to \infty} \frac{\langle (z(t) - z(0))^2 \rangle}{2t} \quad (6.18)$$

This approach is reasonable for the interfacial region of the membrane and solvent. As a result, on the time scales of MD simulations it is unlikely to observe permeation of certain types of molecules through the membrane. Therefore some other approach is necessary for the calculation of the diffusion constant.

The fluctuation-dissipation theorem provides another method to evaluate the local diffusion constant.\textsuperscript{205,206} The autocorrelation function of the random forces $\Delta F(t)$ on a molecule is related to the friction coefficient $\xi(z,t)$ with

$$\xi(z,t) = \frac{\langle \Delta F(z,t) \Delta F(z,0) \rangle}{RT} \quad (6.19)$$

Taking the integral with respect to time of equation 6.19 yields the local static friction coefficient $\xi^s(t)$. The static friction coefficient can be used to find the local diffusion coefficient through the Einstein-Smoluchowski relation:

$$D(z) = \frac{RT}{\xi^s(z)} = \frac{(RT)^2}{\int_0^\infty \langle \Delta F(z,t) \Delta F(z,0) \rangle dt} \quad (6.20)$$
Conveniently, the local random forces can be obtained from the forces on the position restrained molecules as described in section 6.2.1. The random force can be calculated from the average force with

\[ \Delta F(z, t) = F(z, t) - \langle F(z, t) \rangle \] (6.21)

Thus, local diffusion coefficient may be obtained simultaneously with the potential of mean force in MD simulations.

### 6.4 Constrained molecular dynamics

A variety of constraint methods have been applied toward the calculation of the potential of mean force. Constrained molecular dynamics is included in various MD programs. One of the most common molecular dynamics constraints is that described by the SHAKE algorithm originally introduced by Ryckaert, Ciccotti, and Berendsen. This algorithm has been rigorously tested in many simulations because these simulations use SHAKE to constrain bond distances that involve hydrogen.

Essex et al also present a constraint algorithm applicable to potential of mean force calculations. However, this method suffered from issues related to the translation of the center of mass of the system which were later addressed.

The SHAKE algorithm is a numerical algorithm that takes place during integration of the equations of motion of a system with a set of holonomic constraints. The SHAKE algorithm fulfils the holonomic constraints by iteratively solving the set of constraint equations derived from the method of Lagrangian multipliers.

The following section describes the SHAKE constraint algorithm within the context of the Amber MD package. The algorithm was originally described by Ryckaert et al, and its implementation in Amber is explained here. Currently, SHAKE
can only be used in Amber for bonds involving hydrogen. The SHAKE algorithm fits within Verlet-based integration methods, and for this dissertation was implemented in the Amber leap-frog integrator.\textsuperscript{90}

The description of the SHAKE algorithm begins with a system of $N$ interacting particles controlled by $l$ holonomic constraints:

$$\sigma_k(\{r\}) = 0 \quad (k = 1, \ldots, l) \quad (6.22)$$

Each particle contains contributions of $F_i$, a force from the potential energy, and $G_i$, a force from the constraint $\sigma_k$ that includes the $i$th particle. $G_i$ can be written in terms of a set of $l$ Lagrangian multipliers:

$$G_i = -\sum_{k=1}^{l} \lambda_k(t) \nabla_i \sigma_k \quad (6.23)$$

In order to be used in the Verlet integration algorithms, the derivatives of $G$ must also be known. Ryckaert describes the integration of the cartesian equations of motion with the method of Lagrangian multipliers.\textsuperscript{92} The full method of constrained MD with Lagrangian multipliers is included in Ryckaert’s paper.\textsuperscript{92}

This section continues with a description of the SHAKE algorithm. From the method of Lagrangian multipliers the $k$th constraint involving particles $(i, j)$ is described by

$$2(r'_j - r'_i) \cdot (\delta r_j - \delta r_i) + (\delta r_j - \delta r_i)^2 = d_{ij}^2 - (r'_j - r'_i)^2 \quad (6.24)$$

where $\delta r_i$ are given in terms of

$$\delta r_i(t_0 + \Delta t, \{\gamma_k\}) = \frac{1}{m_i} \frac{(\Delta t)^n}{n!} \sum_{k=1}^{l} \gamma_k(\nabla_i \sigma_k)_{t_0} \quad (6.25)$$

There is now a set of $l$ equations of the form of equation 6.24 that describe a
matrix to solve for the vector \{\gamma_k\}. While these equations are nonlinear, it is possible to solve the equations via iteration. Solving the set of equations requires a matrix inversion and therefore is computationally expensive.

The alternative to this is to iteratively apply each constraint. This forms the basis of the SHAKE constraint method. The SHAKE constraint begins with the \(k\)th constraint defined as:

\[
\sigma_k = (\mathbf{r}_i - \mathbf{r}_j)^2 - d^2_{ij} = 0
\]  

(6.26)

During the normal course of MD leap-frog integration, the position at \(t + \Delta t\) is \(\mathbf{r}'_i\) in an unconstrained time step. The constraint is fulfilled by adding a displacement \(\delta \mathbf{r}_i\) to \(\mathbf{r}'_i\) resulting in \(\sigma_k = 0\) for \(\mathbf{r}_i = \mathbf{r}'_i + \delta \mathbf{r}_i\).

The constraint force between the two particles \(i\) and \(j\) are the same magnitude and in the opposite direction along the vector between \(i\) and \(j\). Let \(g_{ij}\) be the contribution of the \(k\)th constraint to the displacement \(\delta \mathbf{r}_i\). \(\gamma_k\) is uniquely determined and therefore \(g_{ij}\) is also uniquely determined. It is necessary to satisfy all constraints with displacements set as

\[
\delta \mathbf{r}_i = \sum_j g_{ij}\mathbf{r}_{ij}(t_0)/m_i \]  

(6.27)

Each individual constraint is considered in sequence. The algorithm corrects the position of each pair of particles for the \(k\)th constraint according to

\[
\delta^k \mathbf{r}_i = g_{ij}\mathbf{r}_{ij}(t_0)/m_i \]  

(6.28a)

\[
\delta^k \mathbf{r}_j = -g_{ij}\mathbf{r}_{ij}(t_0)/m_j \]  

(6.28b)

Thus the unconstrained position \(\mathbf{r}'_i\) is corrected with \(\sum_k \delta^k \mathbf{r}_i\). A quadratic
equation for each constraint \( g_{ij} \) can be obtained. Define the following:

\[
r' = r'_i + \sum_{k'<k} \delta^{k'} r_i - \left( r'_j + \sum_{k'<k} \delta^{k'} r_j \right) \quad (6.29a)
\]

\[
\delta r = \delta^k r_i - \delta^k r_j \quad (6.29b)
\]

\[
r = r_i(t_0) - r_j(t_0) \quad (6.29c)
\]

\[
g = g_{ij} \quad (6.29d)
\]

\[
d = d_{ij} \quad (6.29e)
\]

The constraint can be refactored as

\[
(r' + \delta r)^2 - d^2 = 0
\]

\[
\delta r = \left( \frac{1}{m_1} + \frac{1}{m_2} \right) g r \quad (6.30)
\]

which ultimately gives the equation

\[
2 \left( \frac{1}{m_1} + \frac{1}{m_2} \right) \cdot g \cdot (r \cdot r') + \left( \frac{1}{m_1} + \frac{1}{m_2} \right)^2 g^2 r^2 = d^2 - r'^2 \quad (6.31)
\]

The SHAKE constraint is now defined in equation 6.31. The iterative algorithm of the SHAKE constraint solves each quadratic equation within a tolerance. Constraints can depend on each other, so when one constraint is solved, it may disrupt the other constraints. Each constraint is solved until all constraints are simultaneously fulfilled within a specific tolerance. The total correction \( \delta r_i \) includes all individual constraint contributions.
6.4.1 Implementation

The implementation described here is within the context of the Amber MD programs.\textsuperscript{1,55,56,57,58} Amber’s main molecular dynamics programs \textit{sander} and \textit{pmemd} have included the SHAKE algorithm for many years, but applied only to the length of bonds involving hydrogen. Amber also includes many methods for \textit{restraining} molecular geometries. However, Amber has not had an efficient, general constraint method in its main MD program.

The basic leap-frog integration algorithm with simple temperature, pressure coupling, and SHAKE constraints is described in algorithm 6.1. The leap-frog algorithm also includes the basic locations of the available thermostats and barostats in the MD loop. Within the context of the Amber leap-frog integrator, the SHAKE constraint for hydrogens and general groups is included after the position update. Combined with the position update due to the constraints, it is important to correct the velocities after the constraint position update.

The actual general SHAKE constraint method is explained in algorithm 6.2.\textsuperscript{92} The algorithm to calculate the constraint and forces on a molecule with respect to a reference groups is described there. The subscripts \(c\) stands for constraint group and \(r\) stands for reference group. Distances are defined with the symbol \(d\).

The algorithm needs the positions from the previous step \(\mathbf{r}(t)\) and the updated coordinates \(\mathbf{r}'(t + \Delta t)\). The main loop of the constraint conducts multiple iterations over multiple constraints until all the constraints are satisfied within a tolerance. For a system with only one constraint, the constraint will always be fulfilled within one iteration with this method.

Afterwards, the final change in center of mass of constrained and reference groups is calculated. This is used to estimate the force. The virial is then updated with the constraint forces. The final step is to update the actual atom positions
Algorithm 6.1 Amber leap-frog integrator

**Require:** positions $r(t)$, velocities $v(t - 1/2\Delta t)$, box dimensions $l(t)$

- Calculate forces $F(t)$ on all atoms with PME or GB
  
  $F_i(t) = -\nabla E_i(t)$
  $a_i(t) = F_i(t)/m_i$

- if need virial or energy then
  - Calculate virial and energies
  end if

- if Monte carlo barostat then
  - Attempt Monte Carlo barostat exchange
  end if

- if Andersen thermostat then
  - Randomize velocities every $n$ steps
  end if

- Calculate velocities
  
  $v(t + 1/2\Delta t) = v(t - 1/2\Delta t) + a(t)\Delta t$

- if Berendsen thermostat then
  - Scale velocities
else if Langevin thermostat then
  - Adjust velocities via Langevin dynamics
end if

- Update positions
  
  $r'(t + \Delta t) = r(t) + v(t + 1/2\Delta t)\Delta t$

- Apply constraints
  
  $r(t), r'(t + \Delta t) \rightarrow r(t + \Delta t)$

- Calculate constrained velocities
  
  $v(t + 1/2\Delta t) = (r(t + \Delta t) - r(t))/\Delta t$

- if barostat then
  - Scale coordinates and box dimensions
    
    $\mu r(t + \Delta t) \rightarrow r(t + \Delta t)$
  end if
**Algorithm 6.2** General SHAKE algorithm between constrained and reference atoms

**Require:** \( r(t), r'(t + \Delta t), \) virial \( \Xi(t) \)

```for\_constraint \ k \ do```
- Calculate center of mass of constrained atoms \( r_c \) and \( r'_c \)
- Calculate center of mass of reference atoms \( r_r \) and \( r'_r \)

\[
\begin{align*}
    r_{ij} &= r_c - r_r \\
    \text{New center of mass } r^*_c \text{ and } r^*_r \text{ to be constrained:} & \\
    r'_c &\to r^*_c \\
    r'_r &\to r^*_r \\
\end{align*}
\]

```end\_for```

```while\ any\ constraint\ not\ converged\ and\ i < i_{max}\ do```

```for\_constraint \ k \ do```
- \( r^*_{ij} = r^*_c - r^*_r \)
- Solve \( 2\left( \frac{1}{m_1} + \frac{1}{m_2} \right) (r_{ij} \cdot r^*_{ij}) + \left( \frac{1}{m_1} + \frac{1}{m_2} \right) g^2 r^2_{ij} = d^2 - r^2_{ij} \) for \( g \)
- \( \delta^k r_i = g r_{ij}(t_0)/m_i \)
- \( \delta^k r_j = -g r_{ij}(t_0)/m_j \)

- \( r^*_c + \delta^k r_i \to r^*_c \)
- \( r^*_r + \delta^k r_j \to r^*_r \)
- \( r^*_{ij} = r^*_c - r^*_r \)

```if \ |r^*_{ij}| - d < tolerance \ then```
- Constraint \( k \) converged
```end\_if```
```end\_for```
```end\_while```

Now all constraints \( k = 1 \ldots l \) are satisfied

```for\_constraint \ k \ do```
- \( \delta r_c = r^*_c - r'_c \)
- \( \delta r_r = r^*_r - r'_r \)

- \( f = m \delta r_c/(dt)^2 \)
- \( \Xi = \Xi - 0.5 f r_{ij} \)

- \( r_c(t + \Delta t) = r'_c(t + \Delta t) + \delta r_c \)
- \( r_r(t + \Delta t) = r'_r(t + \Delta t) + \delta r_r \)
```end\_for```
Figure 6.2: Constrained water in lipid bilayer. This is a rendering from the trajectory of a DOPC bilayer simulation. The water molecule center of mass was constrained a set distance from the bilayer center of mass.

based on the change in position of the centers of mass. Afterwards, the velocities are recalculated taking into account constrained positions. This constitutes the general SHAKE constraint implemented in Amber.¹

6.5 Test system

A test system for the constraint method was used for validation of the constraint algorithm. The system consisted of a simple 72 DOPC lipid bilayer solvated with explicit water. A single water molecule was used for the constrained molecule. The reference molecules were all the DOPC molecules in the lipid bilayer. The system used the same simulation settings as DOPC bilayers in Lipid14 simulations as described in chapter 4. The system was simulated with and without the constraint. The constraint was applied only in the dimension normal to the bilayer with a tolerance of $10^{-6}$.

Figures 6.2 and 6.3 show the water constrained in the bilayer. In the unconstrained system, the water molecule is free to move relative to the bilayer. In the
Figure 6.3: Constrained water in lipid bilayer over time. The distance in the dimension normal to the bilayer between the water molecule center of mass and the lipid bilayer center of mass is plotted. The lipid bilayer and water system was simulated with and without the constraint.

constrained system, the water is fixed at a certain distance from the bilayer center of mass and can freely diffuse laterally. In order to calculate the diffusion coefficient and potential of mean force, it is necessary to save the constraint force during the constrained MD simulation. This is saved in a data file and may be processed after MD simulations.

6.6 Graphics processing unit acceleration

Currently, the Amber program *pmemd* is primarily implemented in the Fortran 90. However, several key portions of the code have been ported to the CUDA general purpose GPU language for NVIDIA graphics processing units. The most computationally expensive portion of MD simulations is within the evaluation of non-bonded long range forces such as those from the VDW and electrostatics interactions. Initial code optimization have focused on enhancing performance of the PME and GB force evaluations. Currently, all computationally intensive portions of the *pmemd* code are ported to GPUs and most computation is done on GPU.

Due to the intricacies of new GPU hardware, other factors of the Amber code
affect performance on this hardware architecture. One major advancement for GPU-accelerated MD was the optimization of the precision of floating point integers used for calculations for GPUs. Floating point integers can be represented with various levels of precision, usually with a corresponding tradeoff in computing performance. Because MD depends on numerical integration for many of the simulation algorithms, certain portions of calculations need higher precision for accurate accumulation of forces or other terms. Errors in the accumulation of these values may be manifested in simulations in which the total energy of the system is not conserved.

For this reason, it was important that numerical accumulators on the GPU reproduce the accuracy of previous simulations while optimizing precision. Because of the significant performance difference between single and double precision floating point integers in the GPU hardware, early GPU optimizations used a hybrid precision model. Single point floating point integers were used for most numbers, while double precision was used accumulators. A later precision model included tuned fixed point precision for accumulators.

GPU performance has stemmed from several hardware features in comparison with CPUs. On a fundamental level, GPUs are designed for parallel code in which multiple instructions are run on multiple threads, whereas CPUs are optimized for sequential code running on a single thread. CPUs have been used for years for general applications that use serial logic pathways. In some cases, however, such as intensive numerical floating point calculations, GPUs benefit from their optimized architecture. Basic performance gains can be obtained through algorithms that take advantage of data parallelism.

Another major contribution to GPU performance is the increased memory bandwidth of GPUs in relation to CPUs. GPU memory bandwidth has led CPU memory bandwidth for several years and continues to grow. While memory band-
Figure 6.4: GPU memory architecture. Threads are shown as colored arrows. Thread registers are marked with “R” and thread local memory is marked with “LM.”Blocks are the red rectangles marked with a “B” and block shared memory is marked with a “SM.” Grids are the blue rectangles and global memory is marked with a “GM.”
width is fast relative to CPUs, it is still a major issue in accelerated application performance. The GPU contains several cache memories for the bandwidth requirements of accessing system memory, but these are still relatively small caches. Applications may achieve the highest level of performance by optimizing high-latency memory transfers from system memory. The basic graphics processing unit memory architecture is shown in figure 6.4. One method to alleviate this is to interleave portions of calculations on the GPU with memory downloads and uploads.

6.7 Performance of constrained molecular dynamics

A main issue with the method presented here to evaluate permeability through membranes is that a large amount of sampling is necessary to calculate average forces along the constrained dimension. In order to sample the constrained molecular dynamics along a certain dimension, it is necessary to have a great deal of simulation data. While there are challenges with obtaining a large amount of sampling, sampling along each coordinate in the constrained dimension is inherently parallel. Each simulation is independent and can be simulated in any order.

However, the performance of an MD simulation is significantly impacted when one or more constraints is added. The performance penalty is especially severe if the GPU waits for calculations on the CPU. This method is potentially useful for comparing many different molecules crossing the membrane but a great deal of sampling is required for each molecule. Any significant performance limitations hinder the usefulness of this approach.

Given the nature of massively parallel architectures in computer hardware such as graphics processing units, there are several relatively simple modifications to
the implementation of the general SHAKE distance constraint algorithm that may significantly increase performance. In the following section, several optimizations to the algorithm are discussed in greater detail.

6.8 Specific optimizations for constrained molecular dynamics

6.8.1 Upload and download

The initial implementation of a generalized SHAKE constraint within the Amber code base is within the serial CPU portion of the code. That portion of the code normally executes only on CPUs and is limited by the performance of the CPUs. The most obvious optimization for the constrained molecular dynamics code would be to integrate the algorithm into the GPU-accelerated code base.

Currently, the GPU-accelerated implementation of pmemd is based upon the existing Fortran 90 code. It is expanded with a C++ library that contains GPU functions and wrappers for CUDA code. This means that data must be first uploaded to the GPU memory and later downloaded to system memory after all calculations are complete. For optimal performance, data is stored and processed on the GPU memory during execution and transfers between CPU memory and GPU memory are minimized. Amber’s GPU code executes mainly on the GPU with some exceptions.\textsuperscript{56,57,58}

For new portions of CPU code, it is possible to temporarily download and upload GPU data structures for computation on the CPU. This was done in the first implementation of the general SHAKE constraint code described in this chapter. The algorithm requires the positions from the previous time step and the updated
positions, as well as the molecular virial. After constraints have all been satisfied, the positions are updated. The virial is also updated to remove the constraint force from its calculation. Finally, velocities are updated to reflect the constrained positions.

Transferring coordinates from system to GPU memory and visa versa is easily accomplished through upload and download functions.\textsuperscript{1} These functions were previously used in the original GPU implementation of \textit{pmemd} and are defined \texttt{gpu_download_crd} and \texttt{gpu_upload_crd} in the source file \textit{gpu.cpp}. The functions automatically download and upload coordinates into the GPU data structures taking into account the reorganization of coordinate array due to imaging and neighbor lists. This approach allows for GPU performance on most of the code coupled with new CPU functionality.

### 6.8.2 Center of mass

Following preliminary profiling of the constraint code, it was observed that a lengthy portion of the algorithm is the calculations of the center of mass of group of atoms. The calculation of the center of mass each molecular group is defined as

\[
\mathbf{r}_{cm} = \frac{1}{M} \sum_{i=1}^{n} m_i \mathbf{r}_i
\]  \hspace{1cm} (6.32)

As this is essentially an accumulation operation, it is easily parallelizable. CUDA utilizes a single-instruction multiple thread (SIMT) approach to parallel operations.\textsuperscript{212} Thus, it is straightforward for a CUDA kernel to distribute the array operations to multiple threads on the GPU and accumulate the weighted center of masses. A similar CUDA kernal was implemented for center of mass calculations for restraints in Amber.\textsuperscript{1}
6.8.3 Simultaneous constraint evaluation

The SHAKE constraint method is described in algorithm 6.2. The main loop of the constraint cannot be parallelized, but the inner loop over the constraints can.

It is possible to use the parallel architecture of GPUs to accelerate the algorithm. A simple approach is to launch a thread for each constraint simultaneously on the GPU. To implement this in CUDA, a single kernel may be written for the constraint, and then the kernel may be launched simultaneously for all constraints. Figure 6.5 shows the basic approach for this algorithm. Unless there is a massive number of constraints, the GPU can work on each constraint simultaneously for each iteration of the main SHAKE loop. This is feasible and implementable to significantly increase constraint performance.

6.9 Optimized constraint methods

Given the new membrane model in Amber, it would be valuable to simulate membrane-drug interactions and permeation. The inhomogeneous solubility-diffusion
model provides a framework for estimates of drug permeability. However, the general constraint MD methods still suffer from slow performance in Amber. GPU acceleration can significantly increase MD performance for this type of constraint. However, this all depends on the theory, algorithms, and implementation of the method to achieve optimal performance. With the implementation of this method, it is now possible to compare permeability of potential drugs across membranes efficiently in Amber.

6.10 Acknowledgement

Chapter 6 is currently being prepared for submission for publication of the material. Madej, B. D., Dickson, C. J., Gould, I. R., Walker, R. C. The dissertation/thesis author was the primary investigator and author of this material.
Chapter 7

Conclusions

Molecular dynamics is an advanced computational method for simulating lipid bilayers and cellular membranes that needs a force field for each new molecule. Developing such a robust force field for new molecules is not trivial.

For the first time in Amber, an accurate lipid force field for bilayer and membrane simulations has been developed and validated.\textsuperscript{4} The force field is built on a modular framework such that many different types of lipids may be simulated.\textsuperscript{3} Parameters were refined for MD simulations without constant surface tension with anisotropic pressure coupling. The lipid parameters may be easily combined with the other pairwise additive Amber force fields.

In order to investigate complex lipid bilayer types, such as those in the liquid-ordered phase, it was necessary to parameterize other molecules as well. Amber now includes a parameter set for cholesterol that is able to accurately simulate mixed bilayer types.\textsuperscript{213} This gives some insight into the nature of liquid-ordered bilayers. Complex membrane simulations are now possible in Amber MD.

With the implementation of an optimized constraint method in Amber, the potential of mean force and permeation of small molecules across the membrane
Figure 7.1: Extended lipid force field. Additional head group parameters have been developed for charged and zwitterionic head groups. Polyunsaturated fatty acid tails have also been parameterized. Sphingomyelin parameters have also been refined.

can be efficiently calculated.\textsuperscript{214} Given that the permeability largely depends on the interaction of the small molecule with the membrane, a highly accurate force field is necessary. The relative permeabilities for a set of small molecule drugs can be compared for many membrane types. The mechanism of transport across diverse membranes may indicate where molecules localize in cells.

7.1 Related projects

Several projects have stemmed from the initial work presented in this dissertation. These related projects are briefly discussed in the following sections.

7.1.1 Extension of Lipid14

Even with this initial work, there are many more types of lipids in the milieu of cell membranes. In particular, several key lipid types are also important in membrane dynamics.

As previously discussed, lipid bilayers in the liquid-ordered phase and raft assemblies in membranes usually are composed of phospholipids, cholesterol, sphin-
gomyelin, and raft proteins. With parameters for phospholipids and cholesterol already developed, the next necessary component is sphingomyelin. Sphingomyelin has a different structure consisting of a sphingosine backbone. However, the phosphate, head group, and fatty acid tails are similar to glycerophospholipids. Sphingomyelin has been observed to have several variants of saturated fatty acids on one of the tails. It is possible to construct a sphingomyelin model with interchangeable tail groups. Additionally, the amide region is different from the glycerol group and needs appropriate parameters for that region. Parameters will be validated with actual bilayer simulations.

Several other components are important for diverse bilayer types. For instance, there are other charged and zwitterionic head groups found in membranes. Headgroups include phosphatidylglycerol, phosphatidic acid, and phosphatidylserine. While Lipid11 included the topology and charges for those head groups, the parameters were not refined. Furthermore, polyunsaturated fatty acids are found in many types of phospholipids.\textsuperscript{215,216} In particular, docosahexanoic acid contains four double bonds within the fatty acid. As observed in previous simulations, unsaturated lipid tails influence bilayer structure due to their tail conformations.

7.1.2 All-atom self-assembly of lipid bilayers

Lipids are generally found in aggregated structures such as bilayers or vesicles rather than in solution.\textsuperscript{12} Phospholipids in an aqueous environment spontaneously associate to form these structures. Observations of lipids in solution indicate that it is energetically favorable to form bilayer conformations. The polar head groups interact with the aqueous solvent, while the nonpolar tails self-aggregate within the interior of the bilayer.

It was hypothesized that lipids would spontaneously assemble into a bilayer
The time scales of formation are generally on the order of hundreds of nanoseconds. More self-assembled lipid bilayers have been simulated including the DOPS bilayer shown in figure 7.2.

The self-assembly of lipid bilayers with the current force field is an interesting validation of the parameter set. Bilayers formed from self-assembly were well-formed and matched available experimental properties. Self-assembly provides an alternative method for predicting lipid bilayer structure.

### 7.1.3 Membrane proteins

Another obvious direction for membrane simulations is to simulate membrane bound proteins. This dissertation establishes the necessary framework for membrane protein simulations.
Figure 7.3: Potassium channel structure. The KCSA protein (PDB: 1K4C) is shown in a POPC bilayer.

It is still relatively difficult to construct de novo protein and membrane structures. Several programs and services can build structures, but often suffer from simulation artifacts. Figure 7.3 shows an example of a membrane constructed around an integral membrane protein, the potassium channel KCSA. Advanced tools for building membranes will be necessary for this type of simulation. Ideally, this type of program should be able to build any membrane structure. The orientation of membrane proteins should be calculated as well. It will be necessary to calculate the effect of membrane components (like phospholipids, cholesterol, and sphingomyelin) on protein dynamics.

7.2 The future

Lipid and membrane simulations are improving membrane models. A combination of improvements in experimental technologies as well as computational power
will enable a more detailed description of membranes. Higher resolution experiments will enable a clearer picture of membrane structure. Faster computers will undoubtedly allow for larger or more detailed membrane systems. Fluency in both computational and experimental approaches will be crucial in understanding these incredibly detailed and complex molecular machines.
Appendix A

Lipid11 appendix

The full structures of all Lipid11 residues are included in figure A.1
Lauric acid
(12:0)

Myristic acid
(14:0)

Palmitic acid
(16:0)

Stearic acid
(18:0)

Oleoic acid
(18:1n-9)

Linoleic acid
(18:2n-6)

Linolenic acid
(18:3n-3)

Arachidonic acid
(20:4n-6)

Docosahexanoic acid
(22:6n-3)

Phosphatidylcholine

Phosphatidylethanolamine

Phosphatidylserine

Phosphatidic acid (PO4-)

Phosphatidic acid (PO42-)

Phosphatidyglycerol

Phosphatidylinositol

**Figure A.1:** Lipid11 residues. Full chemical structures for Lipid11 residues. Cholesterol not shown.
References

[1] Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Gohlke, H.; Goetz, AW; Gusarov, S; Homeyer, N; Janowski, P; Kaus, J; Kolossvary, I; Kovalenko, A; Lee, TS; LeGrand, S; Luchko, T; Luo, R; Madej, B; Merz, KM; Paesani, F; Roe, DR; Roitberg, A; Sagui, C; Salomon-Ferrer, R; Seabra, G; Simmerling, CL; Smith, W; Swails, J; Walker, RC; Wang, J; Wolf, RM; Wu, X; and Kollman, PA. *Amber 14*. 2014.


[67] Morse, P. M. Physical Review 1929. 34 (1), 57–64.


[88] Morse, P. M. and Feshbach, H. *Methods of theoretical physics* **1953.**


[101] Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, AW; Kolossvary, I; Wong, KF; Paesani, F; Vanicek, J; Wolf, RM; Liu, J; Wu, X; Brozell, SR; Steinbrecher, T; Gohlke, H; Cai, Q; Ye, X; Wang, J; Hsieh, M-J; Cui, G; Roe, DR; Mathews, DH; Seetin, MG; Salomon-Ferrer, R; Sagui, C; Babin, V; Luchko, T; Gusarov, S; Kovalenko, A; and Kollman, PA. *Amber 12* **2012**.


[129] Lafleur, M.; Cullis, P. R.; Fine, B.; and Bloom, M. *Biochemistry* 1990. 29 (36), 8325–8333.


