

EXPLORING THE STRUCTURE OF THE STRATUM CORNEUM LIPID  
MATRIX VIA MOLECULAR DYNAMICS SIMULATIONS

By

Timothy C. Moore

Dissertation

Submitted to the Faculty of the  
Graduate School of Vanderbilt University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

May 11, 2018

Nashville, Tennessee

Approved:

Clare M<sup>c</sup>Cabe, Ph.D.

Peter T. Cummings, Ph.D.

Paul E. Laibinis, Ph.D.

Anne K. Kenworthy, Ph.D.

To my parents, for always believing in me.

## ACKNOWLEDGMENTS

I have been fortunate to encounter many individuals who have made this work possible, and to whom I owe much gratitude. First and foremost, I thank Clare McCabe for providing funding, resources, and opportunities that have allowed me to grow into the scientist I am today. Chris Iacovella played a major role in helping me get started in the lab, and I have enjoyed many fruitful discussions with him that have strengthened this research. I have been fortunate to work with Annette Bunge, whose expertise on skin has helped put my research into a larger context. I owe many thanks to the participants of the Gordon Research Conference on the Barrier Function of Mammalian Skin, especially David Moore, Joke Bouwstra, Reinhard Neubert, and Rich Mendelsohn.

The members of the McCabe and Cummings research groups have provided invaluable support throughout graduate school. Remco Hartkamp worked closely with me for two years, and my writing and approach to research are significantly stronger as a result of his mentorship. For that, I am truly grateful. Christoph Klein worked with me in developing software used in my research, and this collaboration has led to more robust and reproducible work. Other members of the McCabe and Cummings groups have made my time both more enjoyable and more productive; I would like to specifically thank Andrew Summers, Matt Thompson, Gaurav Das, Will French, Alex Yang, and Justin Gilmer.

I am endlessly grateful for each of these individuals; I would not be who I am today without them.

## TABLE OF CONTENTS

	Page
DEDICATION.....	ii
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES .....	viii
LIST OF FIGURES.....	x
Chapter	
1 INTRODUCTION .....	1
1.1 Bibliography .....	4
2 SKIN BACKGROUND.....	6
2.1 Bibliography .....	14
3 SKIN SIMULATION BACKGROUND .....	18
3.1 Simulation Methodology .....	19
3.2 Force Fields .....	20
3.3 SC Lipid Simulation Specifics.....	23
3.4 Atomistic Molecular Dynamics Simulations of SC Lipid Systems.....	24
3.5 Coarse-Grained Molecular Dynamics Simulations of SC Lipid Systems .	28
3.6 Bibliography .....	30
4 THE EFFECT OF CERAMIDE TAIL LENGTH ON MODEL MEMBRANE PROPERTIES .....	35
4.1 Introduction.....	35
4.2 Material and Methods .....	38
4.2.1 Model.....	38
4.2.2 Methods .....	39
4.2.3 Analysis .....	43
4.3 Results.....	43
4.3.1 Random Walk Molecular Dynamics .....	43
4.3.2 Dependence of Structural Properties on Bilayer Composition....	50
4.4 Discussion .....	59
4.5 Conclusions.....	63

4.6	Bibliography .....	64
5	DEVELOPMENT OF COARSE-GRAINED FORCE FIELDS VIA MULTI-STATE ITERATIVE BOLTZMANN INVERSION .....	70
5.1	Introduction .....	70
5.2	Method and Simulation Details .....	73
5.2.1	Single State Iterative Boltzmann Inversion .....	73
5.2.2	Multistate Extension of IBI .....	74
5.2.3	Simulation Model .....	77
5.3	Results .....	79
5.3.1	Monatomic Lennard-Jones Fluid .....	79
5.3.2	Propane .....	80
5.3.3	<i>n</i> -dodecane .....	84
5.3.4	Tuning CG Force Fields .....	86
5.4	Conclusions .....	88
5.5	Bibliography .....	91
6	DEVELOPMENT OF A COARSE-GRAINED WATER FORCE FIELD VIA MULTISTATE ITERATIVE BOLTZMANN INVERSION .....	97
6.1	Introduction .....	97
6.2	Methods .....	100
6.2.1	<i>k</i> -means Clustering Algorithm .....	100
6.2.2	Multistate Iterative Boltzmann Inversion Method .....	101
6.2.3	Models .....	102
6.3	Results and Discussion .....	103
6.3.1	Modified Chiu Potential .....	103
6.3.2	Potential Derivation and Validation .....	104
6.3.3	Validation and Comparison to Other Models .....	108
6.4	Conclusions .....	112
6.5	Bibliography .....	114
7	DEVELOPMENT OF A COARSE-GRAINED MODEL OF CERAMIDE NS AND FFA .....	118
7.1	Introduction .....	118
7.2	Models and Methods .....	121
7.2.1	Atomistic Simulations .....	121
7.2.2	CG Simulations .....	126
7.2.3	CG Mappings .....	126
7.2.4	CG Force Field Derivation .....	128
7.2.5	Analysis .....	133
7.3	Results .....	133
7.3.1	Pure CER .....	134

7.3.2	Optimization of Lipid–Water Interactions .....	142
7.3.3	Validation .....	147
7.4	CER NS C24 Self-Assembly.....	150
7.5	Conclusion.....	156
7.6	Bibliography .....	158
8	DEVELOPMENT OF THE CROSS-INTERACTIONS FOR COARSE- GRAINED MODELS OF CER NS AND FFA .....	165
8.1	Introduction.....	165
8.2	Methods and Model .....	167
8.3	Simulations.....	167
8.3.1	Model Description and Force Field Optimization .....	169
8.3.2	Analysis .....	171
8.4	Results and Discussion .....	171
8.4.1	Force Field Derivation and Validation .....	171
8.4.2	Self-assembly Into Lamellar Structures .....	174
8.5	Conclusion.....	177
8.6	Bibliography .....	178
9	DEVELOPMENT OF COARSE-GRAINED MODELS FOR A MINIMAL STRATUM CORNEUM LIPID MIXTURE.....	182
9.1	Introduction.....	182
9.2	Models and Methods .....	183
9.2.1	Atomistic Simulations .....	183
9.2.2	CG Simulations .....	184
9.2.3	CG Mappings .....	188
9.2.4	CG Force Field Description and Optimization .....	189
9.3	Results.....	192
9.3.1	Forcefield Derivation .....	192
9.3.2	Force Field Validation.....	196
9.3.3	Self-Assembly Into Lamellar Structures.....	198
9.4	Bibliography .....	203
10	CONCLUSIONS AND OUTLOOK .....	205
10.1	Recommended Future Work .....	207
10.2	Bibliography .....	209
Appendices		
A	Analysis .....	210
A.1	General Structural Properties.....	210

A.2	Contact Angle.....	212
A.3	Hydrogen Bonding .....	213
A.4	Bibliography .....	214
B	Appendix to Chapter 4 .....	215
B.1	Structural Properties .....	215
C	Appendix to Chapter 7.....	220
C.1	4-Site CER NS Headgroup RDFs and Pair Potentials .....	220
C.2	3-Site CER NS Headgroup RDFs and Pair Potentials .....	229

## LIST OF TABLES

Table	Page
4.1	CHOL–CHOL coordination number over the final 50 ns of MD simulation. 46
4.2	Area per lipid of systems as a function of initial configuration and equilibration methodology. .... 47
4.3	Total number of hydrogen bonds in each system. .... 57
5.1	Structural properties of monolayers..... 88
6.1	Density of water calculated with different models. .... 108
6.2	Diffusion of CG water. .... 112
7.1	States used in the pair potential optimizations for pure CER NS and FFA and their interactions with water. .... 123
7.2	Weights given to each state in the pair potential optimizations. .... 124
7.3	Parameters for bond-stretching and angle-bending potentials. .... 130
7.4	In-plane hydrogen bonding in the atomistic, dehydrated CER NS C24:0 bilayer..... 138
7.5	Structural properties of dehydrated CER NS bilayers..... 141
7.6	Structural properties of CER NS C16/CER NS C24 bilayers at 305 K and 1.0 atm simulated using the CG models derived here and the CHARMM36-CER all-atom model (AA)..... 148
7.7	Properties of self-assembled bilayers as a function of expansion and compression time. .... 152
8.1	Comparison of structural properties of CER NS–FFA bilayers simulated using atomistic and CG models. .... 174
9.1	States used in the pair potential optimizations for CHOL. .... 185

9.2	Weights given to each state in the pair potential optimizations. ....	186
9.3	Parameters for bond-stretching potentials in CHOL. ....	190
9.4	Parameters for angle-bending potentials in CHOL. ....	191
9.5	Comparison of structural properties of two-component bilayers simulated with the CG and atomistic models. ....	200
9.6	Percentage of CER molecules adopting extended conformations. ....	202
B.1	Hydrogen bonds between lipid components for the ternary system with a 1:1 eCER:uCER ratio.....	215
B.2	Hydrogen bonds between lipid components for the 2:1 CER NS:CHOL system with a 1:1 eCER:uCER ratio. ....	216
B.3	Hydrogen bonds between lipid components for the 1:1 CER NS:CHOL system with a 1:1 eCER:uCER ratio. ....	217

## LIST OF FIGURES

Figure	Page
2.1 Layers of the epidermis. ....	8
2.2 Intercellular lipids in the stratum corneum. ....	9
2.3 Chemical structure of the ceramides present in human stratum corneum. ....	10
2.4 A schematic of the sandwich model of Bouwstra. ....	12
3.1 Number of SC simulation papers published per year as a function of time. ....	19
3.2 A schematic of the various potential energy terms used in molecular dynamics simulation. ....	22
4.1 Chemical structure of the molecules simulated. ....	39
4.2 A representative plot of temperature versus time during the RWMD simulation. ....	41
4.3 System configurations and CHOL-CHOL CN as a function of equilibration. ....	45
4.4 CER backbone rotations. ....	48
4.5 Evolution of the CHOL-CHOL coordination number as a function of equilibration methodology. ....	49
4.6 Tilt angle and in-plane area of lipid bilayers. ....	51
4.7 Mass density profiles of lipids across the various bilayer systems. ....	53
4.8 Bilayer thickness versus low density tail thickness. ....	54
4.9 Density profiles of specific molecular groups. ....	56
4.10 Comparison of the neutron scattering length density (NSLD) profiles from simulation and experiment. ....	61
5.1 Regions of good potential phase space. ....	73
5.2 RDFs and potentials derived for the LJ system. ....	81

5.3	RDFs and potential derived for the single-site CG model of propane using MS IBI. ....	82
5.4	Different initial guesses and the resulting derived potentials for propane optimizations. ....	83
5.5	Comparison of a structural metric between the CG and UA models of <i>n</i> -dodecane. ....	85
6.1	Schematic of the <i>k</i> -means algorithm. ....	101
6.2	RDFs from simulations using the original and modified Chiu potentials. ....	105
6.3	RDFs and potentials from the MS IBI potential derivation. ....	107
6.4	$f_{\text{fit}}$ from Equation 5.5 as a function of iteration in the pair potential optimization. ....	108
6.5	Interaction potentials from various CG models of water. ....	109
6.6	System snapshots during the droplet simulations. ....	110
6.7	Configurations from simulations in the presence of a nucleation site with the MARTINI and MS IBI models. ....	111
6.8	RDFs of the MARTINI model and the atomistic TIP3P model mapped to the CG level for the bulk NVT state. ....	112
7.1	CG mappings for CER NS and FFA. ....	125
7.2	TAIL-TAIL RDFs and pair potential. ....	135
7.3	RDFs and pair potentials for CER headgroup beads. ....	137
7.4	The effect of weights given to each state on the AMIDE-AMIDE RDFs and pair potentials. ....	140
7.5	TAIL-WATER RDFs and pair potential. ....	144
7.6	Snapshots of FFA in water with different force fields. ....	145
7.7	Snapshots from the droplet simulations. ....	146
7.8	Orientation of MHEAD2-OH vectors in the CER NS C24 bilayer. ....	150

7.9	Self-assembly of CER NS C24 into an ordered bilayer. ....	152
7.10	Self-assembly of pure CER NS C24 into a multilayer structure. ....	154
7.11	Angle and end-to-end distance between the two tails of CER NS in single molecule vacuum simulations. ....	156
8.1	CG mappings for FFA and CER NS.....	169
8.2	RDFs and derived pair potentials for CER NS–FFA headgroup beads at the NPT states. ....	172
8.3	RDFs and derived pair potentials for CER NS–FFA headgroup beads at the NVT states.....	173
8.4	Snapshots of the CER NS–FFA bilayer self-assembly process. ....	175
8.5	Snapshots of the CER NS–FFA multilayer self-assembly process. ....	177
9.1	CG mapping for CHOL.....	189
9.2	RING–RING RDFs and pair potential. ....	193
9.3	TAIL–RING RDFs and pair potential.....	196
9.4	CHEAD–FHEAD RDFs and pair potential. ....	197
9.5	TAIL–CHEAD RDFs and pair potential. ....	198
9.6	CERHEADs–CHEAD RDFs and pair potential. ....	199
9.7	A self-assembled multilayer composed of equimolar amounts of CER NS C24, CHOL, and FFA C24:0. ....	201
9.8	Inverse micellar structures in the system with 62.5 mol % CHOL. ....	202
A.1	Schematic of bilayer thickness calculation.....	212
B.1	Nematic order parameter of the lipid tails as a function of system composition.....	215
B.2	Bilayer thickness as a function of system composition. ....	216

B.3	Total lipid density profile, and density profiles of the terminal part of the CER NS C24 fatty acid and FFA C24:0 tails.....	216
B.4	Thickness of the lipid–water interface as a function of composition.....	217
B.5	CHOL–CHOL coordination number as a function of system composition.	218
B.6	A comparison of the localization of specific groups between experiment and simulation.....	219
C.1	AMIDE–AMIDE RDFs and pair potential.....	220
C.2	AMIDE–MHEAD2 RDFs and pair potential. ....	221
C.3	MHEAD2–MHEAD2 RDFs and pair potential. ....	222
C.4	MHEAD2–OH1 RDFs and pair potential. ....	223
C.5	MHEAD2–OH2 RDFs and pair potential. ....	224
C.6	OH1–OH1 RDFs and pair potential. ....	225
C.7	OH2–OH2 RDFs and pair potential. ....	226
C.8	AMIDE–WATER RDFs and pair potential. ....	227
C.9	MHEAD2–WATER RDFs and pair potential.....	227
C.10	OH1–WATER RDFs and pair potential.....	228
C.11	OH2–WATER RDFs and pair potential.....	228
C.12	NHEAD–NHEAD RDFs and pair potential.....	229
C.13	NHEAD–MHEAD RDFs and pair potential. ....	230
C.14	NHEAD–SHEAD RDFs and pair potential. ....	230
C.15	MHEAD–MHEAD RDFs and pair potential.....	231
C.16	MHEAD–SHEAD RDFs and pair potential.....	231
C.17	SHEAD–SHEAD RDFs and pair potential.....	232

## CHAPTER 1

### INTRODUCTION

As the largest organ of the human body, the skin acts as a crucial barrier between the internal and external environments of the body. This barrier function is of vital importance for life on land, as it prevents rapid dehydration and keeps foreign, potentially hazardous particles outside the body. Owing to its critical location at the body–environment interface, a thorough understanding of how the skin barrier manifests itself is important for both restoring and selectively reducing it.<sup>1</sup> Consider, for example, that many skin diseases can result in an impaired skin barrier, coincidental with abnormalities in lipid composition, organization, and structure.<sup>2–11</sup> While such diseases themselves are not often viewed as particularly life-threatening, they often lead to systemic health problems beyond the original disease. Atopic dermatitis (AD, an inflammatory, itchy skin disorder that affects an estimated 30% of the United States population<sup>12</sup>) often coincides with other, more serious diseases. Up to 75% of children with AD will develop asthma,<sup>12</sup> and up to 67% of children with AD will develop rhinitis.<sup>2,13,14</sup> Ichthyosis vulgaris, a scaly skin disorder, also commonly co-occurs with AD.<sup>15,16</sup> Thus, to prevent such skin diseases from progressing to more serious health issues, treatment needs to focus on restoring the skin barrier rather than simply treating the symptoms, and hence requires a thorough, molecular-level understanding of structures that make up that barrier.

Understanding the skin barrier on a molecular level is not only important for restoring a defective barrier, but also for selectively reducing the barrier, *e.g.*, for transdermal drug delivery. Transdermal drug delivery is an attractive alternative to oral and hypodermic drug delivery for a myriad of reasons, including the avoidance of first-pass metabolism, the lack of pain associated with the transdermal route compared

to a needle, the reduction of dangerous medical waste (*e.g.*, contaminated needles), the ability for self-administration, and the potential for a slow-release mechanism.<sup>17</sup> Despite these advantages, candidate drugs for transdermal delivery must be small, have a low effective dose, and be highly lipophilic. The use of penetration enhancers to temporarily reduce the skin's barrier properties could have a major impact on transdermal drug delivery, but is still relatively underutilized to date. A clear understanding of the molecular-level details of the skin barrier would allow penetration enhancers to be tailored to have specific properties, making the transdermal route possible for a wider variety of drugs.

While much has been learned about the molecular origins of the skin barrier, experimental difficulties make a full molecular-level understanding elusive. One of the major limitations to experimental studies on the skin barrier is the resolution available to scattering and diffraction methods, making it difficult to fully resolve the structures involved. In this regard, molecular simulation becomes a useful tool for studying these types of systems, as molecular-level resolution is readily available. However, the full potential of molecular simulation for studying skin barrier systems has yet to be realized, owing to a number of challenges related to the proper sampling of these systems.

This thesis describes work done on two fronts. First, the challenges hindering progress in skin barrier research are addressed. Next, the methods developed are applied to simulate systems resembling the skin barrier to glean molecular-level details of its structure.

In Chapter 2, background information related to the skin barrier, including results from experimental studies, is presented. In Chapter 3, background information related to simulation studies of SC lipids is given, including the basics of simulation methodology and a brief review of the SC simulation literature. In Chapter 4, a novel equilibration methodology for equilibrating skin barrier systems is presented and

applied to such systems. In Chapter 5, a novel method for developing coarse-grained (CG) force fields, multistate iterative Boltzmann inversion (MS IBI), is presented, and its utility illustrated. In Chapter 6, MS IBI is applied to develop a CG model of water; this model is shown to be more accurate than previously-used CG water models. In Chapters 7 to 9, MS IBI is applied to develop CG models for key components of the human skin barrier. These models are validated against atomistic simulation data, and their self-assembly is explored. Finally, in Chapter 10, the findings presented in this thesis are summarized, and future directions for extending this work are suggested.

## 1.1 Bibliography

- [1] WERTZ, P. W. Lipids and barrier function of the skin. *Acta Dermato-Venereologica Supp* **208**, 2000, pp. 7–11.
- [2] SEGRE, J. A. Epidermal barrier formation and recovery in skin disorders. *Journal of Clinical Investigation* **116**, 2006, p. 1150.
- [3] ELIAS, P. M., HATANO, Y., AND WILLIAMS, M. L. Basis for the barrier abnormality in atopic dermatitis: outside-inside-outside pathogenic mechanisms. *Journal of Allergy and Clinical Immunology* **121**, 2008, pp. 1337–1343.
- [4] ELIAS, P. M., WILLIAMS, M. L., HOLLERAN, W. M., JIANG, Y. J., AND SCHMUTH, M. Thematic review series: skin lipids. Pathogenesis of permeability barrier abnormalities in the ichthyoses: inherited disorders of lipid metabolism. *Journal of Lipid Research* **49**, 2008, pp. 697–714.
- [5] CHOI, M. J. AND MAIBACH, H. I. Role of ceramides in barrier function of healthy and diseased skin. *American Journal of Clinical Dermatology* **6**, 2005, pp. 215–223.
- [6] YAMAMOTO, A., SERIZAWA, S., ITO, M., AND SATO, Y. Stratum corneum lipid abnormalities in atopic dermatitis. *Archives of Dermatological Research* **283**, 1991, pp. 219–223.
- [7] DI NARDO, A., WERTZ, P. W., GIANNETI, A., AND SEIDENARI, S. Ceramide and cholesterol composition of the skin of patients with atopic dermatitis. *Acta Derm Venereol (Stockh)* **78**, 1998, pp. 27–30.
- [8] MATSUMOTO, M., UMEMOTO, N., SUGIURA, H., AND UEHARA, M. Difference in ceramide composition between “dry” and “normal” skin in patients with atopic dermatitis. *Acta Dermato-Venereologica* **79**, 1999, pp. 246–247.
- [9] IMOKAWA, G., ABE, A., JIN, K., HIGAKI, Y., KAWASHIMA, M., AND HIDANO, A. Decreased level of ceramides in stratum corneum of atopic dermatitis: an etiologic factor in atopic dry skin? *Journal of Investigative Dermatology* **96**, 1991, pp. 523–526.
- [10] PILGRAM, G. S. K., VISSERS, D. C. J., VAN DER MEULEN, H., KOERTEN, H. K., PAVEL, S., LAVRIJSEN, S. P. M., AND BOUWSTRA, J. A. Aberrant lipid organization in stratum corneum of patients with atopic dermatitis and lamellar ichthyosis. *Journal of Investigative Dermatology* **117**, 2001, pp. 710–717.
- [11] MOTTA, S., MONTI, M., SESANA, S., CAPUTO, R., CARELLI, S., AND GHIDONI, R. Ceramide composition of the psoriatic scale. *Biochimica et Biophysica Acta - Molecular Basis of Disease* **1182**, 1993, pp. 147–151.
- [12] NIAMS. Handout on health: Atopic dermatitis (2003).

- [13] BOGUNIEWICZ, M., LEUNG, D. Y. N., ORANGE, J., PINZONE, K., HARBECK, R., LEUNG, S. B., AND SPERGEL, J. Total and specific IgE are elevated in Atopic March (AM). *Journal of Allergy and Clinical Immunology* **121**, 2008, p. S31.
- [14] SPERGEL, J. M. Atopic march: link to upper airways. *Current Opinion in Allergy and Clinical Immunology* **5**, 2005, pp. 17–21.
- [15] MCGRATH, J. A. Filaggrin and the great epidermal barrier grief. *Australasian Journal of Dermatology* **49**, 2008, pp. 67–74.
- [16] MCGRATH, J. A. AND UITTO, J. The filaggrin story: novel insights into skin-barrier function and disease. *Trends in Molecular Medicine* **14**, 2008, pp. 20–27.
- [17] PRAUSNITZ, M. R. AND LANGER, R. Transdermal drug delivery. *Nature Biotechnology* **26**, 2008, pp. 1261–1268.

## CHAPTER 2

### SKIN BACKGROUND

In human physiology, skin plays the vital role of separating the inside of the body from its environment, protecting the body from pathogen invasion and dehydration. The skin broadly comprises two layers, the dermis and the epidermis. The dermis is the innermost of the two layers, and contains the connective tissues—collagen and elastin—that give the skin its strength and elasticity. The dermis also contains nerve endings, hair follicles, sweat glands, sebaceous glands, and blood vessels, and is hence very much connected with the rest of the body. The blood vessels provide nourishment and waste removal for the dermal and epidermal cells. Therefore, once a substance passes into the dermis, it may enter systemic circulation; this mechanism forms the basis of transdermal drug delivery.

The epidermis is the outermost, visible layer of skin, and itself consists of several layers, or *strata*, as shown in Figure 2.1. The innermost layer at the interface of the dermis and epidermis is the *stratum basale*, which is a 1–3 cell-thick layer of basal keratinocytes. On top of the *stratum basale* is the *stratum spinosum*, where the keratinocytes start to connect to one another through desmosomes. Also in this layer, the keratinocytes begin to produce lamellar bodies containing glycosphingolipids, sterols, and phospholipids, as well as catabolic enzymes. In the *stratum granulosum*, the keratinocytes lose their nuclei and contain protein-enriched granules, which serve to bind keratin filaments to form a keratin network. In the upper *stratum granulosum*, the lamellar bodies are secreted by the keratinocytes. After secretion, the catabolic enzymes convert the glycosphingolipids into ceramides and the phospholipids into free fatty acids. The lipids are then arranged into lamellae parallel to the cell surface, forming a lipid envelope around the keratinocytes. In the *stratum corneum* (SC), the

keratinocytes become terminally differentiated into corneocytes, and are full of keratin filaments. The corneocytes are flat, polyhedral, nonviable cells that contain a protein envelope on the surface, to which the lipid envelope is attached. The corneocytes are stacked, usually between 10–40 layers thick, and the intercellular space of  $\sim 100$  nm is composed of a highly organized lipid matrix, as shown in Figure 2.2. This arrangement of lipids surrounding the corneocytes is often referred to as a brick and mortar-like arrangement, where the corneocytes serve as the bricks and the lipid matrix serves as the mortar. Because of its unique organization, the barrier properties of skin are mainly localized to the SC.<sup>1</sup> It remains unclear whether or not transcellular diffusion plays a role in trans-SC diffusion,<sup>2,3</sup> but the lipid phase presents the only continuous path through the SC. Therefore, it is widely thought that a functioning skin barrier relies on a properly organized SC lipid matrix.<sup>4-6</sup>

The composition of the SC lipid matrix has been well characterized, and consists of three main classes of lipids: the ceramides (CERs), cholesterol (CHOL), and free fatty acids (FFAs).<sup>1</sup> These lipids are present in 1:1:1 to 2:2:1 molar ratios, although this composition varies, even within a given specimen. Ceramides are a class of two-tailed sphingolipids consisting of a sphingoid base linked to a fatty acid chain through an amide bond. Physiologically, CERs have been implicated in a number of biological processes, including cellular apoptosis and trafficking.<sup>8</sup> In the SC, CERs originate in the *stratum granulosum*, where the large, polar glucosyl headgroup of glucosylceramide is enzymatically cleaved to form CER. To date, 15 classes of CERs have been identified in skin, differing in headgroup chemistry and fatty acid composition.<sup>9</sup> Motta *et al.* proposed a nomenclature based on the composition of the fatty acid and sphingoid base,<sup>10</sup> and this nomenclature is explained graphically in Figure 2.3. The fatty acid is mostly either nonhydroxy (N) or  $\alpha$ -hydroxy (A). A small subset of the CERs, the acylceramides, contain a linoleic acid esterified to the  $\omega$ -carbon of the fatty acid chain (EO). The sphingoid base is either sphingosine (S), dihydrosphingosine (dS),

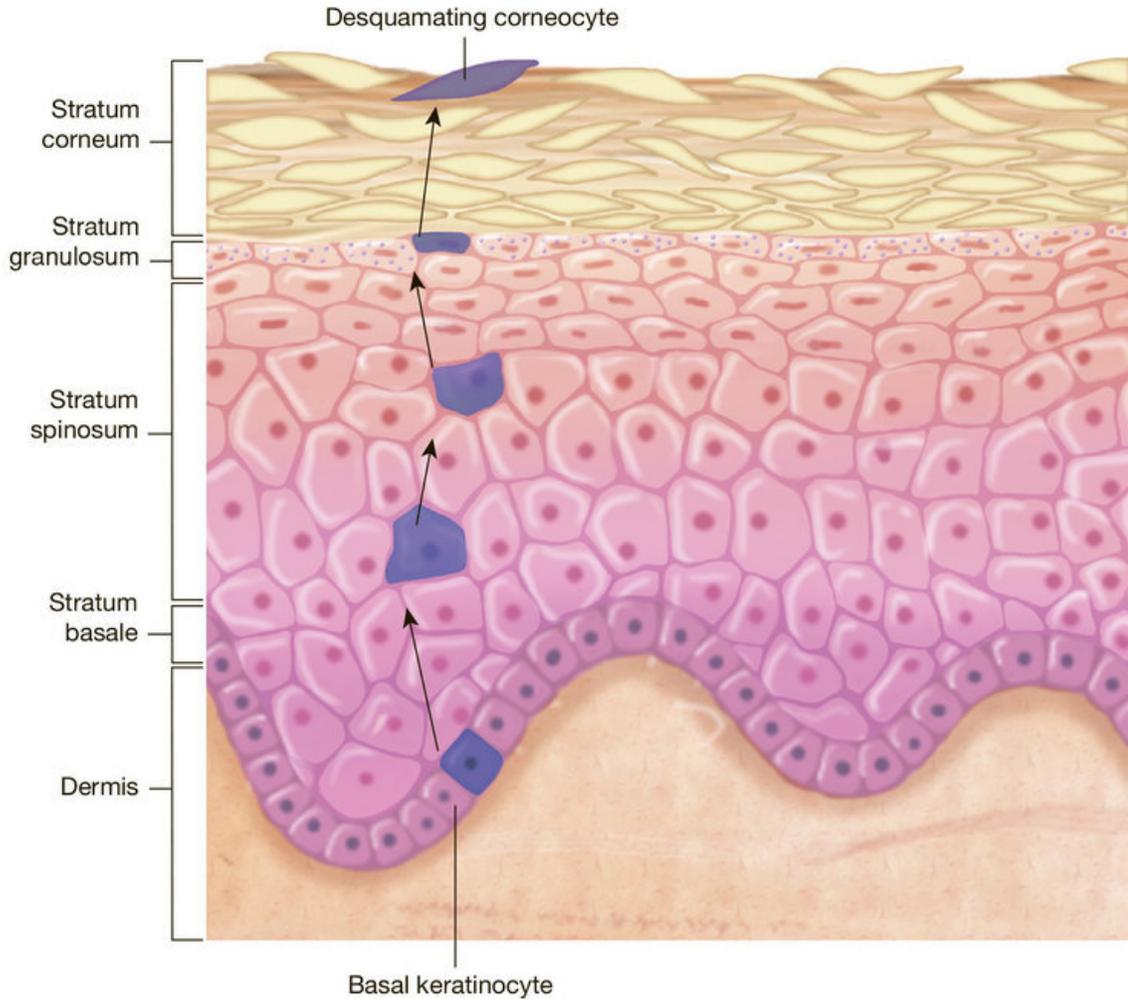


Figure 2.1: Layers of the epidermis. Reproduced from Reference 7 with permission from Wolters Kluwer.

phytosphingosine (P), or 6-hydroxysphingosine (H). Using this nomenclature, CER NS C24 denotes a CER consisting of a sphingosine base linked to a nonhydroxy fatty acid that is 24 carbons in length (*i.e.*, lignoceric acid). While the fatty acid tails generally show a polydisperse tail length distribution, with 22 to 26 being the most prevalent, the sphingoid base is usually 18 carbons in length. CHOL, the most abundant individual lipid in the SC,<sup>11</sup> originates in the lower epidermis, which is an active site of CHOL synthesis.<sup>12</sup> CHOL is a steroid, and thus contains a rigid ring structure. CHOL also contains a hydrophilic hydroxyl group on the A ring, a hydrocarbon tail attached to

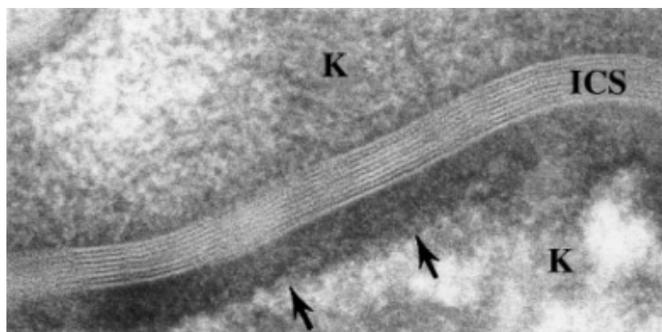


Figure 2.2: Electron micrograph showing the highly organized lipid matrix within the intercellular space (ICS) of two corneocytes (K). Reproduced from Reference 1 with permission from Elsevier.

the D ring, and two methyl groups sticking out of the alpha-face. While CHOL often rigidifies liquid-crystalline lipid bilayers,<sup>13</sup> it is thought to provide some fluidity to the SC lipid matrix, which would otherwise be rigid and perhaps brittle.<sup>14</sup> FFAs in the SC are long and mostly saturated.<sup>11</sup> In the SC, FFAs originate as phospholipids in the lamellar bodies of the *stratum spinosum*, which are enzymatically converted to FFAs in the intercellular space after secretion in the *stratum granulosum*. Like the CER fatty acid tail, the FFAs show a polydisperse tail length distribution, with 22 to 26 being the most common.<sup>11</sup>  $pK_a$  values of 10 have been reported for long, saturated FFAs,<sup>15</sup> and hence the FFAs are mostly protonated in the SC, which has a  $pH \sim 5$ .<sup>16</sup> Other minor components, such as CHOL sulfate, make up a small portion of the SC lipid matrix, but are not thought to play a major role in the barrier properties of the SC.<sup>1</sup>

Experimental studies have revealed much information regarding the organization of lipids within the SC, which are arranged in highly ordered lamellar structures. The lipids are attached to the surface of the corneocytes at the cornified lipid envelope, creating a template for lamellae oriented parallel to the cell's surface.<sup>11</sup> The lipid organization is characterized by the *lamellar organization*, which describes how the lamellae are arranged, and the *lateral organization*, which describes how the lipids are arranged within individual lamellae. Two coexisting lamellar phases are present in

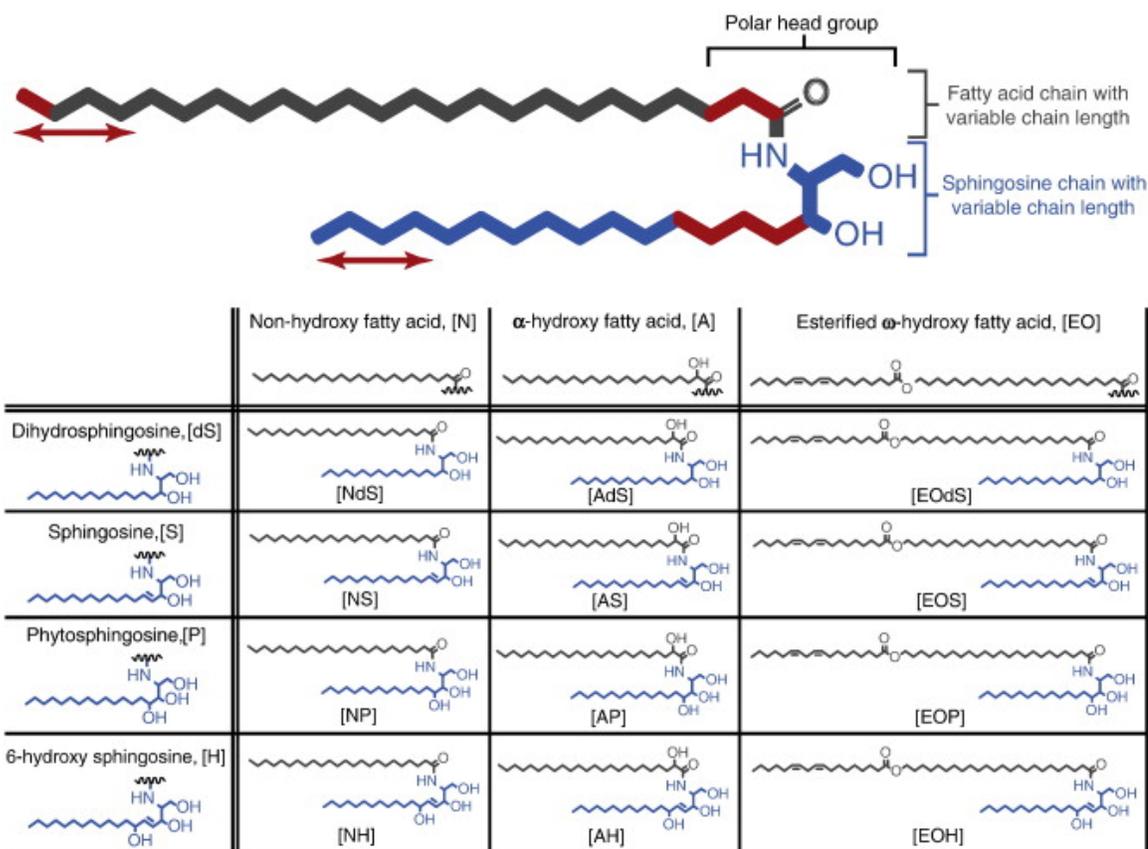


Figure 2.3: The molecular structure of 12 of the ceramide classes present in human stratum corneum, including an explanation of the nomenclature. The newly discovered 1-O-acylceramides<sup>9</sup> are not shown in this figure. Reproduced from Reference 17 with permission from Elsevier.

the SC, the long periodicity phase (LPP) and the short periodicity phase (SPP), with lamellar repeat distances of 13 nm and 6 nm, respectively.<sup>18,19</sup> The lateral organization of lipids consists of dense, crystalline and/or gel-like phases, which differs from the liquid-crystalline behavior seen in other biological membranes. The phases are described as either hexagonal—also called gel—with characteristic lattice spacings of 4.1 Å, or orthorhombic—also called crystalline—with characteristic lattice spacings of 3.7 Å and 4.2 Å.<sup>20,21</sup>

Based on the details of the composition and organization of the SC, several models describing the molecular organization of the SC have been proposed. Swartzendruber *et al.* first proposed the *stacked monolayer model* to describe the lamellar organization

of the lipids in the SC.<sup>22</sup> In this model, the CERs all adopt an extended conformation, where the two tails point in opposite directions. This arrangement results in a trilayer molecular arrangement, which is consistent with the SPP and LPP. Later, Forslind proposed the domain mosaic model, where the SC lipid matrix consists of gel-phase regions surrounded by liquid-crystalline boundaries.<sup>23</sup> This arrangement was used to explain the low permeability yet high malleability of skin.<sup>23</sup> Several versions of a *sandwich model* have been independently proposed by Wertz and colleagues and Bouwstra and colleagues. These models describe a trilayer arrangement, in which a liquid-like lamellar unit that contains the CER EOS linoleate chains is sandwiched between two crystalline lamellar units containing only saturated chains and CHOL, as shown in Figure 2.4. The sandwich model of Bouwstra *et al.* exhibits a 3 nm fluid region surrounded by two 5 nm crystalline regions,<sup>24</sup> whereas the sandwich model of Wertz *et al.* exhibits a 4.3 nm fluid region sandwiched by two 4.3 nm crystalline lamellar units.<sup>25</sup> Norlén proposed the *single gel phase model*, which describes a single, coherent gel phase of the lipids with no phase separation.<sup>26</sup> The model of McIntosh shows a stacked bilayer arrangement of the lipids, with an unequal distribution of CHOL and the linoleate chains of the acylceramides.<sup>27</sup> A similar model proposed by Schröter *et al.* exhibits a homogeneous distribution of CHOL and the linoleate chains of the acylceramides.<sup>28</sup> Neither the single gel phase model, the sandwich models, nor the models of McIntosh or Schröter make any predictions on conformations of the CERs.<sup>24,25,27,28</sup> Finally, in 2012, Iwai *et al.* proposed a model, based on cryoelectron microscopy data, in which all CERs adopt extended conformations and CHOL associates with the sphingoid chains and FFA associates with the fatty acid chains.<sup>29</sup> However, this model does not account for the acylceramides, and only relies on matching a 1-dimensional electron density profile. While many models of the molecular organization of the SC lipids have been proposed, experimental validation of any of these models remains extremely difficult.

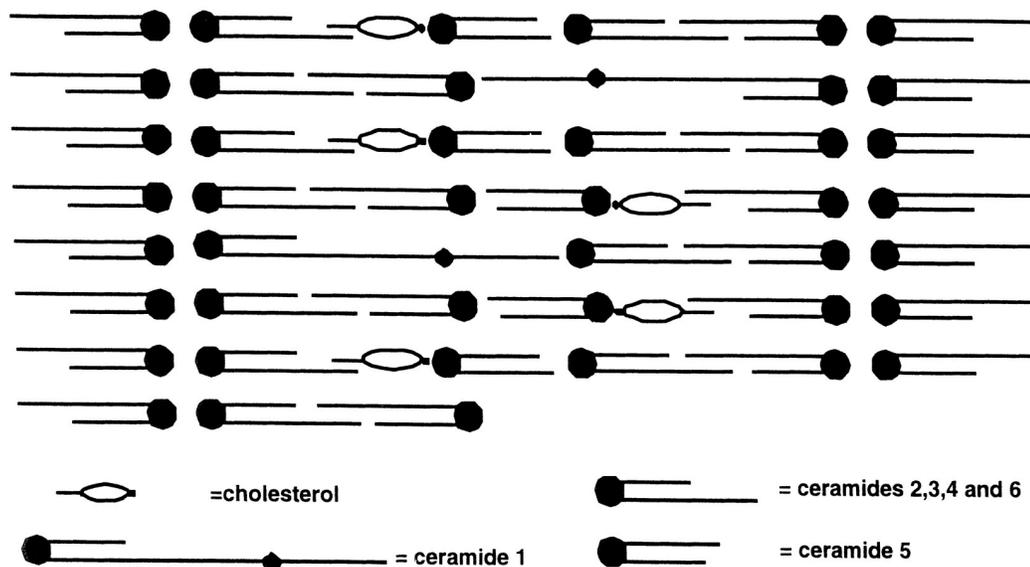


Figure 2.4: A schematic of the proposed sandwich of Bouwstra *et al.*, showing the fluid-like region in the middle of the unit cell surrounded by two, more crystalline regions at the boundaries of the unit cell. Ceramides correspond to Figure 2.3 as follows: ceramide 1 is CER EOS; ceramides 2, 3, 4, and 6 correspond to CER NS, CER NP, CER AS, and CER AP, respectively, each with a fatty acid tail length of 24 carbons; and ceramide 5 corresponds to CER AS C16.

Owing to the varied nature of the CERs and the polydispersity in tail lengths, the SC lipid mixture is very complex, with over 300 individual components identified,<sup>30</sup> making it difficult to isolate the effects of a given lipid on the structural and barrier properties of SC; it therefore becomes difficult to draw relationships between the SC lipid composition, structure, and function. To address this challenge, researchers have begun designing simple mixtures of synthetic SC lipids that exhibit the same phase behavior as the native SC lipid matrix. These mixtures make it so that the lipid composition can be precisely controlled, enabling composition-driven structural changes can be measured. This approach has been successful, as several lipid mixtures have been identified that show the desired properties, and a lot of information has been gleaned from studies using these models. For example, Bouwstra *et al.* showed that formation of the LPP requires an appropriate lipid composition, including the presence of CER EOS and CHOL.<sup>31</sup> Several studies have shown that CHOL tends to be situated

away from the unit cell boundaries.<sup>32,33</sup> Janůšová *et al.* showed that an SPP can be formed with just 3 individual lipids: CER NS C24, CHOL, and FFA C24:0.<sup>34</sup> Recently, Uchiyama *et al.* showed that in model SC membranes, the FFA tail length distribution plays a much larger role in the lateral phase behavior and permeability than the CER composition.<sup>35</sup> While model SC membranes have provided a wealth of information on the relationships between lipid composition, structure, and function in the SC, experiments are still expensive and tedious, often requiring selective deuteration of specific components and creative interpretation of scattering and diffraction data. Interpretation of the data ultimately relies on some assumptions about the molecular organization in the systems. Molecular simulation offers molecular-level resolution of these systems, and hence can be a useful complement to experimental methods.

## 2.1 Bibliography

- [1] MADISON, K. C. Barrier function of the skin: "la raison d'être" of the epidermis. *Journal of Investigative Dermatology* **121**, 2003, pp. 231–241.
- [2] BARBERO, A. M. AND FRASCH, H. F. Transcellular route of diffusion through stratum corneum: results from finite element models. *Journal of Pharmaceutical Sciences* **95**, 2006, pp. 2186–2194.
- [3] KASTING, G. B. AND NITSCHKE, J. M. Mathematical models of skin permeability: Microscopic transport models and their predictions. In *Computational biophysics of the skin*. Pan Stanford Publishing, Singapore (2014), pp. 187–216.
- [4] WANG, T.-F., KASTING, G. B., AND NITSCHKE, J. M. A multiphase microscopic diffusion model for stratum corneum permeability. I. formulation, solution, and illustrative results for representative compounds. *Journal of Pharmaceutical Sciences* **95**, 2006, pp. 620–648.
- [5] WANG, T.-F., KASTING, G. B., AND NITSCHKE, J. M. A multiphase microscopic diffusion model for stratum corneum permeability. II. estimation of physicochemical parameters, and application to a large permeability database. *Journal of Pharmaceutical Sciences* **96**, 2007, pp. 3024–3051.
- [6] KASTING, G. B., MILLER, M. A., AND TALREJA, P. S. Evaluation of stratum corneum heterogeneity. In R. L. BRONAUGH AND H. I. MAIBACH, eds., *Percutaneous Absorption: Drugs-Cosmetics-Mechanisms-Methodology*. Taylor & Francis Group (2005), pp. 193–212.
- [7] SMALL, R., HOANG, D., AND LINDER, J. *A Practical Guide to Chemical Peels, Microdermabrasion & Topical Products*. Lippincott Williams & Wilkins (2012).
- [8] ASCHRAFI, A., FRANZEN, R., SHABAHANG, S., FABBRO, D., PFEILSCHIFTER, J., AND HUWILER, A. Ceramide induces translocation of protein kinase c- $\alpha$  to the Golgi compartment of human embryonic kidney cells by interacting with the C2 domain. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1634**, 2003, pp. 30–39.
- [9] RABIONET, M., GORGAS, K., AND SANDHOFF, R. Ceramide synthesis in the epidermis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 2014, pp. 422–434.
- [10] MOTTA, S., MONTI, M., SESANA, S., CAPUTO, R., CARELLI, S., AND GHIDONI, R. Ceramide composition of the psoriatic scale. *Biochimica et Biophysica Acta - Molecular Basis of Disease* **1182**, 1993, pp. 147–151.
- [11] WERTZ, P. W. Lipids and barrier function of the skin. *Acta Dermato-Venereologica* **Supp 208**, 2000, pp. 7–11.

- [12] FEINGOLD, K. R. The regulation and role of epidermal lipid synthesis. *Advances in Lipid Research* **24**, 1991, pp. 57–82.
- [13] MCINTOSH, T. J. The effect of cholesterol on the structure of phosphatidylcholine bilayers. *Biochimica et Biophysica Acta - Biomembranes* **513**, 1978, pp. 43–58.
- [14] WERTZ, P. W. AND VAN DEN BERGH, B. The physical, chemical and functional properties of lipids in the skin and other biological barriers. *Chemistry and Physics of Lipids* **91**, 1998, pp. 85–96.
- [15] KANICKY, J. R. AND SHAH, D. O. Effect of degree, type, and position of unsaturation on the pKa of long-chain fatty acids. *Journal of Colloid and Interface Science* **256**, 2002, pp. 201–207.
- [16] PARRA, J. L. AND PAYE, M. EEMCO guidance for the in vivo assessment of skin surface pH. *Skin Pharmacology and Physiology* **16**, 2003, pp. 188–202.
- [17] VAN SMEDEN, J., JANSSENS, M., GOORIS, G. S., AND BOUWSTRA, J. A. The important role of stratum corneum lipids for the cutaneous barrier function. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 2014, pp. 295–313.
- [18] WHITE, S. H., MIREJOVSKY, D., AND KING, G. I. Structure of lamellar lipid domains and corneocyte envelopes of murine stratum corneum. an X-ray diffraction study. *Biochemistry* **27**, 1988, pp. 3725–3732.
- [19] BOUWSTRA, J. A., GOORIS, G. S., VAN DER SPEK, J. A., AND BRAS, W. Structural investigations of human stratum corneum by small-angle X-ray scattering. *Journal of Investigative Dermatology* **97**, 1991, pp. 1005–1012.
- [20] GARSON, J.-C., DOUCET, J., LÉVÊQUE, J.-L., AND TSOUCARIS, G. Oriented structure in human stratum corneum revealed by X-ray diffraction. *Journal of Investigative Dermatology* **96**, 1991, pp. 43–49.
- [21] BOUWSTRA, J. A., GOORIS, G. S., SALOMONS-DE VRIES, M. A., VAN DER SPEK, J. A., AND BRAS, W. Structure of human stratum corneum as a function of temperature and hydration: a wide-angle X-ray diffraction study. *International Journal of Pharmaceutics* **84**, 1992, pp. 205–216.
- [22] SWARTZENDRUBER, D. C., WERTZ, P. W., KITKO, D. J., MADISON, K. C., AND DOWNING, D. T. Molecular models of the intercellular lipid lamellae in mammalian stratum corneum. *Journal of Investigative Dermatology* **92**, 1989, pp. 251–257.
- [23] FORSLIND, B. A domain mosaic model of the skin barrier. *Acta Dermato-Venereologica* **74**, 1994, pp. 1–6.

- [24] BOUWSTRA, J. A., DUBBELAAR, F. E. R., GOORIS, G. S., AND PONEC, M. The lipid organisation in the skin barrier. *Acta Dermato-Venereologica; Supplement* **208**, 2000, pp. 23–30.
- [25] HILL, J. R. AND WERTZ, P. W. Molecular models of the intercellular lipid lamellae from epidermal stratum corneum. *Biochimica et Biophysica Acta - Biomembranes* **1616**, 2003, pp. 121–126.
- [26] NORLÉN, L. Skin barrier structure and function: the single gel phase model. *Journal of Investigative Dermatology* **117**, 2001, pp. 830–836.
- [27] MCINTOSH, T. J. Organization of skin stratum corneum extracellular lamellae: diffraction evidence for asymmetric distribution of cholesterol. *Biophysical Journal* **85**, 2003, pp. 1675–1681.
- [28] SCHRÖTER, A., KESSNER, D., KISELEV, M. A., HAUSS, T., DANTE, S., AND NEUBERT, R. H. H. Basic nanostructure of stratum corneum lipid matrices based on ceramides [EOS] and [AP]: a neutron diffraction study. *Biophysical Journal* **97**, 2009, pp. 1104–1114.
- [29] IWAI, I., HAN, H., DEN HOLLANDER, L., SVENSSON, S., OFVERSTEDT, L.-G., ANWAR, J., BREWER, J., BLOKSGAARD, M., LALOEUF, A., NOSEK, D., MASICH, S., BAGATOLLI, L. A., SKOGLUND, U., AND NORLÉN, L. The human skin barrier is organized as stacked bilayers of fully extended ceramides with cholesterol molecules associated with the ceramide sphingoid moiety. *Journal of Investigative Dermatology* **132**, 2012, pp. 2215–25.
- [30] MASUKAWA, Y., NARITA, H., SATO, H., NAOE, A., KONDO, N., SUGAI, Y., OBA, T., HOMMA, R., ISHIKAWA, J., TAKAGI, Y., AND KITAHARA, T. Comprehensive quantification of ceramide species in human stratum corneum. *Journal of Lipid Research* **50**, 2009, pp. 1708–1719.
- [31] BOUWSTRA, J. A., GOORIS, G. S., CHENG, K., WEERHEIM, A., BRAS, W., AND PONEC, M. Phase behavior of isolated skin lipids. *Journal of Lipid Research* **37**, 1996, pp. 999–1011.
- [32] MOJUMDAR, E. H., GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., AND BOUWSTRA, J. A. Localization of cholesterol and fatty acid in a model lipid membrane: a neutron diffraction approach. *Biophysical Journal* **105**, 2013, pp. 911–918.
- [33] MOJUMDAR, E. H., GOORIS, G. S., GROEN, D., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., , AND BOUWSTRA, J. A. Stratum corneum lipid matrix: location of acyl ceramide and cholesterol in the unit cell of the long periodicity phase. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 1926–1934.
- [34] JANUŠOVÁ, B., ZBYTOVSKÁ, J., LORENC, P., VAVRYSOVÁ, H., PALÁT, K., HRABÁLEK, A., AND VÁVROVÁ, K. Effect of ceramide acyl chain length on

skin permeability and thermotropic phase behavior of model stratum corneum lipid membranes. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1811**, 2011, pp. 129–137.

- [35] UCHIYAMA, M., OGURI, M., MOJUMDAR, E. H., GOORIS, G. S., AND BOUWSTRA, J. A. Free fatty acids chain length distribution affects the permeability of skin lipid model membranes. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 2050–2059.

## CHAPTER 3

### SKIN SIMULATION BACKGROUND

Despite recent advances in experimental techniques for studying SC lipid systems, a clear understanding of the molecular-level structure of the SC lipid matrix is lacking. Several models have been proposed,<sup>1-8</sup> but difficulties in performing and interpreting experiments makes identifying which is the “correct” model a challenge. In this regard, molecular simulation can be a valuable tool for studying SC lipid systems. Simulation offers advantages over experiment in several aspects. First, and perhaps most obvious, is the molecular-level resolution that simulation offers. This resolution makes it possible, for example, to directly measure hydrogen bonding based on atomic positions instead of relying on the interpretation of FTIR data. Simulation also allows the energetics of particular processes to be probed. For example, one can use simulation to study the permeation of small molecules through SC model membranes, allowing mechanisms to be identified based on calculated thermodynamic and kinetic properties. Finally, simulation allows infinitely precise control over the composition of a system, making it possible to clearly identify composition-dependent trends.

As a result of these benefits, molecular simulation is becoming a popular tool for studying SC lipid systems. This trend is illustrated in Figure 3.1, which shows the number of original research articles published focusing on SC lipid systems per year since the first one was published in 2001.<sup>9</sup> In this Chapter, the body of literature focusing on SC simulations is reviewed. First, the molecular simulation methodology is briefly described, with comments on specific aspects related to SC lipid simulations. Next, I describe measurements that are possible with simulation, with results presented from representative studies. I then review coarse-grained modeling and its application to SC lipid systems.

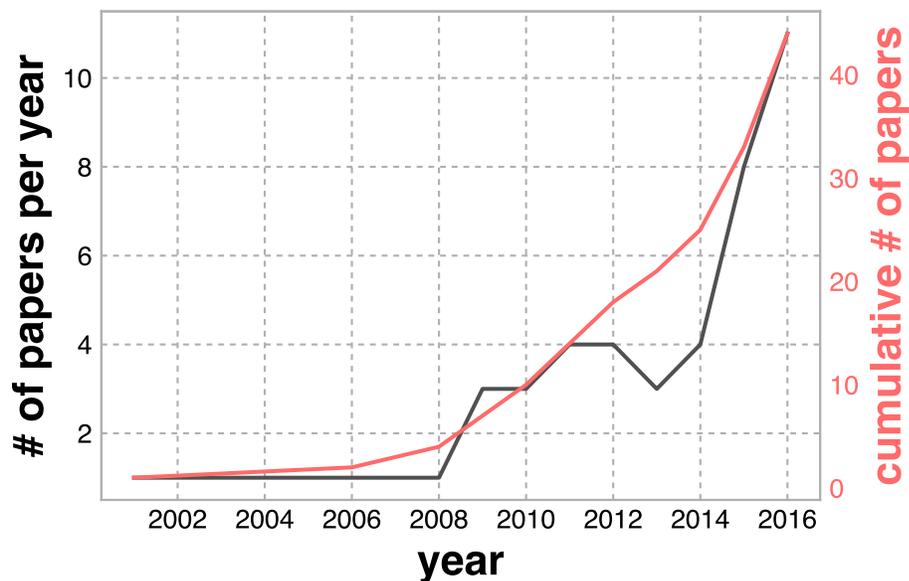


Figure 3.1: Number of SC simulation papers published per year as a function of time. These include original research articles that study SC lipids in the context of skin. Note the different scale on the right axis for the cumulative number of papers.

### 3.1 Simulation Methodology

The main goal of molecular simulation is to examine how molecular interactions give rise to the properties of physical systems. This is achieved by generating a set—or ensemble—of molecular configurations for a given system, and calculating properties from these configurations. These properties can then be connected to macroscopic observables through statistical mechanics. The ensemble can be constructed such that it samples from a given thermodynamic ensemble. For example, since molecular simulations of SC lipid systems aim to compare properties with experimental systems at laboratory conditions, simulations are generally designed to sample from the isothermal–isobaric (or NPT) ensemble.

There are two general types of molecular simulation algorithms to generate configurations. Monte Carlo (MC) methods generate configurations by stochastic displacements of atoms, which are either accepted or rejected based on criteria designed to ensure that the desired ensemble is sampled. As configurations are generated

stochastically, MC simulations have no straightforward connection to time, and hence dynamic properties cannot be calculated. In contrast, molecular dynamics (MD) employs classical mechanics to numerically calculate the trajectory of a system through time based on the forces between the atoms. As MD naturally contains an associated time scale, dynamic properties can be calculated. Additionally, parallelization of MD calculations is more straightforward than MC calculations, allowing very large systems to be simulated more efficiently with MD. Therefore, MD is the predominate method for studying SC lipid systems computationally.

In MD, a system is initialized with a given configuration, and the forces on the atoms are calculated based on the interactions between them; these interactions are encoded in a “force field.” Based on these forces, the positions of the atoms are updated by numerically integrating the equations of motion. Timesteps are generally on the order of 1 to 2 ns, and are limited by the fastest motions in the system (*e.g.*, bond vibrations involving a hydrogen atom). From the initial configuration, the first part of integration is usually regarded as “equilibration,” where the system is allowed to relax from its initial configuration, such that the calculated properties do not depend too heavily on the initial configuration. After equilibration, the “production” portion of the simulation is performed, generally for  $1 \times 10^6$  to  $100 \times 10^6$  timesteps (*i.e.*, 1s–100s of ns of simulation time), and the configurations and thermodynamic properties are saved at specified time intervals. These configurations, when written to disk, are referred to as the “simulation trajectory.” The simulation trajectory is then post-processed to calculate the properties of interest.

### 3.2 Force Fields

As described above, MD simulations rely on force fields to calculate the forces on each atom at each timestep. Thus, the force field (combined with the initial configuration) determines the trajectory of the simulation, and hence also determines

the calculated properties.

Force fields contain terms to account for physical interactions stemming from the chemical nature of the molecules; these terms can broadly be categorized into bonded and nonbonded interactions. The bonded forces represent covalent interactions between atoms. These interactions include bond-stretching and angle-bending to account for the vibrational modes of molecules. Dihedral terms are included to account for the torsional rotations around bonds. Bonds and angles are generally modeled as harmonic springs, and hence these forces vary linearly as a function of displacement from equilibrium. Dihedrals are generally modeled with periodic functions to account for local minima on the torsional potential energy surface. The nonbonded forces are assumed to act pairwise between atoms, and represent noncovalent interactions. van der Waals interactions are usually modeled with a Lennard-Jones potential, which contains a 12<sup>th</sup> order short-range repulsive term to account for the unfavorable overlap of electron clouds and a 6<sup>th</sup> order long-range attractive term to account for the long-range attractions between atoms. Coulomb’s law is used to calculate the electrostatic forces between (partially) charged atoms, *e.g.*, resulting from electronegativity differences between bonded atoms. The total potential energy of the system is the sum of all of these interactions, as shown in Figure 3.2. Note that the terms described above and shown in Figure 3.2 give expressions for the potential energy, whereas MD relies on forces to update the particle positions; the force is simply the gradient of the potential energy function,  $F = -\nabla U$ .

Several force fields have been developed specifically for lipid molecules, and SC lipid simulations are generally based on two common force fields. Notman *et al.*<sup>11</sup> first used the GROMOS87<sup>12</sup>-based Berger force field,<sup>13</sup> which has since been widely used for simulating CERs.<sup>14–27</sup> Guo *et al.*<sup>18</sup> later derived parameters for the headgroups of CER NS and CER NP to fit within the CHARMM36 force field.<sup>28</sup> These parameters were shown to give good agreement with experiment on the thermotropic phase behavior of

$$\begin{aligned}
U = & \sum_{i < j} \sum 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
& + \sum_{i < j} \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \\
& + \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 \\
& + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 \\
& + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi - \delta)]
\end{aligned}$$

Figure 3.2: The total potential energy is the sum over the various bonded and nonbonded interactions. The first two terms represent the nonbonded interactions between atoms  $i$  and  $j$  as a function of their separation  $r_{ij}$ . The Lennard-Jones parameters  $\sigma_{ij}$  and  $\varepsilon_{ij}$  are determined by the force field.  $q_i$  represents the partial charge on atom  $i$ . The last three terms are the bonded interactions. The force constants  $k_b$ ,  $k_a$ , and  $k_\phi$  are specific to a given bond (or angle or dihedral) type for a given force field. Reprinted from Ref. 10 with permission from Elsevier.

pure CER bilayers.<sup>18</sup> Papadimitriou *et al.* compared the structural properties of pure CER NS C24 bilayers simulated with various force fields at skin conditions, finding acceptable agreement with experimental data for all force fields.<sup>21</sup> It should be noted, however, that the CHARMM-compatible CER force field developed by Guo *et al.* showed better agreement in the thermotropic behavior of pure CER bilayers,<sup>18</sup> and hence one take this into consideration if the thermotropic behavior is to be studied.

Up to this point, the discussion has assumed that the models represent each atom explicitly; these are termed “atomistic” force fields. However, the time and length scales that are accessible with atomistic models is limited because of the large number of interactions that are present. To address this issue, simplified, or “coarse-grained” (CG) force fields are often used in simulation to lower the computational expense

and increase the amount of sampling that can be achieved for a given amount of computational resources. CG models thus allow phenomena that occur on large time and length scales, *e.g.*, self-assembly, to be studied computationally. The reduction in computational cost is achieved by lumping groups of atoms into single interaction sites, or “CG beads.” The level of coarse-graining, *i.e.*, how many atoms are grouped into single beads, determines the level of specificity of the CG model and the potential computational speedup. CG models are often parameterized based on atomistic simulation data; for a review on methods for parameterizing CG models, see the excellent review by Brini *et al.*<sup>29</sup> CG modeling in the context of SC lipid simulation is discussed in a separate section below.

### 3.3 SC Lipid Simulation Specifics

While simulations of fluids of small molecules and their mixtures are routine, simulations of SC lipids are more challenging for a number of reasons. First, the SC lipid matrix is a complex mixture containing hundreds of individual lipid species, so a truly representative system would necessarily require on the order of  $1 \times 10^5$  to  $1 \times 10^6$  atoms. While modern computer power allows simulations of millions of atoms, such simulations are computationally very expensive, so studying a series of systems (*e.g.*, to study the effect of a single species on the structure) is still impractical. To avoid this issue, simulations of SC lipid mixtures generally contain a limited number of lipid species, ranging from a single CER to mixtures of several CERs plus CHOL and FFA, greatly reducing the required number of molecules (and hence computational power) needed for a representative system. The dense, gel-like packing of the lipids also presents a problem; the molecular diffusion in these systems is very low,<sup>30</sup> so systems may be stuck in their initial configurations, increasing the likelihood that systems are trapped in metastable configurations. Simulations are often performed at elevated temperatures (relative to skin conditions)<sup>15</sup> to increase the molecular mobility, but

recent work has shown that this approach may be flawed.<sup>31</sup> Therefore, care must be taken to ensure that system configurations represent low-energy structures. Another challenge in simulating SC lipids is related to the lamellar organization of lipids in the SC; the SC lipid matrix is organized as stacked lamellar repeat units. This reality presents two problems; first, the lack of a detailed picture of the molecular-level organization means it is unknown how to initialize lamellar systems. This problem is compounded by the fact that these systems exist in dense, gel-like phases, so observing a lamellar transition is unlikely in any realistic simulation timescale. Next, simulating multiple lipid lamellae would again require very large systems, making it difficult to study a series of related systems. To address these issues, a vast majority of studies examine the SC lipids in preassembled bilayer configurations, and only a handful of simulations have been performed on multilayer systems.<sup>19,26</sup> A different approach is to simulate the self-assembly of SC lipids into multilayer structures, which removes the influence of the initial configuration on the system properties. However, the system sizes and timescales required for self-assembly are computationally prohibitive for atomistic models, so CG models are needed for this approach.

### 3.4 Atomistic Molecular Dynamics Simulations of SC Lipid Systems

The majority of SC lipid simulations have been performed with atomistic models. Hölte *et al.* first studied SC lipids computationally in 2001, examining mixtures of CHOL and FFA with experiment and simulations. The authors observed that the addition of CHOL leads to more chain disorder in the FFAs, especially towards the center of the bilayer, suggesting that CHOL may act as a fluidizing agent in the SC.<sup>9</sup> Pandit and Scott first reported on simulation of a CER bilayer 5 years later in 2006, comparing the lipid–water interface between pure CER NS C16 and sphingomyelin bilayers. They found that the CER structure factor shows significantly less structure than sphingomyelin, and that despite the polar headgroups of CER,

the water near it behaves as though it is near a hydrophobic surface, which may contribute to the good barrier properties of the SC lipids.<sup>32</sup> The authors also observed an increase in lipid–lipid hydrogen bonding in CER compared to sphingomyelin, which was rationalized by the fact that the CER carbonyl oxygen has more conformational freedom than in sphingomyelin. The first simulations of a more realistic CER with an asymmetric tail length were reported by Notman *et al.* in 2007, who simulated a bilayer composed of CER NS C24, and observed a slightly tilted gel-phase bilayer at 283 K and 323 K, and a liquid-crystalline bilayer at 363 K, in good agreement with experimental observations. The same work revealed that the permeation enhancer dimethyl sulfoxide induces a gel-to-fluid transition at skin conditions by interrupting the hydrogen bonds form between the CER headgroups, which suggests a mechanism for its barrier-lowering effects on the skin.<sup>11</sup> Das *et al.* were the first to simulate a realistic SC lipid mixture in bilayer configurations.<sup>15</sup> Systems were composed of CER NS C24, CHOL, and FFA C24:0 at 16 different compositions, ranging from the pure components, to binary mixtures, and ternary mixtures. They found that the addition of CHOL, which is shorter and bulkier than the CER and FFA, tends to “squash” the bilayers, leading to increased interdigitation between the two leaflets and a decreased bilayer thickness. Additionally, it was observed that the 2:2:1 CER:CHOL:FFA system, which most closely resembles the composition of the SC lipid matrix, has the lowest stress and compressibility, and post that these properties may be responsible for the low permeability yet high pliability of the SC. Imai *et al.* reported on the residence time of water on CER and sphingomyelin bilayers, finding that water tends to have a shorter residence time on CER bilayers compared to sphingomyelin, and that incorporation of an unsaturated tail seems to further shorten the residence time.<sup>33</sup> Engelbrecht *et al.* reported on simulations of 4 bilayers composed of CER EOS, CER AP, CHOL, and FFA C22:0 stacked on top of one another with very little interbilayer water to test the stability of a model they propose to describe the molecular organization of

the SC lipid matrix.<sup>34</sup> In the simulation, all CERs were initially placed in hairpin conformations, where the long tail of CER EOS extended into adjacent bilayers, and only 2.5% of the CER EOS tails underwent a flipping transition where the molecule only spanned a single bilayer. In 2011, Hoopes *et al.* reported on simulations of SC lipid mixtures and the effect of adding an unsaturated fatty acid (oleic acid, OA), finding that the addition of OA did not affect the hydrogen bonding in the lipid headgroups.<sup>30</sup> Interestingly, the authors observed that the faster diffusing groups had a higher local density of CHOL, illustrating that CHOL increases the mobility of the lipids. Furthermore, the authors observed that OA increases the CHOL diffusion, which in turn increases the diffusion of the other species, likely a result of the increased free area with the addition of OA.<sup>30</sup> In 2013, Das *et al.* reported on simulations of the self-assembly of very large mixtures of CER NS, CER NP, CER EOS, CHOL, and FFA with realistic tail length polydispersities, finding that inverse lamellar structures tend to form in the bulk, whereas lamellar structures form in the presence of a corneocyte-mimicking template.<sup>17</sup> However, the self-assembly protocol used in this paper may favor the formation of inverse micelles in the bulk, and thus does not represent similar experimental systems, which have been shown to form lamellar structures in the bulk.<sup>35,36</sup> Guo *et al.* in 2013 derived parameters for the CER headgroups within the CHARMM force field, which had been lacking. The newly derived parameters, along with the Berger force field, were used to compare the structural properties of pure CER NS C16 bilayers, as well as the thermotropic phase behavior of the bilayers. Both force fields accurately predicted the structural properties of the bilayers at skin conditions, but the CHARMM-CER force field more accurately reproduced the thermotropic phase behavior.<sup>18</sup> Additionally, bilayers composed of CER NP, which has an extra hydroxyl group compared to CER NS, were found to have a higher order-disorder transition temperature, which was attributed to the increased lipid-lipid hydrogen bonding compared to CER NS. In 2014, Das *et al.* examined the flip-flop motion of

CHOL in stacked bilayers of CER NS, CHOL, and FFA C24:0, providing the first detailed analysis of a stacked multilayer system composed of SC lipids. The authors observed that 42% of the hydrogen bonds formed between headgroups in the middle leaflets were with lipids in the adjacent leaflet, showing that the leaflets are coupled to one another through hydrogen bonds.<sup>19</sup> The calculated CHOL flip-flop rate was found to be faster than in phospholipid systems, which was attributed to a low density region at the center of the bilayers that results from tail length asymmetry between the lipids in the system, allowing a faster reorientation of the CHOL molecules. Paloncyová *et al.* compared the permeability of a small molecule through gel and fluid phase bilayers, where pure CER NS was used to model the gel phase membrane, finding a larger energetic penalty for disrupting the packing in the CER bilayer compared to the fluid phospholipid bilayer.<sup>20</sup> Papadimitriou *et al.* compared force fields for simulating pure CER NS C24:0 bilayers in 2015, and found that of the force fields tested, all yield structural properties that are within acceptable agreement with experiment.<sup>21</sup> Paloncyová *et al.* studied the effect of the CER NS fatty acid tail length on the permeability of bilayers composed of CER NS, CHOL, and FFA, finding that the permeability of water peaks when the CER fatty acid tails are in the range of 4 to 8 carbons long.<sup>22</sup> This trend was rationalized by the fact that these tail lengths induce a less dense bilayer, allowing the formation of small water pores. In 2016, Gupta and Rai found that bilayers composed of mixtures of CER NS, CHOL, and FFA are more thermally stable than pure CER systems, suggesting that the lipid composition may have evolved for stability purposes.<sup>23</sup> Gupta *et al.* further studied the permeation of small molecules through SC lipid bilayers, finding that diffusion plays a larger role for hydrophobic permeants compared to hydrophilic ones, where the free energy barrier mainly determines the permeability.<sup>25</sup> Das *et al.* found that the shape of CER molecule, which is cylindrical because of the relatively small headgroup, prevents a percolating hydrogen bond network from forming, and instead, isolated hydrogen

bonding clusters tend to form between the CER headgroups.<sup>26</sup> Akinshina *et al.* studied the insertion of monoglycerides (oils) and free fatty acids into CER bilayers, and found that while low concentrations of oils can lead to a more dense lipid packing, higher concentrations of oils with a kinked chain can eventually lead to bilayer instability.<sup>27</sup> Gupta *et al.* examined the effect of the CER fatty acid tail length on the permeability of water through CER bilayers, finding that the longer CERs result in a larger free energy barrier for penetration, and hence a larger permeability.<sup>24</sup> In 2017, Moore *et al.* reported on the simulations of a series of CER NS, CER NS/CHOL, and CER NS/CHOL/FFA C24:0 bilayers with varying fractions of shorter CERs to examine the effect of the CER fatty acid tail length on the structural properties of the bilayers. In that work, a simulated-tempering inspired equilibration methodology was introduced and validated to show the importance of thorough equilibration for SC lipid systems. It was further shown that the main effect of the CER tail length is on the middle region of the bilayers, where shorter tails lead to a smaller interdigitation region, and hence a thinner bilayer.<sup>31</sup>

### 3.5 Coarse-Grained Molecular Dynamics Simulations of SC Lipid Systems

While atomistic models have been heavily used since the turn of the millennium to study SC lipid systems, the use of CG models has been relatively more sparse. A major cause of this trend is that unlike atomistic models, generic CG models have not been as widely developed, so researchers must first develop the CG models to be used to simulate a specific system. Therefore, a lot of the work in CG modeling of SC lipid systems has involved the development of the required force fields.

Hadley and McCabe were the first to apply CG modeling to specifically study SC lipid systems. They used iterative Boltzmann inversion (IBI)<sup>37</sup> to develop structurally accurate force fields from atomistic simulations. They also used a clever  $k$ -means-based scheme to dynamically map four water molecules to a single site, enabling the use of

IBI to optimize the pair potentials in a CG water model.<sup>38</sup> Using the CG models they developed,<sup>38-40</sup> the authors studied the self-assembly of mixtures of CHOL and FFA, and elucidated the mechanisms for the hydrophobicity length-matching requirement for their assembly into stable bilayer structures.<sup>41</sup>

The first CG model of CER was developed in the framework of the popular MARTINI force field,<sup>42</sup> and has since been used in several studies. It was first used to clarify the structure of the melted chain phase above the phase transition temperature of the bilayer phase, and described an irregular bicontinuous phase mixed with water droplets.<sup>43</sup> Since then, Gupta and coworkers have used it in several studies to examine the permeability of realistic mixtures of CER, CHOL, and FFA.<sup>44-47</sup>

Concurrently, McCabe and colleagues have been developing CG models of the SC lipids using the multistate IBI method<sup>48</sup> to derive structurally accurate, state-transferable force fields. Emphasis was placed on the multistate nature of the force field development, since self-assembly is inherently a multistate process. The authors have published force fields for water,<sup>49</sup> CER NS,<sup>50</sup> FFA,<sup>50</sup> and mixtures of CER NS and FFA.<sup>51</sup> Using these force fields, the authors observed that CER NS self-assembles into both bilayer and multilayer structures. The CER conformations in the multilayer structures revealed that a small but significant fraction of the CERs adopt extended conformations,<sup>50</sup> and that this fraction of extended CERs increases in mixtures with FFA.<sup>51</sup> Current work, as described in Chapter 9 of this Dissertation, is focused on developing the necessary interactions involving CHOL to examine realistic mixtures composed of CER, CHOL, and FFA.

### 3.6 Bibliography

- [1] SWARTZENDRUBER, D. C., WERTZ, P. W., KITKO, D. J., MADISON, K. C., AND DOWNING, D. T. Molecular models of the intercellular lipid lamellae in mammalian stratum corneum. *Journal of Investigative Dermatology* **92**, 1989, pp. 251–257.
- [2] FORSLIND, B. A domain mosaic model of the skin barrier. *Acta Dermato-Venereologica* **74**, 1994, pp. 1–6.
- [3] BOUWSTRA, J. A., DUBBELAAR, F. E. R., GOORIS, G. S., AND PONEC, M. The lipid organisation in the skin barrier. *Acta Dermato-Venereologica; Supplement* **208**, 2000, pp. 23–30.
- [4] HILL, J. R. AND WERTZ, P. W. Molecular models of the intercellular lipid lamellae from epidermal stratum corneum. *Biochimica et Biophysica Acta - Biomembranes* **1616**, 2003, pp. 121–126.
- [5] NORLÉN, L. Skin barrier structure and function: the single gel phase model. *Journal of Investigative Dermatology* **117**, 2001, pp. 830–836.
- [6] MCINTOSH, T. J. Organization of skin stratum corneum extracellular lamellae: diffraction evidence for asymmetric distribution of cholesterol. *Biophysical Journal* **85**, 2003, pp. 1675–1681.
- [7] SCHRÖTER, A., KESSNER, D., KISELEV, M. A., HAUSS, T., DANTE, S., AND NEUBERT, R. H. H. Basic nanostructure of stratum corneum lipid matrices based on ceramides [EOS] and [AP]: a neutron diffraction study. *Biophysical Journal* **97**, 2009, pp. 1104–1114.
- [8] IWAI, I., HAN, H., DEN HOLLANDER, L., SVENSSON, S., OFVERSTEDT, L.-G., ANWAR, J., BREWER, J., BLOKSGAARD, M., LALOEUF, A., NOSEK, D., MASICH, S., BAGATOLLI, L. A., SKOGLUND, U., AND NORLÉN, L. The human skin barrier is organized as stacked bilayers of fully extended ceramides with cholesterol molecules associated with the ceramide sphingoid moiety. *Journal of Investigative Dermatology* **132**, 2012, pp. 2215–25.
- [9] HÖLTJE, M., FÖRSTER, T., BRANDT, B., ENGELS, T., VON RYBINSKI, W., AND HÖLTJE, H.-D. Molecular dynamics simulations of stratum corneum lipid models: fatty acids and cholesterol. *Biochimica et Biophysica Acta - Biomembranes* **1511**, 2001, pp. 156–167.
- [10] NOTMAN, R. AND ANWAR, J. Breaching the skin barrier—insights from molecular simulation of model membranes. *Advanced Drug Delivery Reviews* **65**, 2013, pp. 237–250.
- [11] NOTMAN, R., DEN OTTER, W. K., NORO, M. G., BRIELS, W. J., AND ANWAR, J. The permeability enhancing mechanism of DMSO in ceramide

- bilayers simulated by molecular dynamics. *Biophysical Journal* **93**, 2007, pp. 2056–2068.
- [12] VAN GUNSTEREN, W. F. AND BERENDSEN, H. J. C. Groningen molecular simulation (GROMOS) library manual. *Biomos, Groningen*, 1987, p. 13.
- [13] BERGER, O., EDHOLM, O., AND JÄHNIG, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **72**, 1997, pp. 2002–2013.
- [14] NOTMAN, R., ANWAR, J., BRIELS, W. J., NORO, M. G., AND DEN OTTER, W. K. Simulations of skin barrier function: free energies of hydrophobic and hydrophilic transmembrane pores in ceramide bilayers. *Biophysical Journal* **95**, 2008, pp. 4763–4771.
- [15] DAS, C., NORO, M. G., AND OLMSTED, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **97**, 2009, pp. 1941–1951.
- [16] DAS, C., OLMSTED, P. D., AND NORO, M. G. Water permeation through stratum corneum lipid bilayers from atomistic simulations. *Soft Matter* **5**, 2009, pp. 4549–4555.
- [17] DAS, C., NORO, M. G., AND OLMSTED, P. D. Lamellar and inverse micellar structures of skin lipids: Effect of templating. *Physical Review Letters* **111**, 2013, p. 148101.
- [18] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>c</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [19] DAS, C., NORO, M. G., AND OLMSTED, P. D. Fast cholesterol flip-flop and lack of swelling in skin lipid multilayers. *Soft Matter* **10**, 2014, pp. 7346–7352.
- [20] PALONCÝOVÁ, M., DEVANE, R., MURCH, B. P., BERKA, K., AND OTYEPKA, M. Rationalization of reduced penetration of drugs through ceramide gel phase membrane. *Langmuir* **30**, 2014, pp. 13942–13948.
- [21] PAPADIMITRIOU, N. I., KAINOURGIAKIS, M. E., KAROZIS, S. N., AND CHARALAMBOPOULOU, G. C. Studying the structure of single-component ceramide bilayers with molecular dynamics simulations using different force fields. *Molecular Simulation* **41**, 2015, pp. 1122–1136.
- [22] PALONCÝOVÁ, M., VÁVROVÁ, K., SOVOVÁ, Ž., DEVANE, R., OTYEPKA, M., AND BERKA, K. Structural changes in ceramide bilayers rationalize increased permeation through stratum corneum models with shorter acyl tails. *The Journal of Physical Chemistry B* **119**, 2015, pp. 9811–9819.

- [23] GUPTA, R. AND RAI, B. Molecular dynamics simulation study of skin lipids: effects of the molar ratio of individual components over a wide temperature range. *The Journal of Physical Chemistry B* **119**, 2015, pp. 11643–11655.
- [24] GUPTA, R., DWADASI, B. S., AND RAI, B. Molecular dynamics simulation of skin lipids: effect of ceramide chain lengths on bilayer properties. *The Journal of Physical Chemistry B* **120**, 2016, pp. 12536–12546.
- [25] GUPTA, R., SRIDHAR, D. B., AND RAI, B. Molecular dynamics simulation study of permeation of molecules through skin lipid bilayer. *The Journal of Physical Chemistry B* **120**, 2016, pp. 8987–8996.
- [26] DAS, C. AND OLMSTED, P. D. The physics of stratum corneum lipid membranes. *Philosophical Transactions of the Royal Society A* **374**, 2016, p. 20150126.
- [27] AKINSHINA, A., DAS, C., AND NORO, M. G. Effect of monoglycerides and fatty acids on a ceramide bilayer. *Physical Chemistry Chemical Physics* **18**, 2016, pp. 17446–17460.
- [28] KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O’CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D., AND PASTOR, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 2010, pp. 7830–7843.
- [29] BRINI, E., ALGAER, E. A., GANGULY, P., LI, C., RODRÍGUEZ-ROPERO, F., AND VAN DER VEGT, N. F. A. Systematic coarse-graining methods for soft matter simulations—a review. *Soft Matter* **9**, 2013, pp. 2108–2119.
- [30] HOOPEES, M. I., NORO, M. G., LONGO, M. L., AND FALLER, R. Bilayer structure and lipid dynamics in a model stratum corneum with oleic acid. *The Journal of Physical Chemistry B* **115**, 2011, pp. 3164–3171.
- [31] MOORE, T. C., HARTKAMP, R., IACOVELLA, C. R., BUNGE, A. L., AND M<sup>c</sup>CABE, C. The influence of ceramide tail length on the structure of bilayers composed of stratum corneum lipids. *Biophysical Journal* **in press**.
- [32] PANDIT, S. A. AND SCOTT, H. L. Molecular-dynamics simulation of a ceramide bilayer. *The Journal of Chemical Physics* **124**, 2006, p. 014708.
- [33] IMAI, Y., LIU, X., YAMAGISHI, J., MORI, K., NEYA, S., AND HOSHINO, T. Computational analysis of water residence on ceramide and sphingomyelin bilayer membranes. *Journal of Molecular Graphics and Modelling* **29**, 2010, pp. 461–469.
- [34] ENGELBRECHT, T., HAUSS, T., SÜSS, K., VOGEL, A., ROARK, M., FELLER, S. E., NEUBERT, R. H. H., AND DOBNER, B. Characterisation of a new ceramide EOS species: synthesis and investigation of the thermotropic phase behaviour and influence on the bilayer architecture of stratum corneum lipid model membranes. *Soft Matter* **7**, 2011, pp. 8998–9011.

- [35] WARTEWIG, S. AND NEUBERT, R. H. H. Properties of ceramides and their impact on the stratum corneum structure: a review. Part 1: ceramides. *Skin Pharmacology and Physiology* **20**, 2007, pp. 220–229.
- [36] KESSNER, D., RUETTINGER, A., KISELEV, M. A., WARTEWIG, S., AND NEUBERT, R. H. H. Properties of ceramides and their impact on the stratum corneum structure. Part 2: stratum corneum lipid model systems. *Skin Pharmacology and Physiology* **21**, 2008, pp. 58–74.
- [37] REITH, D., PÜTZ, M., AND MÜLLER-PLATHE, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24**, 2003, pp. 1624–1636.
- [38] HADLEY, K. R. AND M<sup>C</sup>CABE, C. On the investigation of coarse-grained models for water: balancing computational efficiency and the retention of structural properties. *The Journal of Physical Chemistry B* **114**, 2010, pp. 4590–4599.
- [39] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A structurally relevant coarse-grained model for cholesterol. *Biophysical Journal* **99**, 2010, pp. 2896–2905.
- [40] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A coarse-grained model for amorphous and crystalline fatty acids. *The Journal of Chemical Physics* **132**, 2010, p. 134505.
- [41] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A simulation study of the self-assembly of coarse-grained skin lipids. *Soft Matter* **8**, 2012, pp. 4802–4814.
- [42] MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P., AND DE VRIES, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 2007, pp. 7812–7824.
- [43] SOVOVÁ, Ž., BERKA, K., OTYEPKA, M., AND JUREČKA, P. Coarse-grain simulations of skin ceramide NS with newly derived parameters clarify structure of melted phase. *The Journal of Physical Chemistry B* **119**, 2015, pp. 3988–3998.
- [44] GUPTA, R., KASHYAP, N., AND RAI, B. Transdermal cellular membrane penetration of proteins with gold nanoparticles: a molecular dynamics study. *Physical Chemistry Chemical Physics* **19**, 2017, pp. 7537–7545.
- [45] GUPTA, R. AND RAI, B. Molecular dynamics simulation study of translocation of fullerene C60 through skin bilayer: effect of concentration on barrier properties. *Nanoscale* **9**, 2017, pp. 4114–4127.
- [46] GUPTA, R. AND RAI, B. Effect of size and surface charge of gold nanoparticles on their skin permeability: a molecular dynamics study. *Scientific Reports* **7**.
- [47] GUPTA, R. AND RAI, B. Penetration of gold nanoparticles through human skin: unraveling its mechanisms at the molecular scale. *The Journal of Physical Chemistry B* **120**, 2016, pp. 7133–7142.

- [48] MOORE, T. C., IACOVELLA, C. R., AND M<sup>C</sup>CABE, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140**, 2014, p. 224104.
- [49] MOORE, T. C., IACOVELLA, C. R., AND M<sup>C</sup>CABE, C. Development of a coarse-grained water forcefield via multistate iterative Boltzmann inversion. In *Foundations of Molecular Modeling and Simulation*. Springer (2016), pp. 37–52.
- [50] MOORE, T. C., IACOVELLA, C. R., HARTKAMP, R., BUNGE, A. L., AND M<sup>C</sup>CABE, C. A coarse-grained model of stratum corneum lipids: free fatty acids and ceramide NS. *The Journal of Physical Chemistry B* **120**, 2016, pp. 9944–9958.
- [51] MOORE, T. C., IACOVELLA, C. R., LEONHARD, A. C., BUNGE, A. L., AND M<sup>C</sup>CABE, C. Molecular dynamics simulations of stratum corneum lipid mixtures: a multiscale perspective. *Biochemical and Biophysical Research Communications* **in press**, 2017, pp. 1–6.

## CHAPTER 4

### THE EFFECT OF CERAMIDE TAIL LENGTH ON MODEL MEMBRANE PROPERTIES

This chapter is reproduced in part from Moore *et al.* 2017.<sup>1</sup>

#### 4.1 Introduction

The stratum corneum (SC) layer of the skin acts as the main barrier against chemical penetrants and water loss and is composed of a brick-and-mortar-like arrangement of corneocytes surrounded by a dense, lamellar-structured lipid matrix. The matrix is a complex mixture primarily composed of an equimolar ratio of ceramides (CERs), which are composed of a sphingoid base linked to a saturated fatty acid chain (in most cases), cholesterol (CHOL), and free fatty acids (FFAs).<sup>2,3</sup> Since the extracellular matrix is widely thought to be the only continuous path path for permeation through the SC, proper lipid organization and composition of the lamellar membrane is generally considered crucial for a fully functioning skin barrier. This is supported by the fact that many skin diseases that are characterized by a reduced barrier function also exhibit an altered lipid composition and organization compared to healthy skin.<sup>4-7</sup> For example, a decrease has been observed in the average number of CER carbon atoms in the SC of patients who suffer from atopic eczema<sup>4,8</sup> and, compared with controls, less dense lipid packing and changes in X-ray scattering patterns.<sup>4</sup> Of the two CER tails, reduced CER carbon numbers appear to correspond with a shorter fatty acid (FA) tail rather than the sphingoid base tail.<sup>9</sup>

The complex nature of the human SC lipid matrix, in that it features 15 unique subclasses of CER lipids and a distribution of CER and FFA chain lengths, has made it challenging to establish the influence of individual lipid species on barrier properties.<sup>10</sup>

As a result, many experiments have focused on model systems that are composed of simpler mixtures of synthetic lipids, whose composition and chain lengths can be precisely controlled.<sup>11–22</sup> For example, considering studies of mixtures containing non-hydroxy sphingosine CER (CER NS), CHOL, and FFA, similar to those considered in this work, Mojumdar *et al.*<sup>15</sup> showed that CHOL is essential to the characteristic phase behavior, and hence barrier properties, of such mixtures, and Školová *et al.*<sup>20</sup> found that CER NS-based membranes with short acyl chains had increased permeability compared to long acyl chain CERs. However, while global structural properties of SC lipid lamellae can be inferred from experiment (e.g., from spectroscopic, scattering, and diffraction data), including information about the localization of molecules across the lamellae, the underlying molecular-level details and in-plane morphology often remain elusive. Furthermore, structural examination using neutron scattering is quite laborious, as many independent systems, each with different deuterated lipids and/or concentrations of deuterated lipids, must be considered to provide a single picture of the structure, and the availability of deuterated lipids may limit the study of specific systems.<sup>16,23–25</sup>

Molecular simulation, which offers atomic-level resolution of the entire 3-dimensional structure of the lipids at picosecond timescales, has been used to provide a more direct understanding of the molecular interactions in CER-based membranes,<sup>26–39</sup> with several studies focusing on the dependence of membrane structure on lipid composition. For example, Das *et al.* found that CHOL compresses a CER NS bilayer membrane and increases the interdigitation of CER tails in opposing bilayer leaflets.<sup>27</sup> Gupta and Rai simulated a series of lipid bilayers composed of CER NS, CHOL, and FFA with varying compositions and found that CHOL tends to sit away from the lipid–water interface and increases the thermal stability of the bilayers.<sup>32</sup> The effect of the CER FA tail length on bilayer properties has received considerably less attention. Palonciová *et al.* showed that extremely short CER NS FA tails (*e.g.*, 4–6 carbons)

cause a disruption in the headgroup packing of CER NS bilayers, which gives rise to an increased permeability;<sup>38</sup> however, the CER NS FA tail lengths considered were much shorter than those found in the SC. More recently, Gupta *et al.* showed that the permeability of water across pure CER NS bilayers with more biologically relevant tail lengths decreases monotonically with the CER FA tail length,<sup>31</sup> although a structural explanation for this trend was not presented.

While these computational studies have provided important insight, they also highlight a significant challenge in studying SC lipids with molecular simulation. Specifically, due to the dense packing of the lipid tails, the CER-rich bilayer systems demonstrate negligible molecular diffusion at physiologically relevant temperatures (*i.e.*, 305 K). Even in simulations conducted at elevated temperatures, *e.g.*, 340 K, studies have demonstrated minimal in-plane diffusion of the lipids, despite considering simulations in excess of 200 ns.<sup>33</sup> Additionally, lipids in the gel phase have low rates of rotational relaxations.<sup>40</sup> As such, on typical simulation timescales, systems likely exist in metastable configurations that are highly dependent on the initially bilayer structure. While this may be of little consequence for single component lipid bilayers, many properties of multicomponent lipid mixtures depend on the in-plane morphology of the individual lipid components, *e.g.*, aggregation versus dispersion of lipid species in the bilayer leaflets; furthermore, the morphology on this size scale is typically not known from experiment.<sup>41</sup> While self-assembled bilayers would, in principle, avoid these problems, it is not currently practical to study the self-assembly of multicomponent bilayers with atomistically detailed molecular models due to the high computational cost and a rough free energy landscape that could hinder the formation of equilibrium structures; for example, Das *et al.* found inverted micelle structures formed in mixtures of three different CERs with CHOL and FFA unless constrained by a wall,<sup>28</sup> in contrast to experiments, where lamellar structures form in bulk. Therefore, since preassembled structures remain more practical for simulating atomistic models, care must be taken

that system properties are not biased by the initial configuration, so that reliable conclusions can be made.

In this Chapter, molecular dynamics (MD) simulations are used to examine the structural behavior of model SC membranes, similar to those studied experimentally,<sup>20,42</sup> to examine the role of the CER FA tail length and overall lipid composition on structural properties. This chapter specifically focuses on CER NS, as this is the most abundant CER species in SC and is typically used in model systems. Three sets of systems are studied: pure CER NS, binary mixtures of CER NS and CHOL in 2:1 and 1:1 molar ratios, and equimolar mixtures of CER NS, CHOL, and FFA C24:0. The structure of each of the lipids used in this work is shown in Figure 4.1. Mixtures of different length CERs are used in each set to examine the impact of the CER FA tail length on the systems, which, to date, has not been the focus of any computational studies. To ensure robust results, a random walk MD (RWMD) algorithm is validated and employed to reduce the dependence of the final bilayer configuration on the assumed, initial morphology.

## 4.2 Material and Methods

### 4.2.1 Model

CER NS consists of a sphingosine base, 18 carbons in length, linked to a saturated fatty acid (FA) chain of variable length. Two different FA tail lengths were considered: CER NS C16:0 (Figure 4.1a), in which the FA tail is 16 carbons long and approximately equal in length to the sphingosine chain (this is denoted as equal-length CER NS, or eCER); and CER NS C24:0 (Figure 4.1b), in which the FA tail is 24 carbons long, a typical length in the SC (this is denoted as unequal-length CER NS, or uCER).<sup>12</sup> FFA, when present in this Chapter, is always 24 carbons in length (Figure 4.1d), which has been shown to be a useful model for the FFAs in healthy human SC.<sup>12,43</sup> The fully atomistic CHARMM36 force field<sup>44</sup> supplemented by the CHARMM-compatible CER

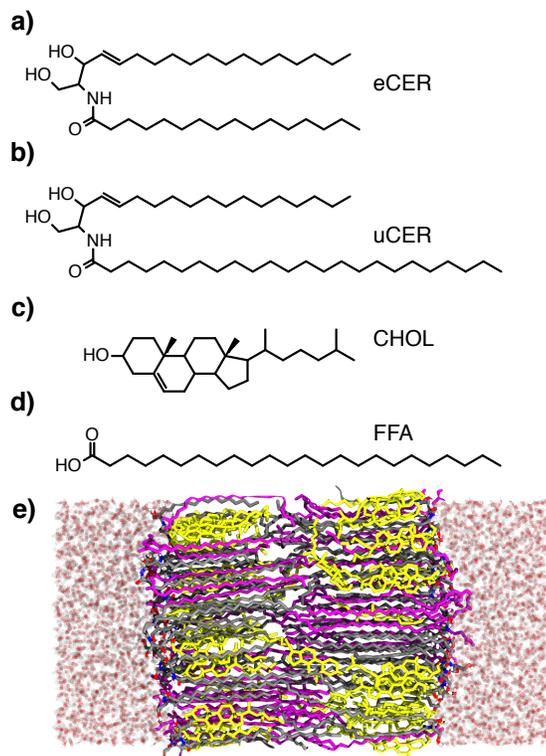


Figure 4.1: Chemical structure of the lipids simulated in this Chapter. a) CER NS C16 (eCER); b) CER NS C24 (uCER); c) cholesterol (CHOL); d) FFA C24:0 (FFA); and e) a snapshot of a typical configuration for the equimolar uCER-CHOL-FFA bilayer. CER is shown in gray, CHOL is shown in yellow, FFA is shown in purple, and water is shown in transparent red (oxygen) and white (hydrogen).

headgroup parameters from Guo *et al.*<sup>30</sup> were used to describe the CER NS, CHOL, and FFA. The CER headgroup parameters have been shown to accurately reproduce the thermotropic phase behavior of bilayers composed of pure CER NS,<sup>30</sup> and have been used to study the hydration of CER NS headgroups in solution.<sup>45,46</sup> Water was modeled using TIP3P.<sup>47</sup>

#### 4.2.2 Methods

Lipid monolayers were first constructed by placing 64 lipids on a rectangular lattice in the  $xy$  plane with  $50 \text{ \AA}$  per lipid. Four initial later distributions of a binary mixture of eCER and CHOL (*i.e.*, different arrangements of lipid species on the lattice) were considered for validating the ST-based approach: completely phase-separated,

consisting of one  $8 \times 4$  block of either lipid species (Figure 4.3a); a coarse-grained checkerboard, consisting of alternating  $4 \times 4$  blocks of each lipid species (Figure 4.3b); randomly mixed, where both lipid species are randomly dispersed throughout the lattice (Figure 4.3c), and a fine-grained checkerboard, consisting of alternating lipid species on the lattice sites (Figure 4.3d). To examine the impact of the CER NS FA chain length on the structural properties, each of the four lipid compositions studied (pure CER NS, 2:1 CER NS-CHOL, 1:1 CER NS-CHOL, and equimolar CER NS-CHOL-FFA) was subdivided into five subcompositions, where the CER NS fraction consists of 0, 25, 50, 75, and 100 mol% eCER. Thus, in total, 20 such systems were considered. For these simulations, the effect of the initial lateral distributions of the lipids was not explored, and the different lipid species were initially randomly dispersed throughout the lattice, with the number of each specific lipid dictated by the system composition.

In all cases, each lipid was rotated about its long axis by a random integer multiple of  $60^\circ$ , since a high degree of alignment between lipid backbones has been shown to cause unphysically large tilt angles in the lipid tails of gel-phase bilayers.<sup>40</sup> The monolayer was then rotated about the  $x$ -axis, and translated in  $z$  such that the ends of the tails in opposing leaflets were in contact, forming a bilayer; note, in this procedure each leaflet starts with the same in-plane configuration. Twenty water molecules per lipid were then added to hydrate the outside of the bilayers (2560 total water molecules for 128 total lipids). System sizes ranged from 19 136 to 24 192 atoms.

The systems were relaxed through a series of energy minimization, NVT simulations, and NPT simulations. First, a steepest decent energy minimization was performed on each initial configuration to reduce the large nonbonded repulsions caused by (inadvertently) overlapping atoms. Next, the systems were simulated for 10 ps in the NVT ensemble at 305 K (*i.e.*, skin temperature), followed by a 10 ns simulation in the NPT ensemble at 305 K and 1 atm. Unless otherwise noted, the systems were

further simulated using the RWMD algorithm. In the RWMD algorithm, the system temperature is adjusted at small time intervals, such that the system takes a random walk through temperature space. Specifically, the sequence is defined within an interval between  $T_{\min}$  and  $T_{\max}$ , with discrete temperatures defined every  $\Delta T$ . In this Chapter, RWMD was performed for 50 ns, with temperature changes of  $\Delta T = 5j$ , ( $j \in -1, 0, 1$ )K every 5 ps,  $T_{\min} = 305$  K, and  $T_{\max}$  set to 355 K for the first 25 ns, and the upper bound linearly reduced to 305 K over the final 25 ns. A representative plot of temperature versus time during the RWMD equilibration is shown in Figure 4.2, where it is worth noting that the sequence is defined such that it samples all temperatures equally. Systems were further simulated at 305 K for at least 150 K after RWMD equilibration. Note that 200 ns of simulation at a fixed temperature range are used for the comparison to the results obtained from simulations with RWMD in Section 4.3.1 (*i.e.*, for the RWMD validation, the 305 K systems were simulated for 200 ns at 305 K, and the RWMD systems were simulated for 200 ns with  $T_{\min} = 305$  K and  $T_{\max} = 355$  K). To avoid modification of the simulation code itself, the random walk was determined separately and the sequence of temperatures provided to the thermostat.

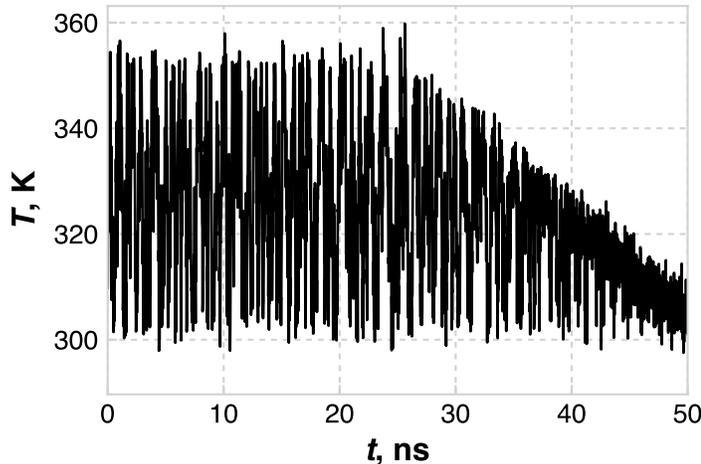


Figure 4.2: A representative plot of temperature versus time during the RWMD simulation.

The RWMD approach is designed to mimic key aspects of the Monte Carlo simulated tempering (ST) algorithm, used to find minima on a rough free energy surface by dynamically adjusting the temperature throughout the course of a simulation.<sup>48</sup> The general idea of the ST algorithm is that increasing the system temperature lowers free energy barriers, allowing different local minima to be explored each time the system returns from an elevated temperature to the (lower) temperature of interest, all while keeping the system at equilibrium because the acceptance criterion used to accept or reject temperature swaps ensures that the configurations sample the correct equilibrium ensemble.<sup>48</sup> The latter distinguishes this approach from the widely-used simulated annealing method, in which a system is driven out of equilibrium by increasing its temperature and simulating for a long time at the elevated temperature, followed by a slow cooling to the temperature of interest. Note that here, the random temperature walk is defined prior to the simulation, in contrast to the original ST algorithm, which involves attempting a temperature swap after each sweep in a Monte Carlo simulation using the standard Metropolis acceptance/rejection criteria, ultimately yielding a random walk through temperature space.

All simulations were performed in GROMACS 5.1,<sup>49</sup> employing the Nosé–Hoover thermostat<sup>50</sup> with a coupling constant of 1 ps and, for the NPT simulations, a Parrinello–Rahman barostat<sup>51</sup> with a coupling constant of 10 ps. The pressure was controlled semi-isotropically for all NPT simulations, where the box lengths of the bilayer lateral directions were coupled. van der Waals interactions were smoothly switched off from 10 Å to 12 Å, beyond which they were neglected. Long-range electrostatic interactions were treated via the PME algorithm<sup>52</sup> with a real-space cutoff of 12 Å. A timestep of 1 fs was used for all simulations. Separate thermostats were used for the lipids and the water, although both groups follow the same temperature walk to avoid introducing temperature gradients into the system.

### 4.2.3 Analysis

System properties were calculated over the final 100 ns of simulation at 305 K. Several properties were calculated to quantify the structure of the bilayers, including the area per lipid (APL), the tilt angle of the lipid tails with respect to the bilayer normal, the area per lipid tail (APT), the nematic order parameter of the lipid tails, the density profiles of various groups across the bilayer normal, the bilayer thickness, the width of the lipid–water interface, the width of the low density tail region, and the hydrogen bonding. The coordination numbers (CNs) of different lipid tail pairs in the bilayer plane are used to describe the lateral distributions of lipids; the center of mass of each tail is projected onto the  $z = 0$  plane for these calculations. Note that the tails were chosen for the CN calculations, as the lipid tails generally pack on a hexagonal lattice, in contrast to the relatively more disordered packing of the lipid headgroups. The lipid backbone orientation is used to describe the rotational motion of the lipids about their long axes. A detailed description of each of these calculations is given in Appendix A.

## 4.3 Results

### 4.3.1 Random Walk Molecular Dynamics

Before using the RWMD scheme in the rest of this Chapter, the efficacy of the approach is first evaluated for equimolar mixtures of eCER and CHOL. This composition was chosen as eCER and CHOL have similar hydrophobic lengths, which should yield bilayers without a considerable interdigitation region in the middle; prior work has shown that CHOL can reside in this interdigitation region in lipid mixtures with uCER.<sup>29</sup> Thus, the lipids (especially CHOL) in this mixture should stay in their canonical bilayer conformation, *i.e.*, with their headgroups at the lipid–water interface and their tails creating a hydrophobic core, therefore allowing evaluation of the ability of RWMD to enhance the in-plane lateral and rotational rearrangements of the lipids.

Figure 4.3 shows the initial configuration of each system studied, the final configurations from 200 ns simulations at 305 K, the final configurations after 200 ns of RWMD, and the evolution of the CHOL–CHOL CNs for each equilibration scheme. Note that the CHOL–CHOL CN was chosen because it is the least biased CN (*e.g.*, the FA–SPH CN is skewed because the FA and SPH tails are part of the same molecule, and hence will show a high level of association). Since lipids of the equimolar eCER–CHOL mixture are expected to mix, the fully separated and coarse-grained checkerboard morphologies (Figure 4.3a,b) serve as “bad” initial configurations, and the randomly mixed and fine-grained checkerboard systems (Figure 4.3c,d) serve as naïve, but reasonable, guesses of how to initialize mixed-lipid bilayers. Each initial morphology was equilibrated for 200 ns with two different schemes: standard MD at 305 K, and RWMD with  $305\text{ K} < T < 355\text{ K}$ . Both visual inspection and the evolution of the CHOL–CHOL CN can be used to compare the in-plane morphologies; and hence equilibration, of the different systems and equilibration procedures. At 305 K, the final morphologies visually resemble the corresponding initial morphologies; this is reflected in the steady-state nature of the CNs at 305 K, which are listed in Table 4.1. This trend is especially notable for the fully-separated and coarse-grained checkerboard systems, where the large CHOL aggregates make it easy to visually establish that there are only slight changes in the shape of the initial aggregates. These results highlight the frozen nature of these systems at 305 K; only small rearrangements occur in 200 ns if RWMD is not applied. Note that in the fine-grained (FG) checkerboard system, small linear aggregates of CHOL form relatively quickly, indicated by the increase in CN during the first 25 ns, although the system does not evolve much after this initial change.

With RWMD, the final lateral distributions of the lipids are visually distinct from the initial configuration, showing that the lipids were able to reorganize within the bilayer leaflets. For the two most separated systems, the larger CHOL aggregates mostly

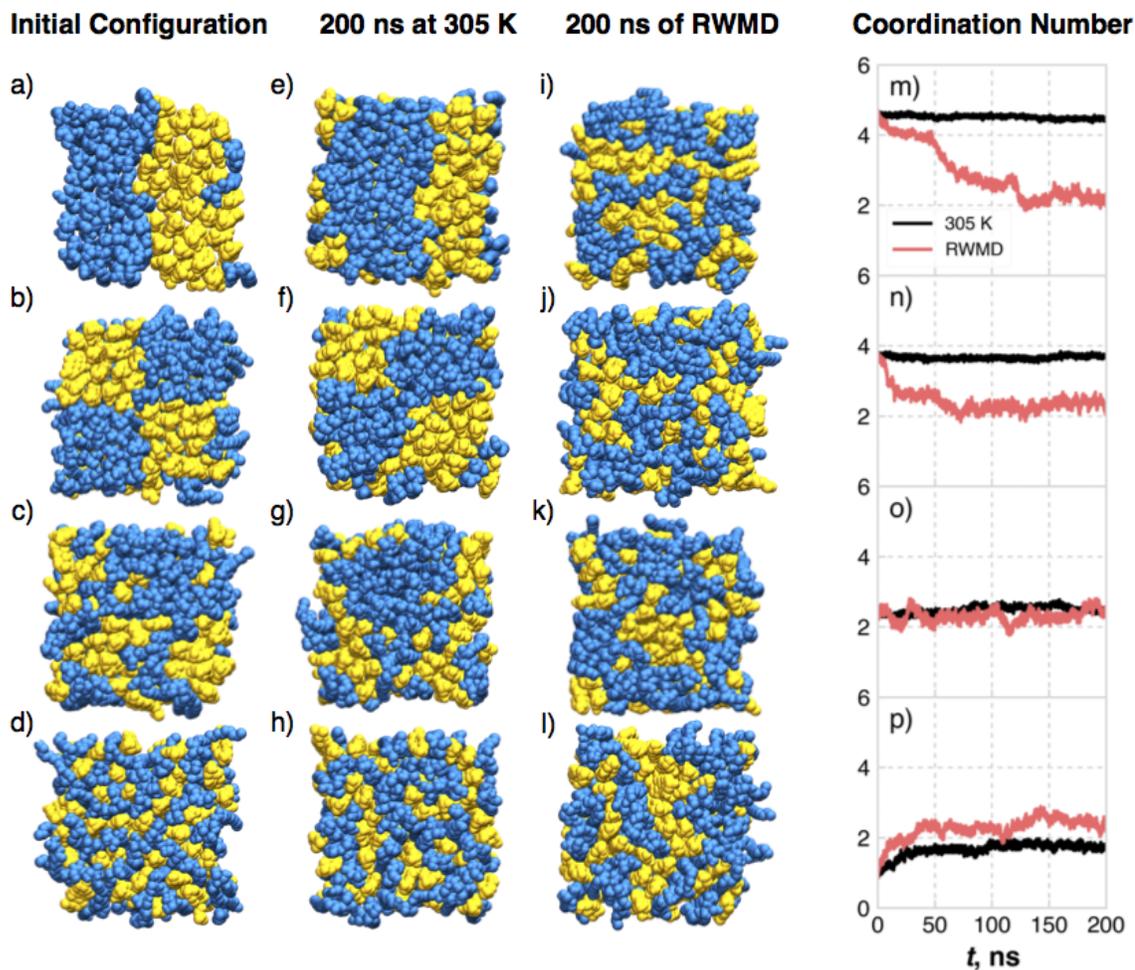


Figure 4.3: System configurations and CHOL–CHOL coordination number (CN) as a function of initial configuration and equilibration procedure for the equimolar mixture of eCER and CHOL. The first column shows the different initial configurations: a) maximum phase separation; b) coarse-grained checkerboard; c) randomly mixed; and d) fine-grained checkerboard. The second column shows the systems after 200 ns of MD at 305 K, the third column shows the systems after 200 ns of RWMD, and the fourth column shows the evolution of the CHOL–CHOL DN for each initial configuration. Snapshots are taken along the  $z$ -axis, *i.e.*, along the bilayer normal direction. eCER is represented as blue spheres, CHOL is represented as yellow spheres, and water is now drawn for clarity.

break up into smaller aggregates (Figure 4.3a,i and Figure 4.3b,j), while small CHOL aggregates form in the fine-grained checkerboard system (Figure 4.3d,l). Importantly, all of the final configurations after RWMD visually resemble each other and have nearly identical CN values, as listed in Table 4.1, and systems that start close to the

Table 4.1: CHOL–CHOL coordination number over the final 50 ns of MD simulation. *Separated* refers to the initial configuration shown in Figure 4.3a, *CG Checker* refers to that in Figure 4.3b, *Random Mix* refers to that in Figure 4.3c, and *FG Checker* refers to Figure 4.3d. The uncertainties here, and in all subsequent tables and figures, are given as the standard error of the measured property.

Equilibration	Initial Configuration			
	Separated	CG Checker	Random Mix	FG Checker
305 K	$4.46 \pm 0.03$	$3.72 \pm 0.04$	$2.53 \pm 0.08$	$1.75 \pm 0.07$
RWMD	$2.3 \pm 0.1$	$2.4 \pm 0.1$	$2.4 \pm 0.2$	$2.5 \pm 0.1$

final morphology (*i.e.*, the random and fine-grained checkerboard) rapidly converge to their final CN. Since each of these systems started from qualitatively different initial configurations, this result shows that the final configurations are decorrelated from the initial configurations and RWMD can reproducibly form the structures with matching properties without a strong bias from the starting configuration. It is worth noting that this does not necessarily indicate that RWMD has found the true free energy minimum of the system, but this does strongly suggest that it is a stable, low energy configuration, given that multiple independent trajectories converge to the same state.

Analysis of the rotational motion of the eCER molecules also suggests that the systems equilibrated with RWMD are more decorrelated from their initial configurations than those equilibrated at 305 K. Figure 4.4 shows the autocorrelation function of the lipid backbone angle in the bilayer plane for the different initial configurations and equilibration methods. The lipid backbone orientations become completely uncorrelated from the initial orientations by 25 ns of RWMD, while slower relaxation is observed for the systems simulated at 305 K. Interestingly, the phase-separated system shows the slowest relaxation of the eCER backbone orientations, likely because the large, dense eCER domain has a larger energy barrier for lipid rotation due to lipid–lipid hydrogen bonding. Since the systems that were equilibrated at 305 K all converge to different morphologies and the lipid backbone orientations are correlated

Table 4.2: Area per lipid (in  $\text{\AA}^2$ ) of systems as a function of initial configuration and equilibration methodology. The systems equilibrated with RWMD were simulated at 305 K for 50 ns for a more direct comparison with the other systems. *Separated* refers to Figure 4.3a, *CG Checker* refers to Figure 4.3b, *Random Mix* refers to Figure 4.3c, and *FG Checker* refers to Figure 4.3d.

Equilibration	Initial Configuration			
	Separated	CG Checker	Random Mix	FG Checker
305 K	$38.8 \pm 0.4$	$38.4 \pm 0.4$	$39.4 \pm 0.2$	$39.6 \pm 0.2$
RWMD	$39.6 \pm 0.5$	$39.7 \pm 0.4$	$39.5 \pm 0.4$	$39.4 \pm 0.3$

to the initial orientations, at least three of the four systems must be metastable configurations. Often in bilayer simulations, a specific property, such as the area per lipid (APL), is monitored over time, and equilibrium is assumed when that property reaches a steady state. This practice, however, may lead to spurious assumptions of equilibrium since other features of the bilayer structure may not exist in the preferred state, but this cannot be easily verified since the preferred state is not always known *a priori*. For example, the APLs of the systems equilibrated at 305 K are compared to the APLs of the systems equilibrated with RWMD (after running at 305 K for a more direct comparison) in Table 4.2, from which it is apparent that the systems equilibrated at 305 K show a larger spread in the APLs than the systems equilibrated with RWMD. The two systems with the highest level of CHOL aggregation have similar APLs, while the other two systems have lower APLs that are similar to each other. Due to the different morphologies and packing densities of the lipids in these systems, they likely have different properties, *e.g.*, permeability, which is related to the in-plane density of the lipids.<sup>53</sup> In contrast, all of the systems equilibrated with RWMD have APLs consistent within error, further illustrating the reproducibility, and hence confidence gained, when equilibrating with RWMD.

This lack of mobility and difficulty in proper equilibration is a recognized problem when simulating CER-based bilayers, with many studies running the systems at

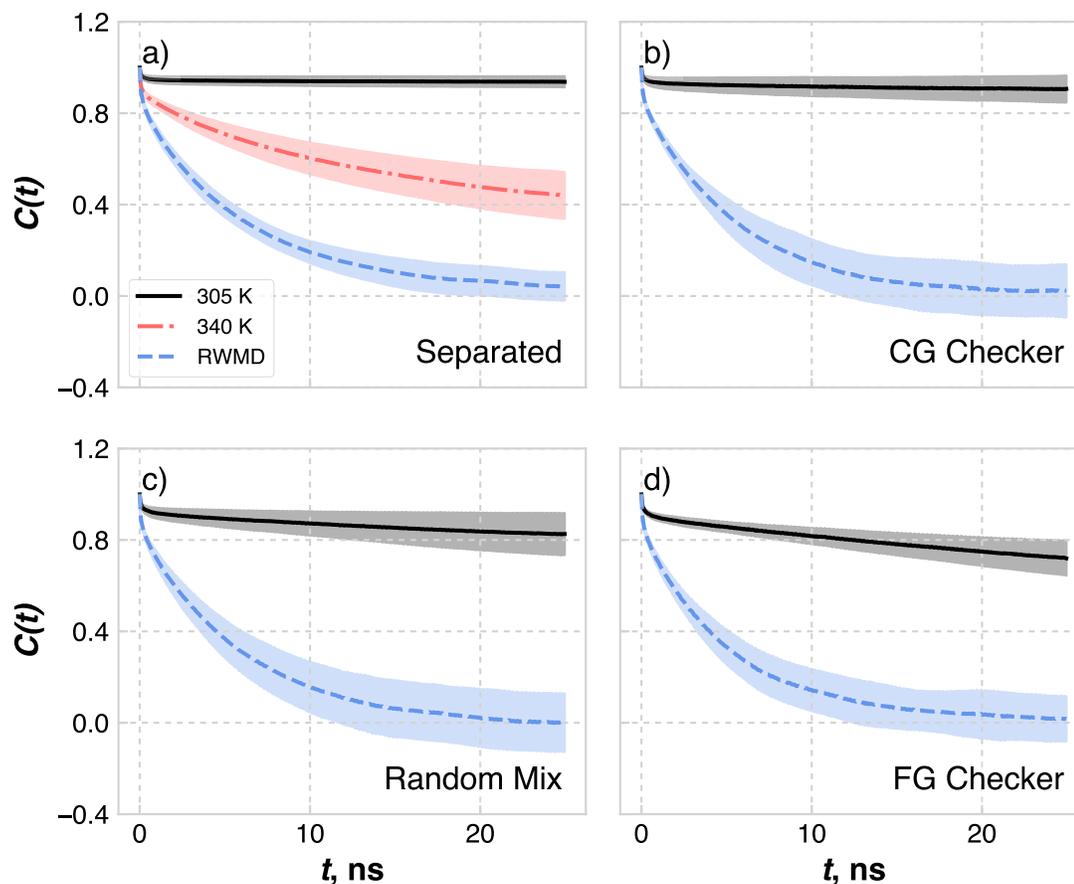


Figure 4.4: Autocorrelation function  $C(t)$  of the eCER backbone orientations during simulations at 305 K and with RWMD from different initial configurations. *Separated* refers to Figure 4.3a, *CG Checker* refers to Figure 4.3b, *Random Mix* refers to Figure 4.3c, and *FG Checker* refers to Figure 4.3d. The shaded area shown is the standard deviation over the rotational autocorrelation function of all eCER molecules in the bilayer.

elevated temperatures in an attempt to overcome mobility issues.<sup>26–29,31,32,35,36</sup> Here, I also compare to equilibration and production runs at 340 K, as several MD studies of SC lipids have used this temperature.<sup>26–29</sup> Figure 4.5 shows the evolution of the CN for the fully separated system (Figure 4.3a) simulated for 600 ns at 305 K, 340 K, and with RWMD. As expected, the system at 305 K is still *frozen* in simulations in excess of 0.5  $\mu$ s, while the system simulated with RWMD reaches its steady state SN at  $\sim$ 110 ns. Perhaps unsurprisingly, the system simulated at 340 K shows behavior somewhere

in between that of the other systems; while lateral reorganization does occur, the CN reaches a steady state value after  $\sim 520$  ns, which is significantly longer than the simulation times typically studied. Note that this system also reaches a pseudosteady state CN between 80 ns to 150 ns; with many studies performed on simulations run for 100 ns or less, this metastable state could be erroneously mistaken for equilibrium, if one were relying on CN to determine convergence. At 600 ns, the rotational relaxation of the lipids for this system at 340 K is higher than 305 K, but still lower than RWMD (Figure 4.4a). While the fully phase-separated system is likely a poorly chosen starting configuration given some level of prior knowledge about this particular system, the actual morphology is typically not known *a priori*. Thus, these results indicate that RWMD is more efficient than simply running at a high temperature to equilibrate multicomponent, gel-phase lipid bilayers and can provide increased confidence in the reproducibility of the results, particularly for systems in which the lateral organization is unknown and expected to play an important role in the properties of interest.

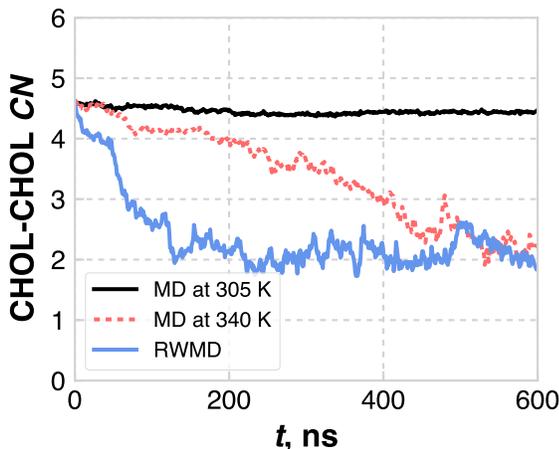


Figure 4.5: Evolution of the CHOL–CHOL coordination number (CN) for the fully phase-separated systems equilibrated with three different methodologies as a function of time ( $t$ ).

### 4.3.2 Dependence of Structural Properties on Bilayer Composition

The effect of the composition of CER-based lipid bilayers on structural properties is now examined. Note, all systems start from a randomly dispersed morphology, given that, in Section 4.3.1, this arrangement relaxed the fastest and appears to be reasonably representative of the morphology for this family of systems; these systems were also relaxed using RWMD as described in Section 4.2.

*General Structural Properties.* All systems exist in dense, highly ordered, gel phases, as indicated by the high nematic order parameters ranging from 0.95 to 0.99, as shown in Figure B.1. The addition of CHOL decreases the nematic order, consistent with Das *et al.*,<sup>27</sup> but the magnitude of this decrease is quite small, (*e.g.*, ranging from 0.988 for pure uCER to 0.954 for 1:1 CER:CHOL with 3:1 eCER:uCER) and the lipids are still highly ordered from visual inspection (Figure 4.1e), indicating no phase change upon addition of CHOL for these compositions.

The tilt angle shows a weak dependence on composition, as shown in Figure 4.6a. The tilt angle depends significantly on the CER FA tail length for the pure CER systems, where a strong increase is seen with increasing eCER for systems with 50% or more eCER. This trend is a result of the balance between headgroup and tail interactions: steric repulsions between lipid headgroups dictate the APL, while van der Waals attractions between the lipid tails cause them to tilt to optimize their spacing.<sup>54</sup> For the CER bilayers that are predominantly uCER, the tail length asymmetry gives rise to a wide *low density tail region* in the center of the bilayer (discussed below), which increases the effective optimal tail packing, leading to a lower tilt angle. The addition of CHOL and FFA causes the tails to tilt less and show almost no change as a function of increasing eCER, likely because CHOL has a bulky ring structure with a small headgroup, and therefore acts as a spacer between the lipid tails, buffering the head–tail effects seen in the pure systems. Hence, there is a similar mismatch between the packing densities of the headgroups and tails, leading to a smaller tilt.

Notably, the 1:1 CER:CHOL and ternary mixtures have very similar tilt angles, which is attributed to the fact that these systems have a similar fraction of alkyl tails.

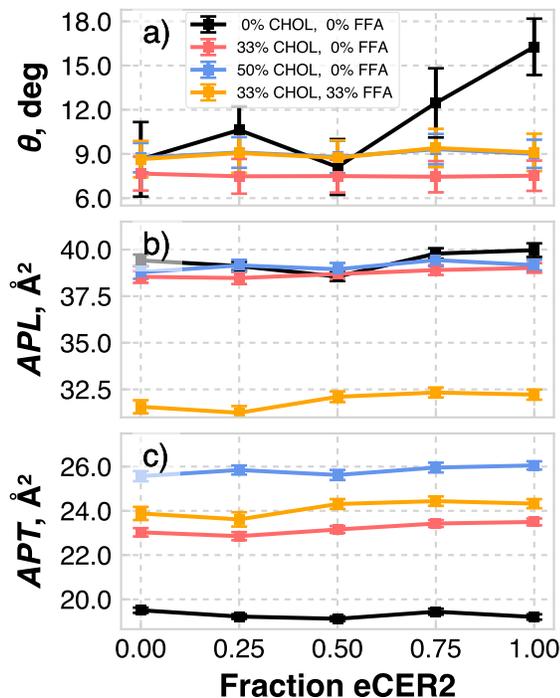


Figure 4.6: (a) Average tilt angle ( $\theta$ ) of the lipid tails with respect to the bilayer normal; (b) area per lipid (APL) and (c) area per tail (APT) as a function of eCER composition. Note that the abscissa is the fraction of CER NS that is eCER, not the total eCER fraction in the system.

The APL and APT trends as a function of the fraction of eCER are shown in Figure 4.6b,c. While the APL describes the density of lipids in the bilayer plane, irrespective of the number of tails that each lipid has, the APT describes the tail packing density in the plane of the lipid tails (*i.e.*, APT accounts for both the number of tails and the tilt angle). All CER NS and CER NS-CHOL systems have similar APLs, from  $38 \text{ \AA}^2$  to  $40 \text{ \AA}^2$ , as expected since CER and CHOL have similar cross-sectional areas.<sup>55</sup> There is only a slight, nonmonotonic dependence of the APL on the CER FA tail length. This trend is expected since the headgroups dictate the APL and these systems have the same headgroups for a given composition. Systems containing

FFA have smaller APLs, since FFA contains a single alkyl chain. The APT increases with the CHOL content, as CHOL and CER NS have similar cross-sectional areas, but CHOL is treated as a single tail. The ternary systems have APTs more similar to the 2:1 CER NS:CHOL systems, which is reasonable considering they have the same ratio of CHOL to alkyl chains. The mixed lipid systems all show a slightly increasing APT with increasing eCER fraction. Interestingly, there appears to be some level of APT non-additivity in these systems. For example, consider the APL and APTs of the 1:1 CER NS:CHOL systems. Since they have similar APLs as the pure CER NS systems, but  $3/4$  the number of tails, ideal mixing would yield APL values that are  $4/3$  that of pure CER NS values. However, the actual values are lower, suggesting a relative attraction between CHOL and the saturated tails of CER NS, similar to what was observed for dimyristoylphosphocholine (DMPC, a phospholipid with saturated tails), but not dioleoylphosphocholine (DOPC, a phospholipid with unsaturated tails).<sup>56</sup> Although perhaps expected based on chain saturation, this behavior is nontrivial for CER systems, since the CER NS headgroups are much smaller than PC headgroups and CHOL presumably has less empty space to occupy (compared to bilayers containing unsaturated lipids).

*Density Profiles.* The bilayer thickness, calculated from the water density profiles, is a function of both overall lipid composition and CER FA tail length, as shown in Figure B.2. The bilayer thickness decreases with increasing eCER and CHOL content, as one would expect since these lipids have shorter tail lengths than uCER or FFA. However, the tail length asymmetry leads to a richer behavior, which can be observed via examination of the lipid density profiles, as shown for each system in Figure 4.7.

The pure CER bilayers exhibit mass density profiles that contain the expected features of a lipid bilayer. Near the lipid–water interface, the profiles contain a peak (Figure 4.7a), which represents the heavy atoms in the CER headgroups. Just inside the headgroup region, there is a high density tail region where the tails are densely

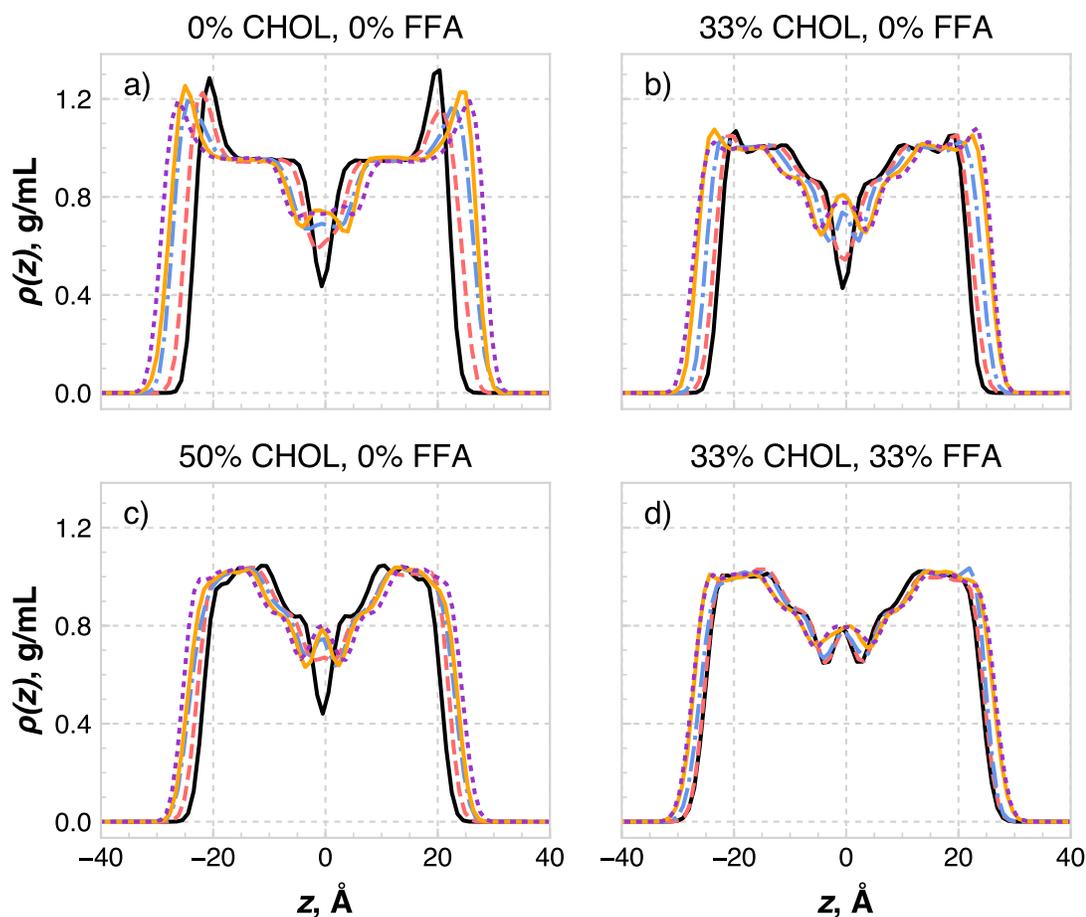


Figure 4.7: Total lipid mass density profiles across the bilayer for all systems considered: a) pure CER NS, b) 2:1 CER NS:CHOL, c) 1:1 CER NS:CHOL, d) equimolar CER NS:CHOL:FFA. The profiles for the 5 subcompositions are shown for all compositions and are labelled as follows: solid black line, 100 mol % eCER; dashed red line, 75 mol % eCER; dot-dashed blue line, 50 mol % eCER; solid orange line, 25 mol % eCER; dotted purple line, 0 mol % eCER. Note that these percentages denote the fraction of CER NS that is eCER, not the total amount of eCER in the system.

packed and highly ordered (see discussion of nematic order above), consistent with prior work.<sup>38</sup> In the center of the bilayer is a low density tail region, where the lipid tails are less densely packed and less ordered. This region exists primarily due to the asymmetry in the lengths of the tails in the systems and is composed of the terminal part of the uCER tail and FFA tails, as shown in Figure B.3. In the systems with less asymmetry (*i.e.*, high eCER and CHOL concentrations), the profile is more V-shaped

in the middle since there is little interdigitation between tails in opposing leaflets. Comparing the high and low density tail regions for the different bilayers, the width of the high density tail region is found to be nearly constant across all compositions, whereas the width of the low density tail region varies with eCER and CHOL content. Figure 4.8 shows the total bilayer thickness as a function of the thickness of this low density tail region in the center of the bilayer. Clearly, the thickness of the low density tail region dictates the total bilayer thickness for a given composition. The width of the high density tail region is determined by the smallest hydrophobic length in the system, where the tails must pack densely to fit in the area dictated by the headgroups. For the lipids in this study, this is roughly 15 carbons, or the length of the sphingosine chain that is not part of the CER NS headgroup, which is also similar to the length of a CHOL molecule.<sup>57</sup>

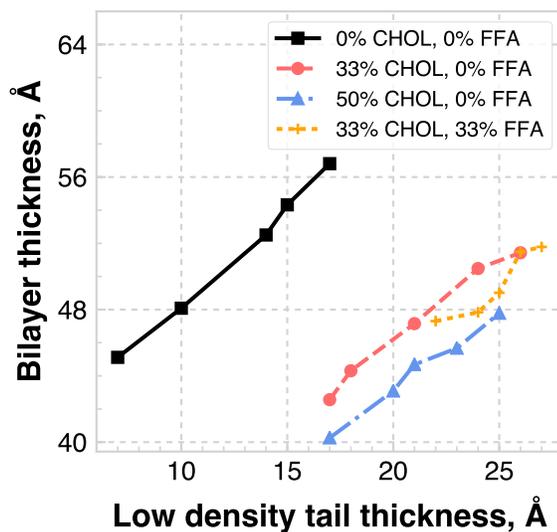


Figure 4.8: Total bilayer thickness as a function of the thickness of the low density tail region, with the different points on each line denoting the different eCER fractions; the smallest values indicate 100 mol % eCER, and the largest 0 mol % eCER.

The presence of CHOL and FFA has two main effects on the density profiles. First, the smaller headgroups of CHOL and FFA lead to smaller peaks in the headgroup region, compared to a pure CER NS bilayer; these peaks are dramatically reduced

going from pure CER NS to 2:1 CER NS:CHOL and are mostly absent in the 1:1 CER NS:CHOL and ternary systems. Second, the width of the low density tail region increases with the addition of CHOL and FFA, as shown in Figures 4.7 and 4.8. Additionally, the low density tail regions tend to have more features in the mixed lipid systems, *e.g.*, shoulders representing intermediate densities due to the different tail lengths, Peaks also appear in the middle of the density profiles for the mixed systems, due to both the long FA chain of uCER (and FFA for the 3 component systems) and the presence of CHOL that has migrated from the ordered bilayer region into the low density tail region.<sup>29</sup> In the eCER-CHOL systems, the lack of significant interdigitation means this peak is absent; however, all systems containing FFA demonstrate a peak in the middle of the density profile because of the long tail of FFA.

Figure 4.9 shows the density profiles for the ternary system with 1:1 eCER:uCER, along the bilayer normal direction, broken down by the contribution of specific lipid components. The CHOL headgroup sits  $\sim 3 \text{ \AA}$  deeper into the bilayer with respect to the CER headgroups, which is consistent with prior simulation and experimental studies.<sup>24,27,32</sup> There is also a small subpopulation of CHOL lying flat in the middle of the bilayer, illustrated by the peak in the middle of the CHOL ring density profile, which is consistent with the work of Das *et al.*<sup>29</sup> Additionally, the FFA headgroups tend to sit slightly further into the water than the CER headgroups.

The interfacial width, which can indicate the hydrophobicity of the surface (*e.g.*, a larger interfacial width would indicate a more hydrophilic surface), is defined as the width of the region over which the water density drops from the bulk value to  $1/e$  of the bulk value. The bilayers in this work appear hydrophobic, having interfacial thicknesses ranging from  $3.5 \text{ \AA}$  to  $6 \text{ \AA}$ , compared to phospholipid bilayers, which are generally greater than  $10 \text{ \AA}$ .<sup>58</sup> There is little dependence on the bilayer composition or CER FA tail length, as shown in Figure B.4. Thus, one can conclude that while the

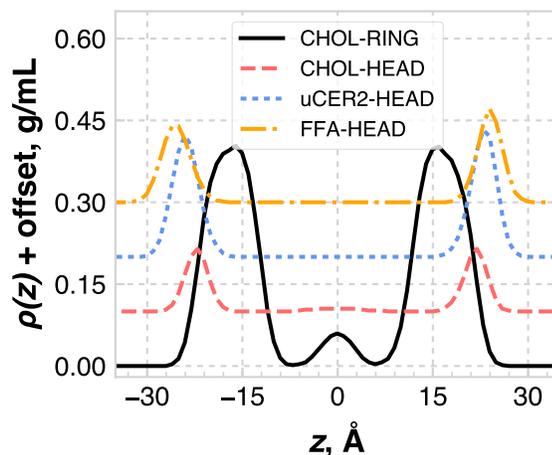


Figure 4.9: Mass density profiles of various groups in the ternary bilayer with a 1:1 eCER:uCER composition. Note that the profiles for different groups are shifted vertically for clarity. Each curve is offset by 0.1 g/mL for each subsequent dataset in the legend.

lamellar organization near the middle of the bilayer depends on the lipid composition, the apparent hydrophobicity of the lipid–water interface does not.

*Hydrogen Bonding.* The total number of lipid–lipid hydrogen bonds in each system is listed in Table 4.3. The lipid–lipid hydrogen bonding is a function of the total CER content, with a reduction in CERs resulting in fewer lipid–lipid hydrogen bonds because CHOL and FFA molecules have fewer hydrogen bonding sites. The lipid–lipid hydrogen bonding is also independent of the eCER fraction, which may be expected since eCER and uCER have identical headgroups and their relative position at the interface is essentially unchanged as a function of the eCER fraction.

The hydrogen bonding between specific pairs of lipids for the ternary mixture with 1:1 eCER:uCER is listed in Table B.1. Negligible CHOL–CHOL hydrogen bonding is observed, which can be rationalized by the fact that CHOL only has a single hydrogen bond donor and acceptor, and two neighboring CHOL molecules would have to adopt a strained configuration to form a hydrogen bond. There are also very few CHOL–FFA hydrogen bonds, likely a result of CHOL sitting deeper into the bilayer.

Table 4.3: Total number of lipid–lipid and lipid–water hydrogen bonds for each system studied.

Composition		Fraction eCER				
		0	0.25	0.50	0.75	1.0
Lipid–lipid	Pure CER NS	63	62	61	63	63
	2:1 CER NS:CHOL	40	37	38	39	38
	1:1 CER NS:CHOL	28	26	28	28	27
	Ternary	19	19	19	19	19
Lipid–water	Pure CER NS	173	168	173	176	171
	2:1 CER NS:CHOL	173	169	173	175	176
	1:1 CER NS:CHOL	168	170	164	167	173
	Ternary	154	152	158	158	155

An appreciable number of CER–CHOL hydrogen bonds are seen, with slightly more eCER–CHOL (4.3) than uCER–CHOL (3.7) hydrogen bonds. Interestingly, the level of CER–CHOL hydrogen bonding is similar in the 2:1 and 1:1 CER:CHOL systems with, for example, 11.4 and 11.1 total CER–CHOL hydrogen bonds for the 2:1 and 1:1 CER–CHOL systems with 50 mol% eCER (see Tables B.2 and B.3); this trend may suggest that the CER–CHOL hydrogen bonding is saturated at or below 33 mol% CHOL. Additionally, when CER is replaced by CHOL (*i.e.*, comparing the pure CER NS and the 2:1 CER NS:CHOL or 1:1 CER NS:CHOL systems), there are fewer hydrogen bonds relative to the total number of hydrogen bonding sites in the system, indicating that CER–CER hydrogen bonds are preferred over CER–CHOL hydrogen bonds. There are comparatively fewer CER–CHOL hydrogen bonds in the ternary systems, since FFA is present and is competing with CHOL to form hydrogen bonds with CER. There is a surprisingly low amount of FFA–FFA hydrogen bonding, given that FFA prefers to be near other FFA molecules (discussed below); however, FFA sits deeper into the water and thus forms more hydrogen bonds with water.

The amount of lipid–water hydrogen bonding in each system is also listed in

Table 4.3. Despite the relatively hydrophobic lipid–water interface (based on the low interfacial thickness), there are significantly more lipid–water hydrogen bonds than lipid–lipid hydrogen bonds. As with the lipid–lipid hydrogen bonds, the lipid–water hydrogen bonding is independent of the CER FA tail length. Interestingly, the pure CER and 2:1 CER:CHOL systems have similar amounts of lipid–water hydrogen bonds, despite the fact that CHOL has significantly fewer hydrogen bonding sites than CER, while the 1:1 CER:CHOL and ternary systems have comparably less lipid–water hydrogen bonds, as expected since these systems have less CER (and hence fewer hydrogen bonding sites) than the pure CER and 2:1 CER:CHOL systems. This trend in lipid–water hydrogen bonding is perhaps explained by the “spacer” effect of CHOL, as discussed above with respect to the tilt angles. Although less CER results in fewer available hydrogen bonding sites, the spacing effect of CHOL allows the water to more thoroughly hydrate the CER headgroups and hence results in more lipid–water hydrogen bonds *per hydrogen bonding site*.

*In-plane Morphology.* For a general view of the in-plane morphology of each system, the coordination numbers (CNs) between specific lipid tails in the bilayer plane are examined. The CHOL–CHOL CNs depend on the amount of CHOL and available neighbors in a given system, and not on the CER FA tail length, as detailed in Figure B.5. This result suggests a random distribution of CHOL throughout the bilayer leaflets. For example, there are equal numbers of three different types of tails in the 1:1 CER:CHOL systems (*i.e.*, CER–FA, CER–SPH, and CHOL). Random mixing would give the observed CHOL–CHOL CN of 2. Applying this logic to the 2:1 CER:CHOL and ternary systems, random mixing would imply CHOL–CHOL CNs of 1.2 and 1.5 for each composition, respectively, which is observed. In contrast to CHOL, FFA shows a preference for specific neighbors, and thus has a less random distribution throughout the bilayer leaflets. In the uCER:CHOL:FFA system, the FFA–CER<sub>FA</sub> and FFA–CER<sub>SPH</sub> CNs are  $1.39 \pm 0.08$  and  $1.20 \pm 0.07$ , respectively (where CER<sub>FA</sub>

and  $\text{CER}_{\text{SPH}}$  represent the fatty acid and sphingosine tails of CER, respectively). This result suggests that FFA has a preference for the FA chain of CER compared to the SPH chain. Additionally, a significant preference for FFA to be near other FFA molecules is observed, with a FFA–FFA CN of  $2.3 \pm 0.1$ , compared to  $1.39 \pm 0.08$  for FFA– $\text{CER}_{\text{FA}}$ , the next highest. This trend has also been observed experimentally, with FFA-enriched domains forming with the tails tightly packed on an orthorhombic lattice.<sup>20</sup> However, unlike experimental results on similar systems,<sup>42</sup> we do not see any chain length-dependent behavior in regards to the mixing of CER and FFA. We note that comparable experimental systems consider multilamellar structures, whereas we consider a single hydrated bilayer, which could account for the differences observed. We also note that the systems studied in the current work are several orders of magnitude smaller than comparable experimental systems (nm versus  $\mu\text{m}$ ), and that an in-depth study of lateral distributions of lipids is beyond the scope of this work.

#### 4.4 Discussion

*Relevance to Experimental Models of the Short Periodicity Phase.* The localization of different groups within an experimental model of the short periodicity phase (SPP) of the SC has been studied with neutron diffraction.<sup>23,24</sup> These experimental systems resemble the systems studied in this Chapter, with a few differences. First, this Chapter only considers CER NS, whereas experimental systems tend to include a mixture of CERs with different headgroups, although uCER typically accounts for the majority ( $\sim 60\%$ ) of the CERs.<sup>12,23,24</sup> Secondly, we also only consider FFA C24, whereas experimental systems often include FFAs with a distribution of tail lengths. We note, however, recent work has shown that CER headgroup chemistry does not play a strong role in the lamellar organization of SPP models, and that systems with a distribution of FFA tail lengths had the same lamellar organization as a similar system with only FFA

C24:0.<sup>12</sup> Perhaps the biggest different between experimental systems and the systems studied here is that this Chapter only considers single bilayers, which is the typical approach to studying these systems with molecular simulations. Experimental systems usually contain membranes with numerous diffraction orders detected, indicating multilayer structures. Nonetheless, the neutron scattering length density (NSLD) profiles from experiment and calculated via simulation can be compared. Additionally, the localization of specific groups in the experimental and simulation systems can be compared.

In Figure 4.10, the total NSLD profiles from experiment<sup>23,24</sup> and simulation are compared. The most obvious features of each are the large peaks at  $\pm 27 \text{ \AA}$  (experiment) and  $\pm 25 \text{ \AA}$  (simulation). Given the experimental resolution of  $\sim 5.4 \text{ \AA}$ , the locations of these peaks are in good agreement. It is worth noting that the NSLD profiles calculated from simulation were averaged over multiple subtrajectories from the simulation, and the calculated uncertainty was smaller than the line thickness; this small error is likely due to the fact that the simulation samples 10s of  $\text{\AA}$  over 10s of ns, so small variability in density profiles should be expected. Both simulation and experiment show valleys in the NSLD at  $\sim 7 \text{ \AA}$ , which notably align with the dips in the low tail density region from Figure 4.7d. Additionally, subtle shoulders in the NSLDs are present at  $\pm 15 \text{ \AA}$  in both simulation and experiment, which aligns with the edges of the low tail density region in Figure 4.7d.

Focusing next on the localization of specific groups within the bilayer, the mass density profiles from simulation and the difference profiles between the protonated and deuterated NSLD from experiment are compared, and shown in Figure B.6. In this manner, the localization of specific groups in the simulations and experiments can be compared. Comparing the location of the FA tail of uCER that was measured in Groen *et al.*,<sup>23</sup> good agreement is found with the simulation result, with a single broad peak spanning  $\pm 7.5 \text{ \AA}$  (Figure B.6), which corresponds to the region where the

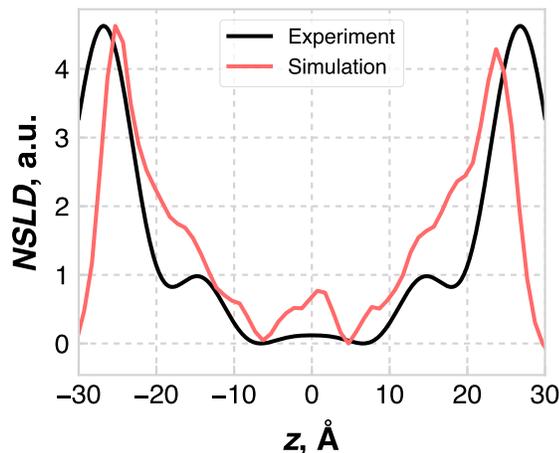


Figure 4.10: Neutron scattering length density (NSLD) profiles, reconstructed from experiment<sup>23</sup> and calculated from simulation. Note that the profiles were shifted such that the minimum lies at 0, and the simulation curve was scaled to have the same height as the experimental curve.

uCER FA tails from opposing leaflets interdigitate. The CHOL tails are localized at  $\pm 2 \text{ \AA}$  in both simulation and experiment (Figure B.6).<sup>24</sup> The CHOL headgroups show some deviation between experiment and simulation, localized at  $\pm 23 \text{ \AA}$  and  $\pm 21 \text{ \AA}$ , respectively (Figure B.6). This discrepancy, however, is again within the experimental resolution of  $5.4 \text{ \AA}$ .<sup>24</sup> Despite the fact that this work studies single bilayers, the locations of specific groups within the bilayer agree very well with experimental data on systems with similar compositions. Therefore, it can be concluded that the model systems accurately approximate the model SPP systems from experiment, and conclusions from this work also likely apply to model SPP systems, and also the SPP in the SC.

*Effects on Barrier Properties of SC.* In this study, the lipid composition and CER FA tail length were found to have the largest impact on lamellar organization, with the bilayer thickness and shape of the density profile in the low density tail region most affected. If we consider how these changes would affect the barrier properties (*i.e.*, the permeability) of these bilayers, we can provide some insight into the observed

experimental behavior. The permeability of a membrane is a product of the solute diffusion and partitioning, where the partitioning has an exponential dependence of the free energy of solvation. Due to this exponential dependence, small changes in the local packing (*i.e.*, depth-dependent density) can have a dominant impact on the bilayer permeability. Since the lipid mass density profiles are not constant across the bilayers, the resistance experienced by a permeant molecule would vary with the bilayer depth. Moreover, since the low density tail regions are qualitatively different for the bilayers with different compositions, these depth-dependent partition coefficients would also change, albeit nontrivially. Thus, if one was interested in comparing the permeability of these different systems, a simple homogeneous solubility–diffusion model would not suffice. This is indeed not unrecognized in the field, as the few simulation studies of SC lipid permeability in the literature to date have used an inhomogeneous solubility–diffusion model.<sup>31,53</sup>

While the structure of the lipids at the lipid–water interface changes with composition, it does not change with CER FA tail length. This is illustrated by the fact that the thickness of the interfacial region and the lipid density profiles near the headgroups are unchanged with the CER FA tail length for a given composition (Figure 4.9 and Figure B.4). Additionally, both the total number of lipid–water and lipid–lipid hydrogen bonds is constant for a given composition. Thus, one should expect any changes in the barrier properties of model SC membranes with CER FA tail length to be a result of changes in the structure of the bilayers in the hydrophobic core, and that changes in the headgroup region would play only a small role. This observation is consistent with the permeability measurements reported by Uchiyama *et al.*, which show that the permeability of ethyl-*p*-aminobenzoic acid through membranes composed of synthetic mixtures of CER, CHOL, and FFA is independent of the CER composition, but strongly dependent on the FFA tail length dispersity.<sup>12</sup>

## 4.5 Conclusions

This Chapter used molecular dynamics simulations to examine the structure of lipid bilayers relevant to the stratum corneum of the layer of the skin, and a protocol to thoroughly relax such systems is proposed and validated. I first illustrated several difficulties in simulating SC lipid mixtures and multicomponent gel-phase bilayers in general, *e.g.*, systems tend to be frozen at physiological temperatures with only small lateral rearrangements possible. Since the most realistic morphologies are not generally known *a priori*, systems must be allowed to find the most realistic morphology. This task is shown to be computationally impractical at physiological temperatures and inefficient at commonly used elevated temperatures, as well as likely to lead to metastable states being confused to equilibrium states. The proposed random walk MD methodology, based on simulated tempering, was shown to be an efficient and reproducible method for minimizing the influence of an assumed, initial membrane configuration for multicomponent, gel-phase bilayers. One should therefore expect this method to be especially useful as simulations incorporate more complex lipid mixtures. Using the random walk MD protocol, a series of SC lipid bilayers with varying lipid compositions and CER FA tail lengths was examined. It was shown that the lamellar organization is most affected by the lipid composition and CER FA tail length. Subtle changes to the lateral organization of the lipid tails as a function of the CER FA tail length were also seen. Additionally, since it was observed that the behavior at the lipid–water interface does not change with the CER FA tail length, it is reasonable to speculate that any tail length-dependent changes in barrier function are a result of changes in the lipid tail region and not the headgroup region.

## 4.6 Bibliography

- [1] MOORE, T. C., HARTKAMP, R., IACOVELLA, C. R., BUNGE, A. L., AND M<sup>c</sup>CABE, C. The influence of ceramide tail length on the structure of bilayers composed of stratum corneum lipids. *Biophysical Journal* **in press**.
- [2] MADISON, K. C. Barrier function of the skin: "la raison d'être" of the epidermis. *Journal of Investigative Dermatology* **121**, 2003, pp. 231–241.
- [3] WEERHEIM, A. AND PONEC, M. Determination of stratum corneum lipid profile by tape stripping in combination with high-performance thin-layer chromatography. *Archives of Dermatological Research* **293**, 2001, pp. 191–199.
- [4] JANSSENS, M., VAN SMEDEN, J., GOORIS, G. S., BRAS, W., PORTALE, G., CASPERS, P. J., VREEKEN, R. J., HANKEMEIER, T., KEZIC, S., WOLTERBEEK, R., LAVRIJSEN, A. P., AND BOUWSTRA, J. A. Increase in short-chain ceramides correlates with an altered lipid organization and decreased barrier function in atopic eczema patients. *Journal of Lipid Research* **53**, 2012, pp. 2755–2766.
- [5] LAVRIJSEN, A. P. M., BOUWSTRA, J. A., GOORIS, G. S., WEERHEIM, A., BODDÉ, H. E., AND PONEC, M. Reduced skin barrier function parallels abnormal stratum corneum lipid organization in patients with lamellar ichthyosis. *Journal of Investigative Dermatology* **105**, 1995, pp. 619–624.
- [6] PILGRAM, G. S. K., VISSERS, D. C. J., VAN DER MEULEN, H., KOERTEN, H. K., PAVEL, S., LAVRIJSEN, S. P. M., AND BOUWSTRA, J. A. Aberrant lipid organization in stratum corneum of patients with atopic dermatitis and lamellar ichthyosis. *Journal of Investigative Dermatology* **117**, 2001, pp. 710–717.
- [7] VAN SMEDEN, J., JANSSENS, M., BOITEN, W. A., VAN DRONGELEN, V., FURIO, L., VREEKEN, R. J., HOVNANIAN, A., AND BOUWSTRA, J. A. Intercellular skin barrier lipid composition and organization in netherton syndrome patients. *Journal of Investigative Dermatology* **134**, 2014, pp. 1238–1245.
- [8] ISHIKAWA, J., NARITA, H., KONDO, N., HOTTA, M., TAKAGI, Y., MASUKAWA, Y., KITAHARA, T., TAKEMA, Y., KOYANO, S., YAMAZAKI, S., AND HATOMICHI, A. Changes in the ceramide profile of atopic dermatitis patients. *The Journal of Investigative Dermatology* **130**, 2010, pp. 2511–2514.
- [9] JOO, K.-M., HWANG, J.-H., BAE, S., NAHM, D.-H., PARK, H.-S., YE, Y.-M., AND LIM, K.-M. Relationship of ceramide-, and free fatty acid-cholesterol ratios in the stratum corneum with skin barrier function of normal, atopic dermatitis lesional and non-lesional skins. *Journal of Dermatological Science* **77**, 2015, pp. 71–74.
- [10] RABIONET, M., GORGAS, K., AND SANDHOFF, R. Ceramide synthesis in the epidermis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 2014, pp. 422–434.

- [11] DE JAGER, M. W., GROENINK, W., GUIVERNAU, R. B. I., ANDERSSON, E., ANGELOVA, N., PONEC, M., AND BOUWSTRA, J. A. A novel in vitro percutaneous penetration model: evaluation of barrier properties with p-aminobenzoic acid and two of its derivatives. *Pharmaceutical Research* **23**, 2006, pp. 951–960.
- [12] UCHIYAMA, M., OGURI, M., MOJUMDAR, E. H., GOORIS, G. S., AND BOUWSTRA, J. A. Free fatty acids chain length distribution affects the permeability of skin lipid model membranes. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 2050–2059.
- [13] GARIDEL, P., FÖLTING, B., SCHALLER, I., AND KERTH, A. The microstructure of the stratum corneum lipid barrier: mid-infrared spectroscopic studies of hydrated ceramide:palmitic acid:cholesterol model systems. *Biophysical Chemistry* **150**, 2010, pp. 144–156.
- [14] LAFLEUR, M. Phase behaviour of model stratum corneum lipid mixtures: an infrared spectroscopy investigation. *Canadian Journal of Chemistry* **76**, 1998, pp. 1501–1511.
- [15] MOJUMDAR, E. H., GOORIS, G. S., AND BOUWSTRA, J. A. Phase behavior of skin lipid mixtures: the effect of cholesterol on lipid organization. *Soft Matter* **11**, 2015, pp. 4326–4336.
- [16] MOJUMDAR, E. H., GOORIS, G. S., GROEN, D., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., , AND BOUWSTRA, J. A. Stratum corneum lipid matrix: location of acyl ceramide and cholesterol in the unit cell of the long periodicity phase. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 1926–1934.
- [17] MOORE, D. J., REREK, M. E., AND MENDELSON, R. Role of ceramides 2 and 5 in the structure of the stratum corneum lipid barrier. *International Journal of Cosmetic Science* **21**, 1999, pp. 353–368.
- [18] MÜLLER, J., SCHRÖTER, A., STEITZ, R., TRAPP, M., AND NEUBERT, R. H. H. Preparation of a new oligolamellar stratum corneum lipid model. *Langmuir* **32**, 2016, pp. 4673–4680.
- [19] ROWAT, A. C., KITSON, N., AND THEWALT, J. L. Interactions of oleic acid and model stratum corneum membranes as seen by  $^2\text{H}$  nmr. *International Journal of Pharmaceutics* **307**, 2006, pp. 225–231.
- [20] ŠKOLOVÁ, B., JANŮŠOVÁ, B., ZBYTOVSKÁ, J., GOORIS, G. S., BOUWSTRA, J. A., SLEPIČKA, P., BERKA, P., ROH, J., PALÁT, K., HRABÁLEK, A., AND VÁVROVÁ, K. Ceramides in the skin lipid membranes: length matters. *Langmuir* **29**, 2013, pp. 15624–15633.
- [21] ŠKOLOVÁ, B., JANŮŠOVÁ, B., AND VÁVROVÁ, K. Ceramides with a pentadecaphingosine chain and short acyls have strong permeabilization effects on skin and model lipid membranes. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 220–232.

- [22] VELKOVA, V. AND LAFLEUR, M. Influence of the lipid composition on the organization of skin lipid model mixtures: an infrared spectroscopy investigation. *Chemistry and Physics of Lipids* **117**, 2002, pp. 63–74.
- [23] GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., VAN MECHELEN, J. B., DEMÉ, B., AND BOUWSTRA, J. A. Disposition of ceramide in model lipid membranes determined by neutron diffraction. *Biophysical Journal* **100**, 2011, pp. 1481–1489.
- [24] MOJUMDAR, E. H., GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., AND BOUWSTRA, J. A. Localization of cholesterol and fatty acid in a model lipid membrane: a neutron diffraction approach. *Biophysical Journal* **105**, 2013, pp. 911–918.
- [25] MOJUMDAR, E. H., KARIMAN, Z., VAN KERCKHOVE, L., GOORIS, G. S., AND BOUWSTRA, J. A. The role of ceramide chain length distribution on the barrier properties of the skin lipid membranes. *Biochimica et Biophysica Acta - Biomembranes* **1838**, 2014, pp. 2473–2483.
- [26] AKINSHINA, A., DAS, C., AND NORO, M. G. Effect of monoglycerides and fatty acids on a ceramide bilayer. *Physical Chemistry Chemical Physics* **18**, 2016, pp. 17446–17460.
- [27] DAS, C., NORO, M. G., AND OLMSTED, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **97**, 2009, pp. 1941–1951.
- [28] DAS, C., NORO, M. G., AND OLMSTED, P. D. Lamellar and inverse micellar structures of skin lipids: Effect of templating. *Physical Review Letters* **111**, 2013, p. 148101.
- [29] DAS, C., NORO, M. G., AND OLMSTED, P. D. Fast cholesterol flip-flop and lack of swelling in skin lipid multilayers. *Soft Matter* **10**, 2014, pp. 7346–7352.
- [30] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>C</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [31] GUPTA, R., DWADASI, B. S., AND RAI, B. Molecular dynamics simulation of skin lipids: effect of ceramide chain lengths on bilayer properties. *The Journal of Physical Chemistry B* **120**, 2016, pp. 12536–12546.
- [32] GUPTA, R. AND RAI, B. Molecular dynamics simulation study of skin lipids: effects of the molar ratio of individual components over a wide temperature range. *The Journal of Physical Chemistry B* **119**, 2015, pp. 11643–11655.
- [33] HOOPES, M. I., NORO, M. G., LONGO, M. L., AND FALLER, R. Bilayer structure and lipid dynamics in a model stratum corneum with oleic acid. *The Journal of Physical Chemistry B* **115**, 2011, pp. 3164–3171.

- [34] IMAI, Y., LIU, X., YAMAGISHI, J., MORI, K., NEYA, S., AND HOSHINO, T. Computational analysis of water residence on ceramide and sphingomyelin bilayer membranes. *Journal of Molecular Graphics and Modelling* **29**, 2010, pp. 461–469.
- [35] NOTMAN, R., ANWAR, J., BRIELS, W. J., NORO, M. G., AND DEN OTTER, W. K. Simulations of skin barrier function: free energies of hydrophobic and hydrophilic transmembrane pores in ceramide bilayers. *Biophysical Journal* **95**, 2008, pp. 4763–4771.
- [36] NOTMAN, R., DEN OTTER, W. K., NORO, M. G., BRIELS, W. J., AND ANWAR, J. The permeability enhancing mechanism of DMSO in ceramide bilayers simulated by molecular dynamics. *Biophysical Journal* **93**, 2007, pp. 2056–2068.
- [37] PALONCÝOVÁ, M., DEVANE, R., MURCH, B. P., BERKA, K., AND OTYEPKA, M. Rationalization of reduced penetration of drugs through ceramide gel phase membrane. *Langmuir* **30**, 2014, pp. 13942–13948.
- [38] PALONCÝOVÁ, M., VÁVROVÁ, K., SOVOVÁ, Ž., DEVANE, R., OTYEPKA, M., AND BERKA, K. Structural changes in ceramide bilayers rationalize increased permeation through stratum corneum models with shorter acyl tails. *The Journal of Physical Chemistry B* **119**, 2015, pp. 9811–9819.
- [39] PANDIT, S. A. AND SCOTT, H. L. Molecular-dynamics simulation of a ceramide bilayer. *The Journal of Chemical Physics* **124**, 2006, p. 014708.
- [40] UPPULURY, K., COPPOCK, P. S., AND KINDT, J. T. Molecular simulation of the DPPE lipid bilayer gel phase: coupling between molecular packing order and tail tilt angle. *The Journal of Physical Chemistry B* **119**, 2015, pp. 8725–8733.
- [41] SPARR, E., ERIKSSON, L., BOUWSTRA, J. A., AND EKELUND, K. Afm study of lipid monolayers: III. phase behavior of ceramides, cholesterol and fatty acids. *Langmuir* **17**, 2001, pp. 164–172.
- [42] ŠKOLOVÁ, B., HUDSKÁ, K., PULLMANOVÁ, P., KOVÁČIK, A., PALÁT, K., ROH, J., FLEDDERMAN, J., ESTRELA-LOPIS, I., AND VÁVROVÁ, K. Different phase behavior and packing of ceramides with long (C16) and very long (C24) acyls in model membranes: infrared spectroscopy using deuterated lipids. *The Journal of Physical Chemistry B* **118**, 2014, pp. 10460–10470.
- [43] OGURI, M., GOORIS, G. S., BITO, K., AND BOUWSTRA, J. A. The effect of the chain length distribution of free fatty acids on the mixing properties of stratum corneum model membranes. *Biochimica et Biophysica Acta - Biomembranes* **1838**, 2014, pp. 1851–1861.
- [44] KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O’CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D., AND PASTOR, R. W. Update of the CHARMM all-atom additive force

- field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 2010, pp. 7830–7843.
- [45] GILLAMS, R. J., BUSTO, J. V., BUSCH, S., GOÑI, F. M., LORENZ, C. D., AND MCLAIN, S. E. Solvation and hydration of the ceramide headgroup in a non-polar solution. *The Journal of Physical Chemistry B* **119**, 2014, pp. 128–139.
- [46] GILLAMS, R. J., LORENZ, C. D., AND MCLAIN, S. E. Comparative atomic-scale hydration of the ceramide and phosphocholine headgroup in solution and bilayer environments. *The Journal of Chemical Physics* **144**, 2016, p. 225101.
- [47] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 1983, pp. 926–935.
- [48] MARINARI, E. AND PARISI, G. Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters* **19**, 1992, pp. 451–458.
- [49] ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B., AND LINDAHL, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 2015, pp. 19–25.
- [50] HOOVER, W. G. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A* **31**, 1985, pp. 1695–1967.
- [51] PARRINELLO, M. AND RAHMAN, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *Journal of Applied Physics* **52**, 1981, pp. 7182–7190.
- [52] ESSMANN, U., PERERA, L., BERKOWITZ, M. L., DARDEN, T., LEE, H., AND PEDERSEN, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 1995, pp. 8577–8593.
- [53] DAS, C., OLMSTED, P. D., AND NORO, M. G. Water permeation through stratum corneum lipid bilayers from atomistic simulations. *Soft Matter* **5**, 2009, pp. 4549–4555.
- [54] HARTKAMP, R., MOORE, T. C., IACOVELLA, C. R., THOMPSON, M. A., BULSARA, P. A., MOORE, D. J., AND M<sup>C</sup>CABE, C. Investigating the structure of multicomponent gel-phase lipid bilayers. *Biophysical Journal* **111**, 2016, pp. 813–823.
- [55] GUNSTONE, F. D., HARWOOD, J. L., AND DIJKSTRA, A. J. *The Lipid Handbook*. CRC Press (2007).
- [56] BOUGHTER, C. T., MONJE-GALVAN, V., IM, W., AND KLAUDA, J. B. Influence of cholesterol on phospholipid bilayer structure and dynamics. *The Journal of Physical Chemistry B* **120**, 2016, pp. 11761–11772.

- [57] SHIEH, H. S., HOARD, L. G., AND NORDMAN, C. E. Crystal-structure of anhydrous cholesterol. *Nature* **267**, 1977, pp. 287–289.
- [58] TIELEMAN, D. P. AND BERENDSEN, H. J. C. Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *The Journal of Chemical Physics* **105**, 1996, pp. 4871–4880.

## CHAPTER 5

### DEVELOPMENT OF COARSE-GRAINED FORCE FIELDS VIA MULTISTATE ITERATIVE BOLTZMANN INVERSION

This chapter is reproduced in part from Moore *et al.* 2014.<sup>1</sup>

#### 5.1 Introduction

The utility of coarse-grained (CG) force fields for use in simulations of soft matter and biological systems has been well-established in the literature, enabling simulation to explore greater length- and timescales than is feasible with fully atomistic models. This is of particular importance when studying the self-assembly of soft matter systems, where the assembly is typically driven by weak forces (*e.g.*, hydrophobicity and entropy)<sup>2-7</sup> and structures often demonstrate hierarchical ordering (*e.g.*, molecules organized into micelles, micelles organized into local/global patterns).<sup>6,8-12</sup> While generic, non-specific CG models have been widely applied,<sup>13-17</sup> providing important information regarding trends and design rules, it is often necessary to use CG models specifically mapped to the system of interest to provide a direct one-to-one correspondence with experiment. While several “transferable” CG force fields, such as TraPPE-CG<sup>18</sup> and MARTINI,<sup>19</sup> haven been developed, akin to force field development at the atomistic level,<sup>20-24</sup> the development of new CG force fields is still often necessary. This is often required since the available force fields may be lacking the necessary molecular species/groupings or may not have been optimized for the properties of interest. This second point is of particular consequence, since, for example, a force field optimized to match phase behavior may not appropriately capture subtle structural features.<sup>18</sup> Generally speaking, direct structural correspondence is needed to accurately transition between different simulation levels (*e.g.*, atomistic to CG), in

order to perform multiscale<sup>25-27</sup> and hybrid-multiscale simulations,<sup>28-34</sup> as well as to recover atomistic details from CG simulations.<sup>35-39</sup>

Several approaches have been developed to derive and optimize CG force fields.<sup>40-45</sup> Among these, the iterative Boltzmann inversion (IBI) method<sup>40</sup> has become a popular choice due to its straightforward nature, general applicability to a wide range of systems, and basis in structural properties. The IBI method relies on self-consistently adjusting a given potential to achieve convergence with target structural data; for nonbonded interactions this target takes the form of the radial distribution function (RDF) between interaction sites and the potential is iteratively updated according to

$$V_{\text{CG}}^{(i+1)}(r) = V_{\text{CG}}^{(i)}(r) - \alpha k_{\text{B}} T \ln \left[ \frac{g^{(\text{target})}(r)}{g^{(i)}(r)} \right], \quad (5.1)$$

where  $V_{\text{CG}}^{(i)}(r)$  is a numerical pair potential;  $i$  represents the current iteration;  $\alpha$  is a damping factor to suppress large changes to the potential, typically varying from 0.2 to unity, where smaller values tend to be necessary to capture dense and/or crystalline states;<sup>46</sup>  $k_{\text{B}}$  is the Boltzmann constant;  $T$  is the absolute temperature;  $r$  is the separation between particles;  $g^{(i)}(r)$  is the pair RDF from the simulation of  $V_{\text{CG}}^{(i)}(r)$ ; and  $g^{(\text{target})}(r)$  is the RDF of the target system mapped to the CG level. Although the CG force fields derived via IBI are typically able to accurately reproduce the target RDFs, they are, in general, only applicable at the state point for which they were derived, due to the structural nature of their derivation (*e.g.*, note the explicit temperature dependence of Equation 5.1, as well as the implicit temperature and density dependence through the  $g(r)$  terms).<sup>47,48</sup> For example, separate potentials were required to capture both the solid and fluid structures of a pure simple lipid.<sup>46</sup> Additionally, several works have shown that CG polymer potentials derived via the IBI method can depend on the chemical environment for which they were derived.<sup>40,49,50</sup> Qian *et al.* found that the potentials derived using IBI for ethylbenzene scale in a

nonlinear fashion with the temperature, however, in the same work, it was found that CG potentials derived for polystyrene were independent of the temperature.<sup>51</sup> Other recent work has shown that some of the CG potentials in a benzene–urea–water system derived via IBI possess a degree of state point transferability, but not all.<sup>52</sup> It is thus unclear why IBI provides transferability for some but not all potentials; this may ultimately depend on a number of factors, including the effective size of the CG beads and the “shape” of the RDFs being fit, but may also be a limitation of the single state methodology. Furthermore, for complex systems, it may not be possible to optimize potentials at the state points of interest, due to time- or length-scale limitations of the atomistic simulations, thus making it difficult to apply the IBI method appropriately given that potentials are not necessarily transferable. Perhaps of most concern is the fact that the IBI method does not guarantee a unique solution, as a multitude of vastly differing potentials may give rise to otherwise matching RDFs. The form of the final derived potential may also vary based on runtime parameters, such as the initial potential guess, potential cutoff, magnitude of the damping factor, etc. . . . Additionally, the derived potential may include artifacts associated with intermediate and long-range structural correlations in the system, *e.g.*, oscillatory behavior in the potential that follows the peaks and valleys in the RDF, which may alter other properties of the system, even if the RDFs match.

In this Chapter, I present the extension of IBI method to perform multistate optimization, *i.e.*, the potential is self-consistently adjusted to achieve simultaneous convergence of target data from multiple states. The general idea, illustrated in Figure 5.1, is that the inclusion of target data from multiple states adds constraints to the optimization problem, such that the derived force field tends toward a single set of pair potentials that can adequately represent all states. For example, potentials in region *i* of Figure 5.1a are able to match the target structure at a single state *i*, potentials in region *ii* are able to reproduce target data at state *ii*, etc. . . , with a

single representative potential lying at the overlap of these regions, shown as region *iv*. To illustrate the efficacy of the proposed multistate method, I first present potential optimizations for the idealized system of a Lennard-Jones (LJ) fluid for which the potential is known, in order to determine if the method resolves the correct potential. Next, to test the method in a system where only nonbonded interactions are present in the CG model, a 3:1 mapped CG force field is optimized for propane using target data generated from united-atom (UA) propane simulations, and compared with a single-site LJ model mapped to the experimental critical point of propane. Next, I apply the approach to *n*-dodecane, a system more representative of the typical application of a CG force field. Finally, a monolayer system of *n*-dodecane is examined, where it is demonstrated that adjustment of the relative weights given to each target in the MS IBI method can be used to tune the potentials to match other measurable system properties beyond the RDF.

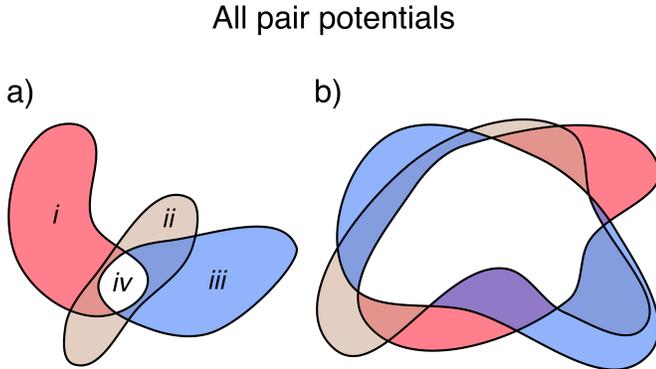


Figure 5.1: Regions of good potential phase space for states with optimal overlap (left) and too much overlap (right).

## 5.2 Method and Simulation Details

### 5.2.1 Single State Iterative Boltzmann Inversion

In the IBI method (which for clarity is referred to as single state, or SS IBI), a numerical pair potential,  $V(r)$ , is iteratively updated according to Eq. 5.1. In this

manner,  $V(r)$  is updated at each separation  $r$  based on whether the RDF overpredicts or underpredicts the target RDF at the given  $r$ , and is repeated until the trial RDF matches the target RDF within some tolerance.<sup>40</sup> The initial guess of the numerical potential is often taken to be the Boltzmann inversion of the target RDF:

$$V_{\text{CG}}^{(0)}(r) = -k_{\text{B}}T \ln g^{(\text{target})}(r). \quad (5.2)$$

While not exact for site–site interactions in molecules,<sup>53</sup> this methodology is motivated by the statistical mechanical relationship between the potential of mean force and the RDF, and provides a reasonable starting potential over which to iterate.

Typically, force fields derived with this method are capable of reproducing the target RDFs with high accuracy, with slight deviations resulting from information lost during coarse-graining. The ease of use of the IBI method and its general applicability make it a powerful tool; given a CG mapping and a target RDF, site–site pair potentials can be readily derived with little user input.

### 5.2.2 Multistate Extension of IBI

Although force fields derived with SS IBI will typically reproduce their target RDFs with high accuracy, caution must be taken when using them. Upon successful convergence of the pair potentials, it is only guaranteed that the CG and target RDFs match, not that the pair potentials are necessarily representative of the “true” underlying potentials (*i.e.*, not necessarily state-independent). It is important to note, especially since information is lost due to coarse-graining, that a multitude of CG potentials may give rise to similar CG RDFs. Only a small portion of the pair potentials that produce matching RDFs may accurately fall within the region of potentials that match the true potential and, since the true potential is typically unknown, it is difficult to assess the accuracy of the derived potentials. If the derived potential falls far outside the true potential region, this may give rise to potentials

that, despite providing a good match for the target RDF, lack transferability and contain artifacts making them incapable of resolving system properties other than the RDF.

The proposed MS IBI method aims to minimize the state dependence of the derived potentials by adding additional constraints to the optimization process such that the derived potentials fall within the region of phase space where potentials are representative of the “true” potential. This approach relies of two key assumptions: (1) different thermodynamic states have different regions of the potential phase space that adequately reproduce their respective target RDFs, and (2) that the true, underlying potential lies within the common overlap between these regions of phase space. As the name suggests, this is accomplished by updating the derived potential to simultaneously match target RDFs at different thermodynamic states, producing a single potential that provides sufficient matching for all target RDFs considered. As shown graphically in Figure 5.1a, the converged potential lies at the intersection of each of the regions representing the target RDFs, as this is the only region where a sufficient match will be found for all states.

The implementation of MS IBI is similar to that of SS IBI, the only additional requirement is more target data. As in SS IBI, the initial potential is assumed to be the Boltzmann inversions of the target RDFs, averaged over the  $N_{\text{states}}$  states used,

$$V_{\text{CG}}^{(0)}(r) = -\frac{1}{N_{\text{states}}} \sum_{\text{states},s} k_{\text{B}}T \ln g_s^{(\text{target})}(r), \quad (5.3)$$

where the subscript  $s$  represents the property at state  $s$ . After a trial CG simulation is run at each state using the potential from Equation (5.3), the potential is updated according to

$$V_{\text{CG}}^{(i+1)}(r) = V_{\text{CG}}^{(i)}(r) - \frac{1}{N_{\text{states}}} \sum_{\text{states},s} \alpha_s(r) k_{\text{B}}T_s \ln \left[ \frac{g_s^{(\text{target})}(r)}{g_s^{(i)}(r)} \right]. \quad (5.4)$$

While in SS IBI,  $\alpha$  represents a damping factor useful for suppressing fluctuations in the potential update, here  $\alpha_s(r)$  also serves as a weighting factor, allowing more or less emphasis to be put on each state. For example, if fitting a potential with three states, where state 1 will ultimately be of most interest, it may make sense to give state 1 a higher  $\alpha$  value; this will be discussed later in Chapter 5.3.4. Additionally, here  $\alpha_s(r)$  is defined as a linear function of the separation  $r$ , with the points  $\alpha_s(0) = \alpha_{\max}$ , and  $\alpha_s(r_{\text{cut}}) = 0$ . When  $V(r_{\text{cut}}) = 0$  is used for the initial potential guess, this ensures that the derived potential smoothly decays to zero at the interaction cutoff  $r_{\text{cut}}$ . Since  $\alpha$  decreases as  $r$  increases, increased emphasis is placed on shorter-range interactions compared to long-range interactions, similar to the radial dependence of the pressure correction formula often used with IBI.<sup>40</sup> This helps to suppress the influence of long-range structural correlations (*e.g.*, the formation of bulk crystals from particles interacting through a short-ranged, truncated potential). For direct comparability in this work, both SS and MS IBI treat the damping factor as a linear function of separation, with a fixed value of 0 at  $r_{\text{cut}}$ . Note that, although bonded interactions may be optimized in a similar manner (*i.e.*, adjusting the potential to match a target distribution), in this work, the assumption is made that bonded and nonbonded interactions are sufficiently independent such that we use analytical bonded potentials, as has been done in previous work.<sup>46,54,55</sup>

The choice of states used in the fitting procedure is naturally important to deriving an accurate potential. To derive a potential most representative of the underlying one, it would not be beneficial to choose states with RDFs that are too similar, as the overlap region would be large, essentially providing minimal additional constraints; this situation is shown in Figure 5.1b. In such a case, there would be no advantage to the multistate fitting. At the other end of the spectrum, there may in fact be no overlap between states, or more specifically, no overlap for a given level of matching (*i.e.*, no overlap without relaxing the tolerance of a RDF similarity test). For some

systems, it may not be possible to define a single potential that accurately reproduces the target structure at all states. This is not a problem unique to coarse-graining, as it applies at all levels of modeling, *e.g.*, classical atomistic potentials may also lack full state-independence given that they do not allow variation in electron density.

### 5.2.3 Simulation Model

In this work, simulations were performed using 3 distinct models: a generic LJ fluid, TraPPE-UA, and CG models derived via IBI. First, simulations of particles interacting through a LJ potential were performed in the canonical ensemble (*i.e.*, fixed number of particles, volume, and temperature), with temperature controlled via the Nosé-Hoover thermostat. These monatomic LJ systems contained 1468 particles initially randomly distributed throughout the box, and were run for  $1 \times 10^6$  steps before target data were collected over  $1 \times 10^5$  steps. A timestep of  $1 \times 10^{-3}$  in reduced time units was used. The interaction parameters used in all LJ simulations were  $\sigma = 1.0$  and  $\epsilon = 1.0$ , with a potential cutoff of  $r_{\text{cut}} = 3\sigma$ . Here, no coarse-graining was applied to the target systems, as these simulations were used simply to test the efficacy of the potential derivation in the ideal case where the true potential is known.

The second model used relies on the united atom (UA) TraPPE-UA force field for simulation.<sup>20</sup> Here, alkanes were simulated in the canonical ensemble. Bulk fluid systems of both propane (1024 molecules) and *n*-dodecane (400 molecules) were simulated at 3 different states, as listed in Chapters 5.3.2 and 5.3.3, and used to generate target RDF data. Although not an all-atom model (as hydrogens are not explicitly modeled), the TraPPE force field was chosen for computational convenience, since, in principle, the target data can come from any source. In all cases, a timestep of 1 fs was used. After an initial equilibration period of 5 ns, data were collected over a 10 ns production run. In addition to the bulk fluid *n*-dodecane simulations, UA simulations were performed of *n*-dodecane gel and fluid monolayers, composed

of 100 *n*-dodecane chains each. These were performed in the same manner as the bulk simulations at 298 K, but with the first head bead of each chain held stationary such that a 2D hexagonally arranged periodic array with a density of 4.10 chains/nm<sup>2</sup> (gel-like) and 3.79 chains/nm<sup>2</sup> (fluid-like) was achieved; these were chosen to match state points commonly used in alkylsilane monolayer simulations and experiments.<sup>56</sup>

The third model used is a CG representation of alkanes. In all cases, a 3-to-1 CG model (*i.e.*, each CG bead represents 3 UA carbon groups) was used to simulate bulk fluid and monolayer systems of alkanes. Pair potentials were derived using the SS and MS IBI methods, using the results of the UA simulations as target data, as discussed in detail in Chapter 5.3.

The bond stretching and angle bending potentials used in the study of dodecane were derived by a Boltzmann inversion of the bonded distributions sampled in the atomistic trajectory mapped to the CG level.<sup>54</sup> Specifically, from a normalized bond length distribution  $p(r)$ , the bond stretching potential is written as

$$V_{\text{bond}}(r) = -k_{\text{B}}T \ln p(r),$$

which, assuming a Gaussian bond length distribution, results in a harmonic potential about the most probable bond length,  $r_{\text{eq}}$ ; note a similar formalism was used for angles, where  $\theta$  is substituted for  $r$ , and the normalization includes a factor of  $\sin(\theta)$ . Since minimal state dependence was found between systems, a single set of bonded parameters was used in all simulations, with  $k/k_{\text{B}} = 15.60 \text{ K}/\text{\AA}$  and  $r_{\text{eq}} = 3.56 \text{ \AA}$  for bonds and  $k/k_{\text{B}} = 0.17 \text{ K}/\text{deg}$  and  $\theta_{\text{eq}} = 174.53^\circ$  for angles.

In all cases, the GPU-enabled HOOMD-Blue<sup>57,58</sup> simulation engine was used to perform the simulations. The high performance of the GPU allows for rapid derivation of potentials. A standard potential optimization using MS IBI required approximately 50 iterations to be well-converged. For the pure LJ systems with 1468 particles, this

convergence took less than 1 h using three Nvidia GTX580 GPUs concurrently. The following fitness function was used to measure how well a trial RDF matched with its target:

$$f_{\text{fit}}(i) = 1 - \frac{\sum_{r=0}^{r_{\text{cut}}} |g^{(i)}(r) - g^{(\text{target})}(r)|}{\sum_{r=0}^{r_{\text{cut}}} [g^{(i)}(r) + g^{(\text{target})}(r)]}. \quad (5.5)$$

An  $f_{\text{fit}}$  value of unity represents a perfect match between the CG and target RDFs. Additionally, in all figures, the following two-point central moving average smoothing function was applied to the derived potential to reduce the noise:

$$V'_n(r) = \frac{1}{3} [V_{n-1}(r) + V_n(r) + V_{n+1}(r)], \quad (5.6)$$

where  $V_n(r)$  is the  $n^{\text{th}}$  element of the numerical potential, and the prime denotes the smoothed value. The application of the smoothing function was not found to significantly influence the behavior or degree of matching.

## 5.3 Results

### 5.3.1 Monatomic Lennard-Jones Fluid

To test the efficacy of the MS IBI method, potentials were derived using RDFs from monatomic LJ spheres as target data, and the results compared to single state potential derivations (*i.e.*, SS IBI). Target data were acquired from the following states: state A, reduced density  $\rho^* = N\sigma^3/V = 0.85$ , reduced temperature  $T^* = k_B T/\varepsilon = 0.5$ ; state B,  $\rho^* = 0.67$ ,  $T^* = 1.5$ ; and state C,  $\rho^* = 0.18$ ,  $T^* = 2.0$ . No coarse-graining was performed since the goal was to test whether the MS IBI method could recover a known potential. In contrast to mapping an atomistic system to the CG level, no information about the system is lost through coarse-graining, ensuring that a single potential is truly applicable to all states and that this potential is known.

While the RDFs match well, as illustrated in Figure 5.2a-c, the potentials derived via SS IBI demonstrate significant state dependence, as shown in Figure 5.2d. For the

more dense states A and B (Figures 5.2a and 5.2b), the SS IBI method was not able to converge to the true potential to the extent that in the most dense system (state A), the converged potential is almost purely repulsive. This result is due to the elevated density of this state, where the structure can be reproduced with a purely repulsive potential.<sup>59</sup> In this case, even though the RDF matches the target well, the overall behavior of the system would be dramatically altered as compared to the target. A similar situation arises in state B where only a weak attraction is required to match the target structure. In state C, however, the low density causes attractive forces to become important, and as such, the attractive portion of the LJ potential is needed to fully reproduce the target data. Thus, the true LJ potential is recovered only for state C. The application of SS IBI to the monatomic LJ system illustrates two points: (1) that potentials derived via SS IBI are state-dependent, and (2) these potentials are not unique, in that both the LJ potential and the vastly differing derived potentials produce matching RDFs.

MS IBI aims to address the aforementioned issues by forcing the potential to sample portions of potential phase space that satisfy all of the constraints, *i.e.*, find a single potential that matches the target structure at multiple states. The results of applying MS IBI to the monatomic LJ fluid are shown in Figures 5.2e-h. The inclusion of target data from multiple states results in closely matching RDFs and a derived potential that accurately reproduces the true LJ potential, as shown in Figure 5.2h. Although this example is simple, as no coarse-graining was performed, it illustrates the ability of MS IBI to recover a known potential and reduce state-dependence of the derived potential.

### 5.3.2 Propane

To further test the MS IBI algorithm, potential optimizations were performed for the pair interactions in a CG model of propane. The chosen 3-to-1 mapping results

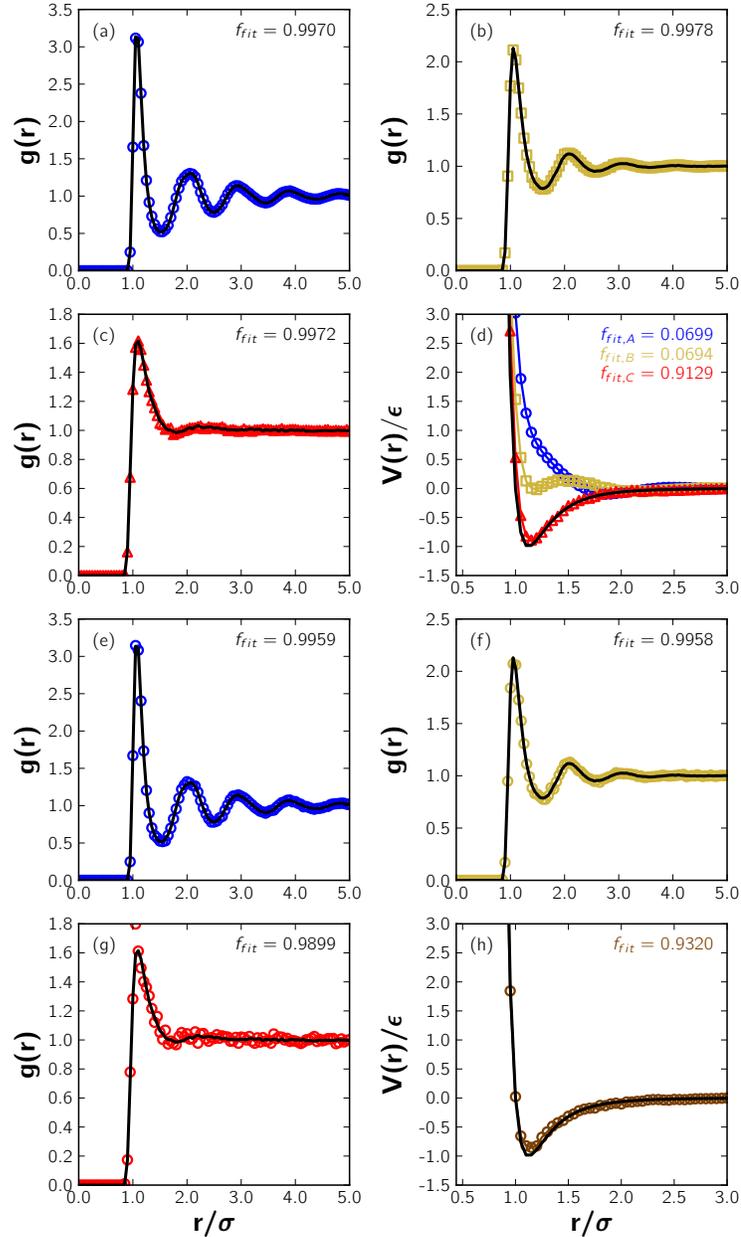


Figure 5.2: RDFs and potentials derived for the LJ system. (a-d) SS IBI results. (e-h) MS IBI results. The  $\alpha$  value used for the MS IBI optimization was 0.7 for each state. The solid black line represents the target RDF (a-c, e-g) or the known potential (d) and (h). Symbols represent the derived RDFs calculated from simulations of the derived potential (a-c, e-g), or the derived potential (d, h).

in a single-site model that can be directly compared to single-site 12-6 LJ models established in the literature.<sup>60</sup> Note, the 12-6 LJ potential should not be considered to be the “true” potential, but rather a good approximation. Target data were acquired

from UA simulations at the following states: state A, 298 K, 0.818 g/mL; state B, 298 K, 0.439 g/mL; and state C, 298 K and 0.14 g/mL. The damping values used in the subsequent MS IBI optimizations were  $\alpha_A(0) = 0.5$ ,  $\alpha_B(0) = 0.7$ , and  $\alpha_C(0) = 0.5$ . The resulting RDFs and (single) pair potential are presented in Figure 5.3. At each state,  $f_{\text{fit}}$  indicates excellent agreement between the target RDFs and those calculated from simulations using the derived potential. Moreover, we find that the derived potential agrees well with a single-site 12-6 LJ model using parameters mapped to the critical point of propane,<sup>60</sup> providing confidence in the MS IBI method. While the match between the two potentials is good, the derived potential does show two small bumps at  $\sim 7 \text{ \AA}$  and  $\sim 9.5 \text{ \AA}$ , which are likely related to orientational effects; that is, likely related to treating the three UA carbon groups as a single, spherically symmetric interaction site.

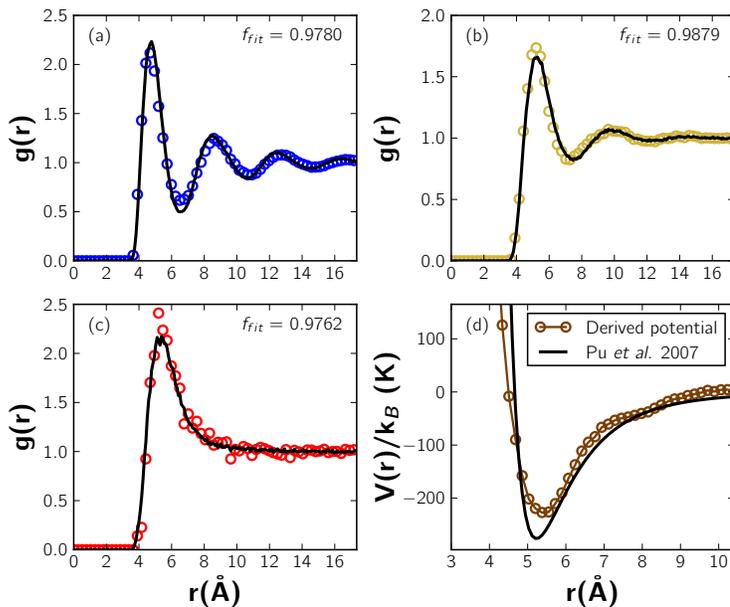


Figure 5.3: RDFs (a-c) and potential (d) derived for the single-site CG model of propane using MS IBI. (a-c) correspond to states A, B, and C in the text, respectively. The  $\alpha$  values were 0.5, 0.7, and 0.5 for states A, B, and C, respectively. The black reference potential in (d) is from fitting a Lennard-Jones potential to have the same critical point as propane.<sup>60</sup>

To illustrate the consistency of the potentials derived via MS IBI (*i.e.*, that the

final potential is insensitive to the initial guess), optimizations were performed using a number of different initial potentials. In addition to the PMF-like quantity of Equation 5.3, three additional initial guesses were used, each a 12-6 LJ potential with vastly differing parameters: (1)  $\varepsilon_1 = 0.46$  kcal/mol,  $\sigma_1 = 4.51$  Å; (2)  $\varepsilon_2 = 0.001\varepsilon_1$ ,  $\sigma_2 = \sigma_1$ ; and (3)  $\varepsilon_3 = 2\varepsilon_1$ ,  $\sigma_3 = \sigma_1$ . The final derived potentials are, in each case, very similar to each other and to the derived potential in Figure 5.3d, as shown Figure 5.4. Particularly, the  $f_{\text{fit}}$  values between each potential and the derived potential shown in Figure 5.3d are 0.986, 0.980, and 0.986, respectively.

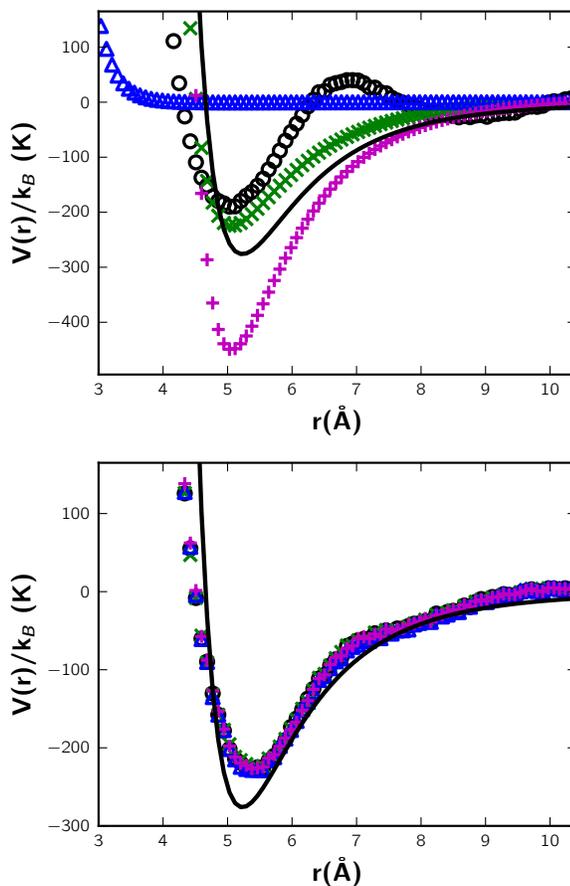


Figure 5.4: Different initial guesses (top) and the resulting derived potentials (bottom) for propane optimizations. As in Figure 5.3d, the solid black reference potential is a Lennard-Jones potential with parameters chosen to have the same critical point as propane.<sup>60</sup>

### 5.3.3 *n*-dodecane

To examine a more complex system and test the state-independence of the derived force field, *n*-dodecane was examined in the bulk state. Intermolecular pair potentials were derived for the beads of a CG model of *n*-dodecane, again using a 3-to-1 mapping. The resulting 4-site model contains two middle beads and two terminal beads, where middle and terminal beads were treated uniquely, resulting in the need to derive three pair potentials; harmonic bonds and angles were used, as detailed in Chapter 5.2. The target data were collected from UA simulations of *n*-dodecane at the following states: state A, 298 K, 1.04 g/mL; state B, 298 K, 0.74 g/mL; and state C, 370 K, 0.55 g/mL. The damping values used were  $\alpha_A(0) = 0.5$ ,  $\alpha_B(0) = 0.7$ , and  $\alpha_C(0) = 0.5$ . Note that state B corresponds to the experimental density at standard ambient temperature and pressure, and, as such, was given higher weight than the other states in this example. Close agreement with the target RDFs was found, with an  $f_{\text{fit}}$  value greater than 0.98 for each of the nine RDFs calculated (3 unique pair interactions at 3 different states).

As a further test of the quality of the derived force field, the average squared radius of gyration normalized by the average end-to-end distance, denoted by  $R_{\text{chain}}$ , was calculated, providing a measure of the chain conformations at different thermodynamic states. Using the MS IBI-derived force field, good agreement was seen between the UA target data and the CG model for the ratio  $R_{\text{chain}}$ , as shown in Figure 5.5. The MS IBI force field deviates slightly from the target data for larger values of  $R_{\text{chain}}$ , which is likely related for the fact that, for simplicity, only a single Gaussian was used to fit the bond and angle distributions. The bond and angle distributions are well-approximated by a single Gaussian at states A and B, however, at the higher temperature state C, the bond and angle distributions may be better approximated by multiple superimposed Gaussians, as proposed by Milano *et al.*<sup>54</sup> While deviations for the MS IBI force field become more apparent for larger values of  $R_{\text{chain}}$ , the SS IBI force fields shows a larger, systematic deviation over all values of  $R_{\text{chain}}$ . Since both the

MS IBI and SS IBI force fields use the same bond and angle parameters, it would be expected that SS IBI would also show the same increased deviations for larger  $R_{chain}$ . However, the ability of MS IBI to more accurately capture the chain conformations of  $n$ -dodecane appears to be directly attributable to the increased accuracy of the derived pair potentials. Note, in both cases, additional simulations were performed at state points not used in the fitting (state points used in the optimizations are highlighted with open squares in Figure 5.5), showing good transferability of the MS IBI-derived force field.

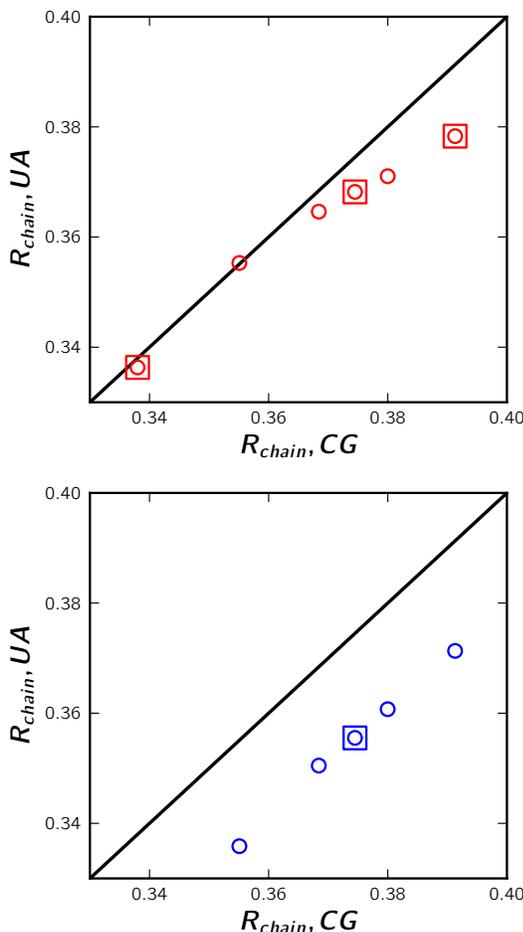


Figure 5.5: Comparison of a structural metric between the CG and UA models of  $n$ -dodecane. The intermolecular pair potentials were derived with MS IBI (top) and SS IBI (bottom). Squares represent data points from simulations at state points used for target data; circles are data points from other state points used for testing the state dependence of the chain conformations.

### 5.3.4 Tuning CG Force Fields

While it is not uncommon to tune pair potentials to reproduce desired behaviors, such tuning is often accomplished by manually adjusting the potentials or mixing potentials derived at different state points, *e.g.*, modifying cross-interactions to ensure stability of CG bilayers.<sup>61</sup> MS IBI provides a more systematic way to optimize the CG force fields by adjusting the relative weights applied to each state in the optimization process, providing a similar effect to manually mixing parameters from optimizations performed independently at different state points. As mentioned in Chapter 5.2, here the damping coefficient  $\alpha_s(r)$  is a function of both separation  $r$  and state  $s$ . Adjusting the  $\alpha_s(0)$  value given to each state  $s$  alters the weight given to the state in the optimization, *i.e.*, more or less emphasis can be placed on matching a given state. While adjusting the relative weights may have only a small effect on the derived RDFs, it may alter subtle features of the derived pair potentials, which can significantly alter other properties. To demonstrate this, as well as to further test the transferability of the derived force field, alkane monolayers were simulated using the 3-to-1 CG model, with pair potentials optimized in the bulk states discussed above, using various values of  $\alpha_s(0)$ . The average tilt angle  $\theta$  (with respect to the surface normal) and nematic order parameter<sup>62</sup>  $S_2$  of the chains were calculated and compared with those values calculated from the corresponding UA simulations.  $S_2$  is defined as  $S_2 = \langle \frac{3}{2} \cos^2 \theta - \frac{1}{2} \rangle$ , where  $\theta$  is the angle between an individual chain in the system and the average direction of all the chains in the system; this yields a value of unity for perfect orientational ordering, decaying to zero as order is lost. Note that the UA monolayer simulations were not used as target data in the force field derivation; they were used only to validate the properties predicted by the derived CG force field.

Unique sets of CG pair potentials were derived over a range of  $\alpha_s(0)$  values, as summarized in Table 5.1. Here, the states A, B, and C are the same states described for *n*-dodecane in Section 5.3.3. Initially, MS IBI was performed with equal weight

given to each state. As shown in Table 5.1, this  $\alpha$  set yields potentials that significantly overpredict the fluid phase order parameter, while underpredicting the chain tilt in the gel-phase monolayer. Since the monolayers are inherently ordered, it would be expected that increasing the relative weight given to the most dense state—state A—would yield a potential that better captures the system behavior. By systematically reducing the weight given to the less dense states, first state C, then state B, a force field that very closely reproduces the monolayer behavior is obtained for  $\alpha_s(0)$  of 0.7, 0.1, and 0.1 for states A, B, and C, respectively. Given the small weights assigned to states B and C, it may be expected that this force field would give results similar to the potentials derived via SS IBI at state A. However, it can be seen in Table 5.1 that this is clearly not the case; the force field derived with target data from only state A shows large deviations from the UA behavior, underpredicting both the average tilt and nematic order in the gel phase monolayer, in stark contrast to the nearly perfect behavior predicted by the MS IBI force field. This result is a direct consequence of using the MS IBI method; even though low weights are given to the other states, the derived potentials will only be considered converged if all states demonstrate good agreement with the target data. That is, state points with lower weights are still able to make changes to the pair potentials, albeit of smaller magnitude; if those changes negatively impact the fitting of higher weighted states, a strong correction will be applied, producing subtle changes to the potentials as compared to systems with equal weight.

While in this case, tilt angle and nematic order demonstrate a strong sensitivity to the weighting factors, not all systems demonstrate such clear dependence. The ability of the CG force field to reproduce the properties of the target model may certainly be linked to the level of coarse-graining applied and therefore deviations of desired properties may be inherent to the CG model, not the choice of weighting functions. Similarly, as previously discussed in Chapter 5.3.3, the fitting of the topological

Table 5.1: Average chain tilt with respect to the surface normal,  $\theta$ , and nematic order parameter,  $S_2$ , for the monolayers in the fluid state (subscript F) and in the gel state (subscript G). The states A, B, and C are the same ones described in Chapter 5.3.3. Values are given as ensemble averages  $\pm$  standard deviation.

$\alpha_A/\alpha_B/\alpha_C$	$\theta_F$	$S_{2,F}$	$\theta_G$	$S_{2,G}$
0.7/0.7/0.7	$19 \pm 6^\circ$	$0.97 \pm 0.01$	$18 \pm 7^\circ$	$0.98 \pm 0.02$
0.7/0.7/0.6	$20 \pm 6^\circ$	$0.97 \pm 0.01$	$20 \pm 4^\circ$	$0.992 \pm 0.002$
0.7/0.7/0.5	$18 \pm 7^\circ$	$0.96 \pm 0.02$	$20 \pm 4^\circ$	$0.991 \pm 0.002$
0.7/0.7/0.4	$15 \pm 7^\circ$	$0.95 \pm 0.02$	$20 \pm 4^\circ$	$0.990 \pm 0.003$
0.7/0.7/0.3	$13 \pm 7^\circ$	$0.94 \pm 0.02$	$19 \pm 4^\circ$	$0.990 \pm 0.004$
0.7/0.7/0.2	$12 \pm 6^\circ$	$0.93 \pm 0.01$	$12 \pm 6^\circ$	$0.96 \pm 0.01$
0.7/0.7/0.1	$10 \pm 6^\circ$	$0.95 \pm 0.01$	$8 \pm 5^\circ$	$0.958 \pm 0.008$
0.7/0.4/0.1	$18 \pm 9^\circ$	$0.83 \pm 0.02$	$29 \pm 4^\circ$	$0.988 \pm 0.003$
0.7/0.1/0.1	$20 \pm 10^\circ$	$0.81 \pm 0.02$	$32 \pm 4^\circ$	$0.989 \pm 0.002$
0.7/-/-	$17 \pm 9^\circ$	$0.85 \pm 0.02$	$15 \pm 7^\circ$	$0.89 \pm 0.01$
-/0.7/-	$23 \pm 5^\circ$	$0.983 \pm 0.005$	$21 \pm 3^\circ$	$0.993 \pm 0.001$
-/-/0.7	$23 \pm 4^\circ$	$0.988 \pm 0.002$	$20 \pm 5^\circ$	$0.987 \pm 0.002$
United atom	$20 \pm 15^\circ$	$0.82 \pm 0.03$	$30 \pm 10^\circ$	$0.97 \pm 0.05$

parameters (*e.g.* bonds, angles, and/or dihedrals) can also influence the measured properties, and thus deficiencies caused by these topological parameters are not likely to be fixed via a simple adjustment of weights. However, adjustment of the weights should provide a means to tune force fields that might otherwise be adjusted manually or would need to be derived by mixing potentials from different state points. The use of MS IBI simply provides a clearer, more systematic framework for this process, even if the weights are modified via trial and error.

## 5.4 Conclusions

A multistate extension of the popular IBI method has been proposed, developed, and validated in this Chapter. In the proposed MS IBI method, multiple thermodynamic states are used in the derivation of single, generally applicable pair potentials. It

was shown that MS IBI is able to recover a known potential where no coarse-graining was performed, in contrast to SS IBI, which was unable to consistently derive the known potential. Through the coarse-graining of propane, it was shown that MS IBI was able to recover a potential very similar to a previously published single-site model, with excellent reproducibility. Furthermore, force fields derived via MS IBI were shown to better reproduce structural conformations of *n*-dodecane than force fields derived via SS IBI. It was also demonstrated that adjusting the relative weights given to each target in the optimizations can be used to tune system properties beyond the RDF; in this case, tuning the weights enabled pair potentials to be derived that provided near perfect agreement between CG and atomistic models when considering the nematic order parameter and tilt angle of an *n*-dodecane monolayer. While pressure, and thermodynamics in general, were not investigated in this work, the standard pressure correction scheme of SS IBI<sup>40</sup> could be trivially applied to MS IBI by calculating the average pressure deviations between all states. As such, the MS IBI stands as an improvement of the typical IBI method, producing more generally applicable force fields that can be tuned to match target properties from experiment or finer-grained simulations.

The improved methodology should be very useful for a host of molecular systems, including, for example, lipid systems, where, not only do systems demonstrate structural heterogeneity within a given state point (*i.e.*, different molecular structures in a single system), but properties such as tilt angle, nematic order, area per lipid etc. . . need to be tuned in order to match atomistic simulation and experiment.<sup>63–67</sup> Given that the MS IBI approach is also capable of deriving force fields which demonstrate increased levels of transferability than SS IBI, force fields can be derived for complex systems with many unique interactions by examining the individual components separately, reducing the number of simultaneous optimizations that need to be performed. Furthermore, this work presents a method to develop potentials that may allow the examination of

phase transitions; in many prior works utilizing SS IBI, different potentials are needed to appropriately model different states, making it difficult to accurately examine the transitions between those states.<sup>19,46,68,69</sup> Additionally, given that multi-GPU machines and GPU-enabled simulation packages are becoming more common,<sup>57,70-72</sup> the potential derivation process can be performed quite efficiently, even if a large number of targets are needed or a large number of iterations must be undertaken to find appropriate weighting functions.

## 5.5 Bibliography

- [1] MOORE, T. C., IACOVELLA, C. R., AND M<sup>c</sup>CABE, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140**, 2014, p. 224104.
- [2] ISRAELACHVILI, J. N., MITCHELL, D. J., AND NINHAM, B. W. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics* **72**, 1976, pp. 1525–1568.
- [3] ISRAELACHVILI, J. N., MITCHELL, D. J., AND NINHAM, B. W. Theory of self-assembly of lipid bilayers and vesicles. *Biochimica et Biophysica Acta - Biomembranes* **470**, 1977, pp. 185–201.
- [4] MUŠEVIČ, I. AND ŠKARABOT, M. Self-assembly of nematic colloids. *Soft Matter* **4**, 2008, pp. 195–199.
- [5] BISHOP, K. J., WILMER, C. E., SOH, S., AND GRZYBOWSKI, B. A. Nanoscale forces and their uses in self-assembly. *Small* **5**, 2009, pp. 1600–1630.
- [6] IACOVELLA, C. R., KEYS, A. S., AND GLOTZER, S. C. Self-assembly of soft-matter quasicrystals and their approximants. *Proceedings of the National Academy of Sciences* **108**, 2011, pp. 20935–20940.
- [7] ZIHERL, P. AND KAMIEN, R. D. Maximizing entropy by minimizing area: Towards a new principle of self-organization. *The Journal of Physical Chemistry B* **105**, 2001, pp. 10147–10158.
- [8] THOMAS, E. L., KINNING, D. J., ALWARD, D. B., AND HENKEE, C. S. Ordered packing arrangements of spherical micelles of diblock copolymers in two and three dimensions. *Macromolecules* **20**, 1987, pp. 2934–2939.
- [9] PHILLIPS, C. L., IACOVELLA, C. R., AND GLOTZER, S. C. Stability of the double gyroid phase to nanoparticle polydispersity in polymer-tethered nanosphere systems. *Soft Matter* **6**, 2010, pp. 1693–1703.
- [10] IACOVELLA, C. R. AND GLOTZER, S. C. Complex crystal structures formed by the self-assembly of ditethered nanospheres. *Nano Letters* **9**, 2009, pp. 1206–1211.
- [11] GLASER, M. A., GRASON, G. M., KAMIEN, R. D., KOŠMRLJ, A., SANTANGELO, C. D., AND ZIHERL, P. Soft spheres make more mesophases. *Europhysics Letters* **78**, 2007, p. 46004.
- [12] ZIHERL, P. AND KAMIEN, R. D. Soap froths and crystal structures. *Physical Review Letters* **85**, 2000, p. 3528.
- [13] LARSON, R., SCRIVEN, L., AND DAVIS, H. Monte carlo simulation of model amphiphile-oil-water systems. *The Journal of Chemical Physics* **83**, 1985, pp. 2411–2420.

- [14] GLOTZER, S. C., HORSCH, M. A., IACOVELLA, C. R., ZHANG, Z., CHAN, E. R., AND ZHANG, X. Self-assembly of anisotropic tethered nanoparticle shape amphiphiles. *Current Opinion in Colloid and Interface Science* **10**, 2005, pp. 287–295.
- [15] GAI, L., MAERZKE, K. A., CUMMINGS, P., AND M<sup>c</sup>CABE, C. A Wang–Landau study of a lattice model for lipid bilayer self-assembly. *The Journal of Chemical Physics* **137**, 2012, p. 144901.
- [16] IACOVELLA, C. R., HORSCH, M. A., ZHANG, Z., AND GLOTZER, S. C. Phase diagrams of self-assembled mono-tethered nanospheres from molecular simulation and comparison to surfactants. *Langmuir* **21**, 2005, pp. 9488–9494.
- [17] LAU, K. F. AND DILL, K. A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* **22**, 1989, pp. 3986–3997.
- [18] MAERZKE, K. A. AND SIEPMANN, J. I. Transferable potentials for phase equilibria- coarse-grain description for linear alkanes. *The Journal of Physical Chemistry B* **115**, 2011, pp. 3452–3465.
- [19] MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P., AND DE VRIES, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 2007, pp. 7812–7824.
- [20] MARTIN, M. G. AND SIEPMANN, J. I. Transferable potentials for phase equilibria. 1. united-atom description of n-alkanes. *The Journal of Physical Chemistry B* **102**, 1998, pp. 2569–2577.
- [21] JORGENSEN, W. L., MAXWELL, D. S., AND TIRADO-RIVES, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society* **118**, 1996, pp. 11225–11236.
- [22] OOSTENBRINK, C., VILLA, A., MARK, A. E., AND VAN GUNSTEREN, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 2004, pp. 1656–1676.
- [23] VANOMMESLAEGHE, K., HATCHER, E., ACHARYA, C., KUNDU, S., ZHONG, S., SHIM, J., DARIAN, E., GUVENCH, O., LOPES, P., VOROBYOV, I., AND MACKERELL, A. D. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 2010, pp. 671–690.
- [24] WANG, J., WOLF, R. M., CALDWELL, J. W., KOLLMAN, P. A., AND CASE, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **25**, 2004, pp. 1157–1174.

- [25] M<sup>C</sup>CABE, C., GLOTZER, S. C., KIEFFER, J., NEUROCK, M., AND CUMMINGS, P. T. Multiscale simulation of the synthesis, assembly and properties of nanostructured organic/inorganic hybrid materials. *Journal of Computational and Theoretical Nanoscience* **1**, 2004, pp. 265–279.
- [26] LIU, H., LI, M., LU, Z.-Y., ZHANG, Z.-G., SUN, C.-C., AND CUI, T. Multiscale simulation study on the curing reaction and the network structure in a typical epoxy system. *Macromolecules* **44**, 2011, pp. 8650–8660.
- [27] PETER, C. AND KREMER, K. Multiscale simulation of soft matter systems—from the atomistic to the coarse-grained level and back. *Soft Matter* **5**, 2009, pp. 4357–4366.
- [28] DI PASQUALE, N., MARCHISIO, D., AND CARBONE, P. Mixing atoms and coarse-grained beads in modelling polymer melts. *The Journal of Chemical Physics* **137**, 2012, p. 164111.
- [29] ENSING, B., NIELSEN, S. O., MOORE, P. B., KLEIN, M. L., AND PARRINELLO, M. Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics. *Journal of Chemical Theory and Computation* **3**, 2007, pp. 1100–1105.
- [30] LIDORIKIS, E., BACHLECHNER, M. E., KALIA, R. K., NAKANO, A., VASHISHTA, P., AND VOYIADJIS, G. Z. Coupling length scales for multiscale atomistics-continuum simulations: atomistically induced stress distributions in si/si 3 n 4 nanopixels. *Physical Review Letters* **87**, 2001, p. 086104.
- [31] MICHEL, J., ORSI, M., AND ESSEX, J. W. Prediction of partition coefficients by multiscale hybrid atomic-level/coarse-grain simulations. *The Journal of Physical Chemistry B* **112**, 2008, pp. 657–660.
- [32] PRAPROTNIK, M., DELLE SITE, L., AND KREMER, K. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annual Review of Physical Chemistry* **59**, 2008, pp. 545–571.
- [33] RZEPIELA, A. J., LOUHIVUORI, M., PETER, C., AND MARRINK, S. J. Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Physical Chemistry Chemical Physics* **13**, 2011, pp. 10437–10448.
- [34] WERDER, T., WALTHER, J. H., AND KOUMOUTSAKOS, P. Hybrid atomistic-continuum method for the simulation of dense fluid flows. *Journal of Computational Physics* **205**, 2005, pp. 373–390.
- [35] HEATH, A. P., KAVRAKI, L. E., AND CLEMENTI, C. From coarse-grain to all-atom: Toward multiscale analysis of protein landscapes. *Proteins: Structure, Function, and Bioinformatics* **68**, 2007, pp. 646–661.
- [36] HESS, B., HOLM, C., AND VAN DER VEGT, N. Osmotic coefficients of atomistic nacl (aq) force fields. *The Journal of Chemical Physics* **124**, 2006, p. 164509.

- [37] LIU, P., SHI, Q., LYMAN, E., AND VOTH, G. A. Reconstructing atomistic detail for coarse-grained models with resolution exchange. *The Journal of Chemical Physics* **129**, 2008, p. 114103.
- [38] RZEPIELA, A. J., SCHÄFER, L. V., GOGA, N., RISSELADA, H. J., DE VRIES, A. H., AND MARRINK, S. J. Reconstruction of atomistic details from coarse-grained structures. *Journal of Computational Chemistry* **31**, 2010, pp. 1333–1343.
- [39] LYMAN, E., YTREBERG, F. M., AND ZUCKERMAN, D. M. Resolution exchange simulation. *Physical Review Letters* **96**, 2006, p. 028105.
- [40] REITH, D., PÜTZ, M., AND MÜLLER-PLATHE, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24**, 2003, pp. 1624–1636.
- [41] ERCOLESSI, F. AND ADAMS, J. B. Interatomic potentials from first-principles calculations: the force-matching method. *Europhysics Letters* **26**, 1994, p. 583.
- [42] IZVEKOV, S. AND VOTH, G. A. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B* **109**, 2005, pp. 2469–2473.
- [43] SHELL, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129**, 2008, p. 144108.
- [44] CHAIMOVICH, A. AND SHELL, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *The Journal of Chemical Physics* **134**, 2011, p. 094112.
- [45] IACOVELLA, C. R., ROGERS, R. E., GLOTZER, S. C., AND SOLOMON, M. J. Pair interaction potentials of colloids by extrapolation of confocal microscopy measurements of collective suspension structure. *The Journal of Chemical Physics* **133**, 2010, p. 164903.
- [46] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A coarse-grained model for amorphous and crystalline fatty acids. *The Journal of Chemical Physics* **132**, 2010, p. 134505.
- [47] FALLER, R. Automatic coarse graining of polymers. *Polymer* **45**, 2004, pp. 3869–3876.
- [48] FU, C.-C., KULKARNI, P. M., SCOTT SHELL, M., AND GARY LEAL, L. A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties. *The Journal of Chemical Physics* **137**, 2012, p. 164106.
- [49] BAYRAMOGLU, B. AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 2012, pp. 9205–9219.

- [50] BAYRAMOGLU, B. AND FALLER, R. Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion. *Macromolecules* **46**, 2013, pp. 7957–7976.
- [51] QIAN, H.-J., CARBONE, P., CHEN, X., KARIMI-VARZANEH, H. A., LIEW, C. C., AND MÜLLER-PLATHE, F. Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures. *Macromolecules* **41**, 2008, pp. 9919–9929.
- [52] GANGULY, P. AND VAN DER VEGT, N. F. A. Representability and transferability of Kirkwood–Buff iterative Boltzmann inversion models for multicomponent aqueous systems. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5247–5256.
- [53] CHAN, E. R., STRIOLO, A., M<sup>c</sup>CABE, C., CUMMINGS, P. T., AND GLOTZER, S. C. Coarse-grained force field for simulating polymer-tethered silsesquioxane self-assembly in solution. *The Journal of Chemical Physics* **127**, 2007, p. 114102.
- [54] MILANO, G., GOUDEAU, S., AND MÜLLER-PLATHE, F. Multicentered Gaussian-based potentials for coarse-grained polymer simulations: linking atomistic and mesoscopic scales. *Journal of Polymer Science Part B: Polymer Physics* **43**, 2005, pp. 871–885.
- [55] HADLEY, K. R. AND M<sup>c</sup>CABE, C. A structurally relevant coarse-grained model for cholesterol. *Biophysical Journal* **99**, 2010, pp. 2896–2905.
- [56] RIVERA, J. L., JENNINGS, G. K., AND M<sup>c</sup>CABE, C. Examining the frictional forces between mixed hydrophobic–hydrophilic alkylsilane monolayers. *The Journal of Chemical Physics* **136**, 2012, p. 244701.
- [57] ANDERSON, J. A., LORENZ, C. D., AND TRAVESSET, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 2008, pp. 5342–5359.
- [58] GLASER, J., NGUYEN, T. D., ANDERSON, J. A., LUI, P., SPIGA, F., MILLAN, J. A., MORSE, D. C., AND GLOTZER, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **192**, 2015, pp. 97–107.
- [59] WEEKS, J. D., CHANDLER, D., AND ANDERSEN, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics* **54**, 1971, pp. 5237–5247.
- [60] PU, Q., LENG, Y., ZHAO, X., AND CUMMINGS, P. T. Molecular simulations of stretching gold nanowires in solvents. *Nanotechnology* **18**, 2007, p. 424007.
- [61] HADLEY, K. R. AND M<sup>c</sup>CABE, C. A simulation study of the self-assembly of coarse-grained skin lipids. *Soft Matter* **8**, 2012, pp. 4802–4814.

- [62] KEYS, A. S., IACOVELLA, C. R., AND GLOTZER, S. C. Characterizing complex particle morphologies through shape matching: Descriptors, applications, and algorithms. *Journal of Computational Physics* **230**, 2011, pp. 6438–6463.
- [63] HELLER, H., SCHAEFER, M., AND SCHULTEN, K. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *Journal of Physical Chemistry* **97**, 1993, pp. 8343–8343.
- [64] EGBERTS, E., MARRINK, S. J., AND BERENDSEN, H. J. C. Molecular dynamics simulation of a phospholipid membrane. *European Biophysics Journal* **22**, 1994, pp. 423–436.
- [65] SAIZ, L. AND KLEIN, M. L. Computer simulation studies of model biological membranes. *Accounts of Chemical Research* **35**, 2002, pp. 482–489.
- [66] SCOTT, H. L. Modeling the lipid component of membranes. *Current Opinion in Structural Biology* **12**, 2002, pp. 495–502.
- [67] VENTUROLI, M., SPEROTTO, M. M., KRANENBURG, M., AND SMIT, B. Mesoscopic models of biological membranes. *Physics Reports* **437**, 2006, pp. 1–54.
- [68] VETTOREL, T. AND MEYER, H. Coarse graining of short polyethylene chains for studying polymer crystallization. *Journal of Chemical Theory and Computation* **2**, 2006, pp. 616–629.
- [69] D’ADAMO, G., PELISSETTO, A., AND PIERLEONI, C. Coarse-graining strategies in polymer solutions. *Soft Matter* **8**, 2012, pp. 5151–5167.
- [70] HESS, B., KUTZNER, C., VAN DER SPOEL, D., AND LINDAHL, E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **4**, 2008, pp. 435–447.
- [71] BROWN, W. M., WANG, P., PLIMPTON, S. J., AND THARRINGTON, A. N. Implementing molecular dynamics on hybrid high performance computers—short range forces. *Computer Physics Communications* **182**, 2011, pp. 898–911.
- [72] BROWN, W. M., KOHLMAYER, A., PLIMPTON, S. J., AND THARRINGTON, A. N. Implementing molecular dynamics on hybrid high performance computers—particle–particle particle–mesh. *Computer Physics Communications* **183**, 2012, pp. 449–459.

## CHAPTER 6

### DEVELOPMENT OF A COARSE-GRAINED WATER FORCE FIELD VIA MULTISTATE ITERATIVE BOLTZMANN INVERSION

This Chapter is reproduced in part from Moore *et al.* 2016.<sup>1</sup>

#### 6.1 Introduction

Coarse-grained (CG) models have proven useful in many fields of chemical research,<sup>2-11</sup> allowing molecular simulations to be performed on larger system sizes and longer timescales than is possible with atomistic-level models, enabling complex phenomena such as hierarchical self-assembly to be described.<sup>12,13</sup> In CG simulations of aqueous systems, especially ones with significant amounts of hydrophobic and/or hydrophilic interactions, the water model is important and can have a major impact on the resulting properties of the system.<sup>14</sup>

While the assignment of atoms to CG beads (*i.e.*, defining the CG mapping) is relatively straightforward for most chemical systems, (*e.g.*, aggregating four methyl groups bonded in sequence into a single CG bead), mapping an atomistic water trajectory to the CG level (*i.e.*, grouping several water molecules into a single CG bead) is not as well-defined given the lack of permanent bonds between water molecules. Even if a mapping were chosen, water molecules will diffuse away their initial clusters over time, such that the initial mapping is no longer representative of the local clustering of water. This ambiguity presents a problem for structure-based methods that require an atomistic configuration to be mapped to the corresponding CG configuration, *e.g.*, to generate a target radial distribution function (RDF) against which the force field is optimized. As such, the majority of many-to-one CG models of water (*i.e.*, where one CG bead represents multiple water molecules) have instead been derived by assuming

a functional form of the force field and optimizing the associated parameters to match selected physical properties of water, such as density, vaporization enthalpy, and surface tension.<sup>14-20</sup> For example, Chiu *et al.* developed a 4:1 CG water force field by optimizing the parameters of a Morse potential to accurately reproduce the surface tension and density of liquid water.<sup>19</sup> Despite capturing the interfacial properties and density, this potential overestimates structural correlations, as one might expect given that structural data was not used in its optimization.

Recently, Hadley and McCabe<sup>21</sup> proposed a method for mapping configurations of atomistic water to their CG representations using the  $k$ -means clustering algorithm. Subsequently, in related work, van Hoof *et al.* developed the CUMULUS method for mapping atoms to CG beads.<sup>22</sup> Both methods enable dynamic mapping of multiple water molecules to a single CG bead, allowing structure-based schemes to be used. Here, dynamic refers to a CG mapping that changes over the course of the atomistic trajectory, *i.e.*, different water molecules are assigned to different CG beads in each frame of the atomistic trajectory. Both works employed the iterative Boltzmann inversion (IBI)<sup>23</sup> method to derive the intermolecular interaction by optimizing a numerical, rather than analytical, potential to reproduce RDFs calculated from the atomistic-to-CG mapped configurations.<sup>21,22</sup> The force fields derived are similar and show good agreement with the structural properties and density of the atomistic water models studied. However, neither model is able to accurately reproduce interfacial properties, since they were derived solely from bulk fluid data. This failure to capture interfacial properties is a consequence of the single-state nature of the IBI approach and may alter the balance of hydrophobic and hydrophilic interactions when using these water models in multicomponent systems.

Recently, the multistate IBI (MS IBI) method<sup>24</sup> was developed as an extension of the original IBI approach, with the goal of reducing state dependence and structural artifacts often found in IBI-based force fields.<sup>14,25,26</sup> While IBI-based force fields have

been derived that show some degree of transferability,<sup>26-28</sup> a significant issue related to the IBI method is that a multitude of potentials can give rise to similar RDFs, and the method cannot necessarily differentiate which of the many potentials is most accurate, as only RDF matching is considered. MS IBI operates based on the idea that different thermodynamic states will occupy different regions of potential “phase space” (*i.e.*, regions where potentials give rise to similar RDFs), and that the most transferable, and thus most accurate, potential lies in the overlap of phase space for the different states. That is, by optimizing a potential simultaneously against target data from multiple thermodynamic state points, MS IBI provides constraints to the optimization, forcing the method to derive potentials that exist in this overlap region, and thus are transferable among the states considered. The MS IBI approach has been shown to reduce the state dependence and improve the quality of the derived potentials, as compared to the original IBI method.<sup>24</sup>

In this Chapter, MS IBI is used to derive an intermolecular potential that captures both bulk and interfacial properties of water, improving upon the CG water model of Hadley and McCabe.<sup>21</sup> Again, optimizations are carried out using the MS IBI method, where both bulk and interfacial systems are used simultaneously as target conditions for the optimization. MS IBI is also used, for the first time, in a multi-ensemble context, enabling optimizations in both the canonical (NVT) and isothermal-isobaric (NPT) ensembles to be performed simultaneously to derive the density-pressure relationship of the system. To further constrain the optimization, a slightly modified version of the Chiu *et al.* CG water force field, originally optimized for surface tension, is used as a starting condition, allowing the MS IBI method to make specific modifications to the potential to improve the structural properties. The remainder of this chapter is organized as follows: in Section 6.2, a brief overview of the  $k$ -means clustering and MS IBI algorithms is given and the models used are described. The potential derivation is then presented, validated, and compared to existing CG water models in Section 6.3

and finally, conclusions are drawn about the applicability of the derived CG model and the broader applicability of the MS IBI method discussed in Section 6.4.

## 6.2 Methods

### 6.2.1 $k$ -means Clustering Algorithm

Mapping a water trajectory to a many-to-one CG level is inherently different than mapping a larger molecule’s trajectory, since for water, atoms mapped to a single CG bead necessarily exist on different molecules. Furthermore, the water molecules mapped to a common bead are not likely to remain associated throughout the full simulation because of thermal diffusion. A dynamic mapping scheme is therefore required to generate CG structures from atomistic configurations for water. Following the work of Hadley and McCabe,<sup>21</sup> the  $k$ -means algorithm has been used to map atomistic water trajectories to the CG level.  $k$ -means is a clustering algorithm that is used to find clusters of data points in a large data set. The algorithm, depicted schematically in Figure 6.1 is as follows: first,  $k$  cluster locations are chosen within the data space (in this case, the simulation box from the atomistic simulation). Here, the initial cluster locations were chosen as randomly selected oxygen molecules in the atomistic configuration. Next, each data point (*i.e.*, water molecule) is assigned to its nearest cluster location. The cluster locations are then shifted to the center of mass of the water molecules assigned to each one. This process of assigning water molecules and shifting the cluster locations is repeated until some criterion is met, which in this work was when the largest cluster displacement was less than 0.1 Å. While the  $k$ -means algorithm can be used to group together any number of water molecules, a 4:1 mapping is chosen, as this was found in prior work to provide the best balance between accuracy and computational efficiency,<sup>21</sup> and 4:1 models are common in the literature.<sup>18,19,21</sup> The final cluster locations are used as the location of the CG water beads for calculating the target RDFs, and this process is repeated for each frame in

the atomistic trajectory.

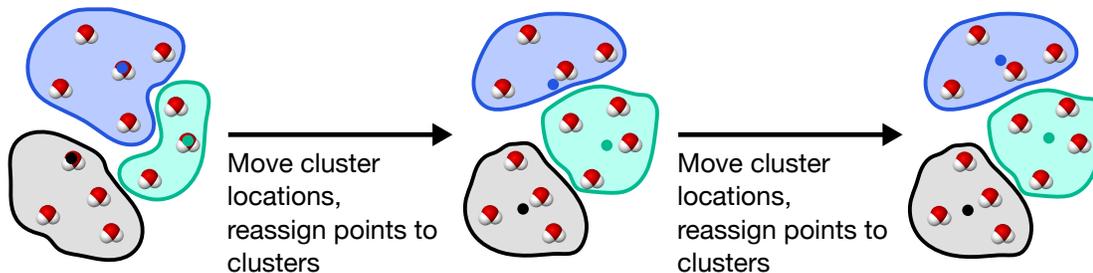


Figure 6.1: A schematic of the  $k$ -means algorithm. The water molecules are shown as red (oxygen) and white (hydrogen) spheres. The cluster locations are shown as solid dots, and the water molecules assigned to each cluster are contained within the colored regions.

## 6.2.2 Multistate Iterative Boltzmann Inversion Method

MS IBI, as detailed in Chapter 5, was used to derive the pair potential between the beads of the CG water model. As in Chapter 5, the weighting factor  $\alpha_s(r)$  was chosen to be a linear function of the form

$$\alpha_s(r) = \alpha_{0,s} \left( 1 - \frac{r}{r_{\text{cut}}} \right),$$

such that  $\alpha_s(r_{\text{cut}}) = 0$  and the pair potential remains 0 for  $r \geq r_{\text{cut}}$ . This form of  $\alpha_s(r)$  also places more emphasis on the short-range part of the potential to suppress long-range structural artifacts.

An initial potential was assumed for the pair interaction. In theory, there are no restrictions on the initial pair potential, so it may take any form; in practice, however, the initial potential is often taken to be the approximate potential of mean force (PMF) between two beads, calculated from a Boltzmann inversion of the RDF. In this Chapter, rather than taking an average of the PMFs over the states used, the initial potential used was chosen to be a slightly modified version of Chiu *et al.*'s water

model,<sup>19</sup> as discussed below. That is, rather than starting from an initial potential that is likely to do a poor job of predicting the behavior, we start from a robust starting point as the Chiu *et al.* model is known to accurately reproduce several properties of water.<sup>19</sup>

As in Chapter 5, the fitness function given in Equation 5.5 was used to define the stopping criteria for the optimization. The specific stopping criteria are given below in Chapter 6.2.3.

### 6.2.3 Models

Atomistic simulations of pure water were performed with the TIP3P model.<sup>29</sup> All atomistic systems contained 5,832 water molecules and were simulated in LAMMPS<sup>30</sup> using a 1 fs timestep. A cutoff distance of 12 Å was used for the van der Waals interactions; long-range electrostatics were handled with the PPPM method with a 12 Å real space cutoff. Three distinct states were simulated: bulk, NVT at 1.0 g/mL and 305 K; bulk, NPT at 305 K and 1.0 atm; and an NVT droplet state at 305 K, where the box from the bulk NVT state was expanded by a factor of 3 in one direction. Each atomistic simulation was run for 7 ns. The atomistic trajectories were mapped to the CG level using the *k*-means algorithm as discussed above. Target RDFs were calculated from the final 5 ns of the mapped trajectory from each state (bulk NVT, bulk NPT, and droplet NVT). MS IBI was performed using the target data from each of the three states. The initial guess of the potential was a Morse potential of the form

$$V(r) = D_e [e^{-2\beta(r-r_{\text{eq}})} - 2e^{-\beta(r-r_{\text{eq}})}], \quad (6.1)$$

where  $r_{\text{eq}}$  is the location of the potential minimum,  $-D_e$  is the value of the potential minimum, and  $\beta$  determines the width of the potential well. Parameters are taken from Chiu *et al.*:  $D_e = 0.813 \text{ kcal/mol}$ ,  $\beta = 0.556/\text{Å}$ , and  $r_{\text{eq}} = 6.29 \text{ Å}$ . However, we note that the potential was adjusted so that  $\beta = 0.5/\text{Å}$  for  $r < r_{\text{eq}}$ . This change

was made to increase sampling at small separations, because numerical issues arise in the potential update when the CG RDF is zero but the target RDF is nonzero. This modification of the potential will slightly alter the properties as compared to the original model, as discussed below. The potential update scaling factor  $\alpha_{0,s}$  was set to 0.7 to avoid large updates to the potential. The optimizations were stopped when  $f_{\text{fit}} \geq 0.98$  and  $f_{\text{fit}}(i) - f_{\text{fit}}(i-1) < 0.001$  for each state.

All optimizations were performed with the open-source MS IBI Python package we developed,<sup>31</sup> which calls HOOMD-Blue<sup>32,33</sup> to run the CG simulations and uses MDTraj<sup>34</sup> for RDF calculations and file-handling. CG simulations were run at the same states as the atomistic systems. Initial CG configurations were generated from the CG-mapped atomistic trajectories at each state. As a result of the 4:1 mapping, CG water simulations contained 1,458 water beads. All CG simulations were run with a 10 fs timestep. The derived CG potential was set to 0 beyond the cutoff of 12 Å.

The surface tension  $\gamma$  of the droplet state was calculated as

$$\gamma = \frac{1}{2}L_z \left\langle P_{zz} - \frac{P_{xx} + P_{yy}}{2} \right\rangle, \quad (6.2)$$

where  $L_z$  is the length of the box in the expanded direction,  $P_{zz}$  is the pressure component in the direction normal to the liquid-vapor interfaces,  $P_{xx}$  and  $P_{yy}$  are the pressure components in the directions lateral to the interfaces, and the angle brackets denote a time average. The factor of  $\frac{1}{2}$  is included to account for the two interfaces that are present in the droplet simulation setup.

## 6.3 Results and Discussion

### 6.3.1 Modified Chiu Potential

Since the MS IBI optimization of water uses a modified version of the Chiu, et al. potential as an initial guess, we first consider the impact of modifying the potential

to create a softer repulsion. Figure 6.2 shows the RDFs of the three target states for the original and modified potentials and the RDF of the 4:1 mapped state (i.e., the target data used later for the MS IBI optimization). The peak location of the NVT state is relatively unchanged; however, upon modification, there is a slight shift in the first peak for the NPT and interfacial states, allowing the model to access smaller separations, as was intended and required for the potential update scheme. The softer potential allows closer contact and thus allows the MS IBI algorithm to modify this region of the potential where the 4:1 mapped atomistic water has non-zero values of the RDF. The density predicted with both potentials is the same ( $0.991 \pm 0.003$  g/mL); however, due to the softening of the potential, the calculated surface tension of the droplet changes from 70.3 mN/m to 45 mN/m after the modification, although this value is still sufficient for the droplet to maintain a stable interface. These surface tension values agree favorably with that of TIP3P water, which is reported to have a surface tension of 52.3 mN/m at 300 K.<sup>35</sup>

### 6.3.2 Potential Derivation and Validation

Starting from the modified Morse potential of Chiu *et al.*, the new water force field is optimized using the bulk NVT and NPT states and the interfacial state. This potential is chosen as the initial starting guess, rather than an arbitrary starting point, as the unmodified version has been shown to accurately reproduce many properties of water (e.g., the density and surface tension), but overestimates the structural correlations. The use of MS IBI should allow for modification of this potential, such that it is able to reproduce structural quantities. The results of the potential derivation are summarized in Figure 6.3, where it is clear that the modified Chiu *et al.* potential (i.e., step 0) overestimates the structural correlations, as was also seen in Figure 6.2 for both the modified and original potentials. After only a few iterations the RDFs match the targets with a high degree of accuracy. This trend is shown in Figure 6.4, which

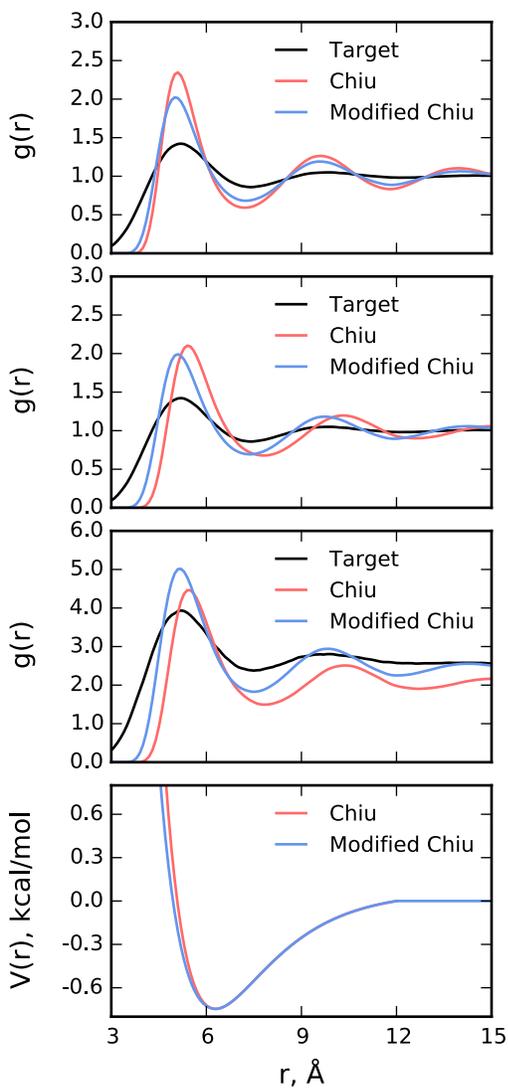


Figure 6.2: RDFs from simulations using the original and modified Chiu potentials. Top: NVT; top-middle: NPT; bottom-middle: interface; bottom: comparison of the two potentials.

plots the fitness value from Eq. 5.5 as a function of iteration. The value of  $f_{\text{fit}}$  changes most rapidly in the first 3 steps of the optimization. After 10 iterations, the stopping criterion is met and the optimization stopped. While changes to the potential are small, there is a noticeable shift in the location of the minimum to a slightly larger  $r$  values and the potential becomes slightly more attractive. Although the shape of the attractive well is mostly unchanged, the potential more rapidly decays to 0 than the

original Morse potential at larger  $r$  values, while the shape of the repulsive regime at small  $r$  is changed slightly. These subtle changes to the potential are sufficient to create significant changes in the RDF and provide excellent convergence of the structural correlations. These changes are made possible by modifying a numerical rather than adjusting parameters for an analytical potential. Note that in Figures 6.2 and 6.3 the RDFs from the interfacial state do not decay to 1 at large  $r$ . This is due to the fact that 2/3 of the box is essentially devoid of particles, but the RDF is normalized based on the volume of the whole simulation box. This has no effect on the potential update scheme, as both the target and CG RDFs are normalized by the same factor, which cancels out in Equation 5.4.

In addition to accurately capturing the RDFs, the multi-ensemble approach provides an accurate estimate of the density at 305 K and 1 atm. NPT simulations performed using the optimized CG force field find a density of  $1.027 \pm 0.006$  g/mL, compared to  $1.037 \pm 0.004$  g/mL for TIP3P water, which was used to generate the target data. This approach is successful because the RDFs will not match if the pressure-density relationship is not satisfied, as the density is implicitly represented in Equation 5.4 through the RDF terms (i.e., the RDFs at the NPT state will not match the target RDFs if the density is significantly different than the density of the target state). In contrast, the original IBI method proposed the use of a pressure correction term of the form  $\Delta V(r) = A(1 - r/r_{cut})$  to account for the pressure.<sup>23</sup> This approach has been successful, but requires a somewhat arbitrary estimate of the parameter  $A$ . While a method exists for estimating  $A$  based on the virial expression,<sup>36</sup> some degree of trial-and-error is still necessary. Furthermore, the multi-ensemble approach within MS IBI does not require direct calculation of the pressure, which often demonstrates considerable fluctuations, providing a simpler route to account for pressure in the CG model.

Calculation of the surface tension of the derived MS IBI potential yields a value

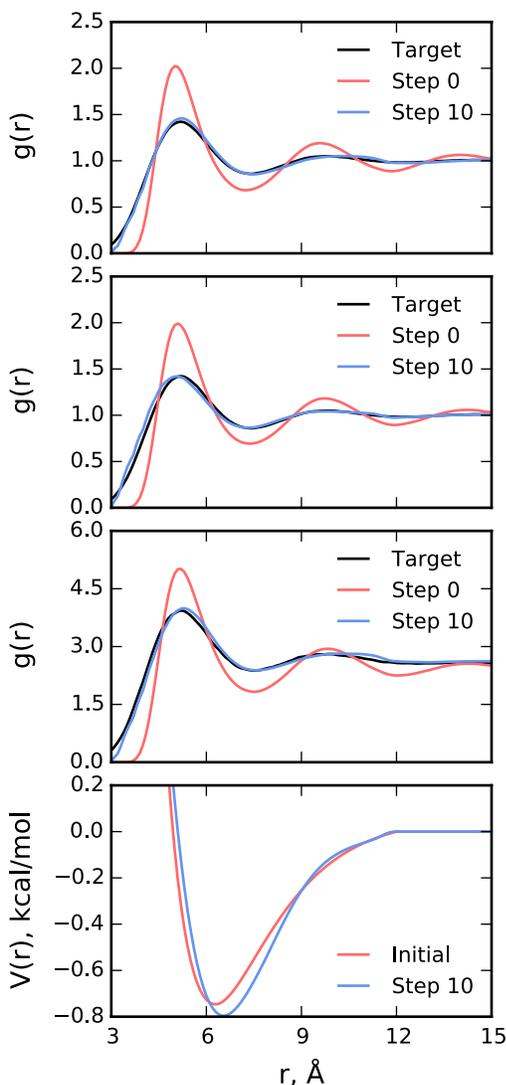


Figure 6.3: RDFs and potentials from the MS IBI potential derivation. Top: NVT; middle-top: NPT; middle-bottom: interface; bottom: potentials. The initial potential shows significant structural correlations missing from the target data. The derived potential at ten iterations shows excellent structural agreement with the target.

of 42 mN/m, lower than the original Chiu *et al.* potential (70.3 mN/m) which was optimized to match experiment, but only slightly perturbed from the modified potential (45 mN/m). This reduction in surface tension appears directly related to the softening of the potential, although, we note that this softening is required to provide an accurate match of the structure and that this value reasonably approximates the surface tension

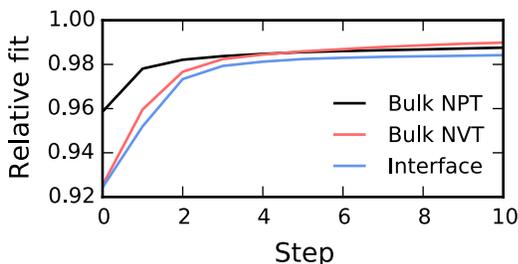


Figure 6.4:  $f_{\text{fit}}$  from Equation 5.5 as a function of iteration in the pair potential optimization. Convergence with the criterion is found after 10 iterations.

of the atomistic TIP3P model used as target data (52.3 mN/m) at 300 K.<sup>35</sup>

### 6.3.3 Validation and Comparison to Other Models

To further explore the efficacy of the MS IBI-derived model, the comparisons are made to other CG water models in the literature, namely, the  $k$ -means based potential of Hadley and McCabe<sup>21</sup> derived via the single state (SS) IBI procedure (here referred to as the SS IBI potential) and the MARTINI potential.<sup>18</sup> These models were chosen because they are short-ranged, non-polarizable, and 4:1 models. For reference, these potentials are plotted in Figure 6.5. Note that the MS IBI and SS IBI potentials are numerical (as they were derived via IBI), while the MARTINI potential is represented by a 12-6 Lennard-Jones potential with a well depth of 1.195 kcal/mol located at a separation of 5.276 Å. Note that all of the potentials considered in this paper provide a close estimate of the density of water at 1 atm and 305 K, as reported in Table 6.1.

Table 6.1: Density of water at 305 K, 1 atm calculated with different models.

Model	Density (g/mL)
TIP3P	$1.037 \pm 0.004$
MS IBI	$1.027 \pm 0.006$
SS IBI	$1.083 \pm 0.008$
MARTINI	$1.015 \pm 0.003$
Chiu	$0.991 \pm 0.003$

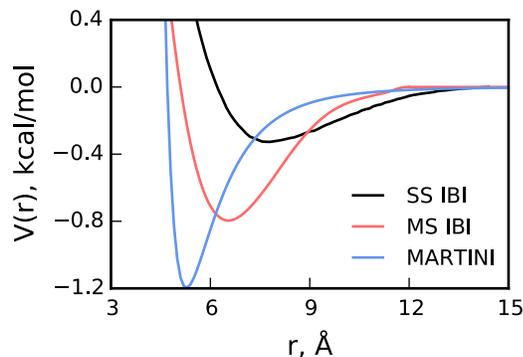


Figure 6.5: Interaction potentials from the CG water models compared in this work. The MS IBI and SS IBI potentials are numerical, derived with structure-based methods. MARTINI is a Lennard-Jones 12-6 potential.

First considering the SS IBI potential, it can be seen that the well depth is approximately 0.5 kcal/mol weaker than the MS IBI potential and shifted to larger separations. While this has little impact on the density or the structural correlations of the bulk states, simulations of droplets show that the interfacial properties are not sufficiently captured. Specifically, as shown in Figure 6.6, simulations of atomistic TIP3P, SS IBI, and MS IBI water were performed with interfaces. From these simulations it can be seen that the SS IBI potential model fills the box, rather than maintaining the interface. In contrast, the MS IBI model maintains a stable interface in agreement with the atomistic model. Thus, while an exact match to the experimental surface tension is not found for the MS IBI potential, as discussed above, it is still sufficiently strong to maintain a clear interface, providing a significant improvement over the SS IBI potential. We note that the difference between the SS IBI and MS IBI potentials is likely related to the aforementioned issue whereby many potentials can give rise to matching RDFs, and SS IBI provides no means to determine which ones are most physical. This limitation is overcome by the use of the interfacial state during the MS IBI optimization.

It is also important that the potential is not so strong that the system can solidify

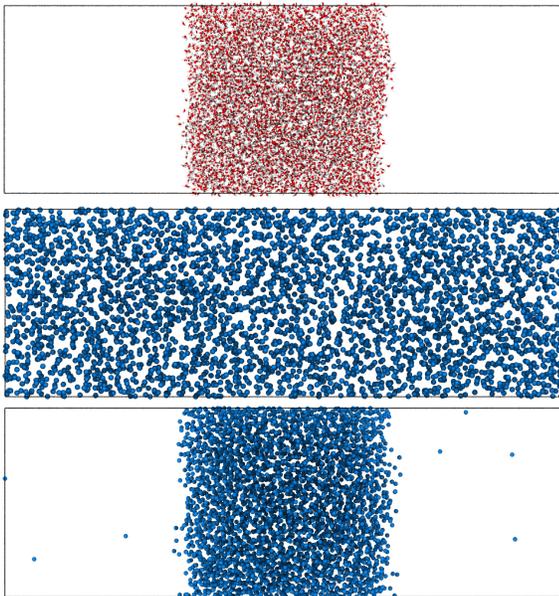


Figure 6.6: Simulation snapshots of droplets using the various models discussed. Top: all-atom TIP3P; middle: SS IBI; bottom: MS IBI. Atomistic and MS IBI models agree, producing a system with a stable interface, whereas SS IBI does not form a stable interface.

at physiological conditions. For example, the MARTINI water model is known to spontaneously crystallize at physiologically relevant temperatures.<sup>18</sup> This phenomenon is enhanced by the presence of interfaces (*e.g.*, a lipid bilayer surface), and requires the addition of unphysical “antifreeze” particles to avoid crystallization. While we note that modifications to the MARTINI water model exist (*e.g.*, adding charge polarization<sup>37,38</sup>), only the original MARTINI model was tested, since it more closely resembles the model derived via MS IBI (*i.e.*, both represent 4 water molecules as a single, spherically symmetric interaction site). To test the crystallization tendency, a nucleation site is generated with the following protocol. A crystalline state is generated by running a simulation with the MS IBI potential in the NVT ensemble. During this simulation, the temperature is decreased from 305 K to 1 K over 10 ns. A subsequent CG simulation is run at 1000 K, where the middle-most 1/8th of the beads are kept fixed, resulting in a configuration that contains a crystal seed surrounded by a fluid of CG water beads. The beads in the crystal seed are kept fixed in the nucleation site

simulations, with interactions identical to the fluid interactions. While neither model shows a tendency to freeze at 305 K in the absence of a nucleation site over a 100 ns simulation, the MARTINI model rapidly crystallizes in the presence of a nucleation site, while the MS IBI potential remains fluid (see Figure 6.7). Note, for a direct comparison with the MS IBI model derived here, antifreeze particles were not used with the MARTINI model. To ensure that the MS IBI system is not an amorphous solid structure, the ratio of the diffusion coefficients with and without a nucleation site were calculated for each model from the slope of the mean-squared displacement. As shown in Table 6.2, the diffusion coefficient of the MS IBI potential model remains relatively unchanged when a nucleation site is added, whereas a significant drop is seen for the MARTINI model resulting from the crystallization. Additionally, Figure 6.8 shows the RDF of the MARTINI model for the bulk NVT state as compared to the 4:1 mapped target data. Clearly, the MARTINI potential does not accurately capture the structural correlations of bulk water, further demonstrating the significant improvement of the MS IBI model in reproducing key properties of water.

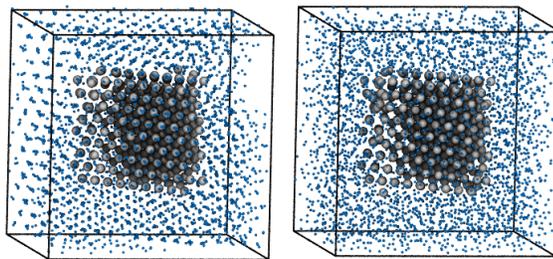


Figure 6.7: Configurations from simulations in the presence of a nucleation site with the MARTINI (left) and MS IBI (right) models. CG water beads colored silver were kept fixed during the simulations, but were treated as the same type as the blue particles (i.e., the color is different to show the nucleation site).

We note that the self-diffusion coefficient of MS IBI water is calculated to be  $16.07 \times 10^{-9} \text{ m}^2/\text{s}$  at 305 K and 1 atm, as compared to  $3.05 \times 10^{-9} \text{ m}^2/\text{s}$  for the atomistic TIP3P water at the same conditions, both run for 5 ns. This factor of  $\sim 5$  difference is not entirely unexpected, given the softening of the free energy landscape

Table 6.2: Ratio of diffusion coefficients from simulations with ( $D_{\text{nuc}}$ ) and without ( $D_{\text{bulk}}$ ) a nucleation site with different potentials.

Model	$D_{\text{nuc}}/D_{\text{bulk}}$
MS IBI	0.88
MARTINI	0.02

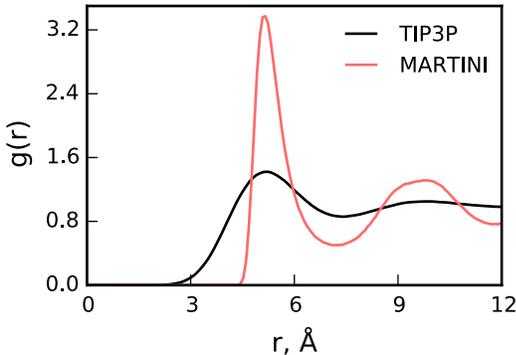


Figure 6.8: RDFs of the MARTINI model and the atomistic TIP3P model mapped to the CG level for the bulk NVT state.

that often comes with CG models and the fact that kinetic data was not used in the optimization. However, we also note that the dynamics of the CG model does not bear a strong connection with the atomistic level behavior, given that each CG bead represents 4 water molecules, but not necessarily the same water molecules through time, due to the lack of permanent bonds between the waters being grouped together.

## 6.4 Conclusions

In this Chapter, the MS IBI method was used to derive the interactions for a 4:1 mapped CG water model, using a modified version of the Chiu *et al.*, potential as an initial guess. An improvement over previous models is made by simultaneously matching the fluid structure to target data from bulk and interfacial states. It was shown that a model that reproduces the structure and density of water does not necessarily reproduce the interfacial properties and that the addition of a droplet

target state constrains the potential to also capture the interfacial properties. The resulting potential is able to accurately predict the density of water at 305 K and 1 atm, interfacial properties, and structural correlations. Additionally, the model shows no tendency to spontaneously crystallize at physiological conditions. This is important, since inaccuracies in a water model can propagate as more potentials are derived against it when simulating mixed systems.

This work highlights a key advantage of deriving potentials via the MS IBI approach. For simulations that cover multiple states, it is important to have a force field that is accurate across the states of interest. MS IBI allows this to be achieved by including target data from states that represent structures present in the states of interest. This is realized here by including a multi-ensemble state to accurately model the pressure-density relationship, and a droplet state to capture the interfacial properties of water. Another case where this would be beneficial is studying systems over multiple phases, *e.g.*, phase transitions in liquid crystals. While clever approaches are taken to capture behavior across multiple states,<sup>39</sup> a more systematic approach would be useful. Based on the results presented here, we foresee this method being useful for deriving CG potentials for a wide range of applications.

## 6.5 Bibliography

- [1] MOORE, T. C., IACOVELLA, C. R., AND M<sup>c</sup>CABE, C. Development of a coarse-grained water forcefield via multistate iterative Boltzmann inversion. In *Foundations of Molecular Modeling and Simulation*. Springer (2016), pp. 37–52.
- [2] WANG, Y. AND VOTH, G. A. Unique spatial heterogeneity in ionic liquids. *Journal of the American Chemical Society* **127**, 2005, pp. 12192–12193.
- [3] BHARGAVA, B. L., DEVANE, R., KLEIN, M. L., AND BALASUBRAMANIAN, S. Nanoscale organization in room temperature ionic liquids: a coarse grained molecular dynamics simulation study. *Soft Matter* **3**, 2007, pp. 1395–1400.
- [4] KARIMI-VARZANEH, H. A., MÜLLER-PLATHE, F., BALASUBRAMANIAN, S., AND CARBONE, P. Studying long-time dynamics of imidazolium-based ionic liquids with a systematically coarse-grained model. *Physical Chemistry Chemical Physics* **12**, 2010, pp. 4714–4724.
- [5] PADDING, J. AND BRIELS, W. Time and length scales of polymer melts studied by coarse-grained molecular dynamics simulations. *The Journal of Chemical Physics* **117**, 2002, pp. 925–943.
- [6] HARMANDARIS, V. A., FLOUDAS, G., AND KREMER, K. Temperature and pressure dependence of polystyrene dynamics through molecular dynamics simulations and experiments. *Macromolecules* **44**, 2010, pp. 393–402.
- [7] SUN, Q. AND FALLER, R. Crossover from unentangled to entangled dynamics in a systematically coarse-grained polystyrene melt. *Macromolecules* **39**, 2006, pp. 812–820.
- [8] MILANO, G. AND MÜLLER-PLATHE, F. Mapping atomistic simulations to mesoscopic models: a systematic coarse-graining procedure for vinyl polymer chains. *The Journal of Physical Chemistry B* **109**, 2005, pp. 18609–18619.
- [9] SHINODA, W., DEVANE, R., AND KLEIN, M. L. Coarse-grained molecular modeling of non-ionic surfactant self-assembly. *Soft Matter* **4**, 2008, pp. 2454–2462.
- [10] LEE, H. AND PASTOR, R. W. Coarse-grained model for pegylated lipids: effect of pegylation on the size and shape of self-assembled structures. *The Journal of Physical Chemistry B* **115**, 2011, pp. 7830–7837.
- [11] SRINIVAS, G., DISCHER, D. E., AND KLEIN, M. L. Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics. *Nature Materials* **3**, 2004, pp. 638–644.
- [12] NGUYEN, H. D. AND HALL, C. K. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2004, pp. 16180–16185.

- [13] IACOVELLA, C. R., KEYS, A. S., AND GLOTZER, S. C. Self-assembly of soft-matter quasicrystals and their approximants. *Proceedings of the National Academy of Sciences* **108**, 2011, pp. 20935–20940.
- [14] HADLEY, K. R. AND M<sup>c</sup>CABE, C. Coarse-grained molecular models of water: a review. *Molecular Simulation* **38**, 2012, pp. 671–681.
- [15] BASDEVANT, N., BORGIS, D., AND HA-DUONG, T. A semi-implicit solvent model for the simulation of peptides and proteins. *Journal of Computational Chemistry* **25**, 2004, pp. 1015–1029.
- [16] BASDEVANT, N., HA-DUONG, T., AND BORGIS, D. Particle-based implicit solvent model for biosimulations: application to proteins and nucleic acids hydration. *Journal of Chemical Theory and Computation* **2**, 2006, pp. 1646–1656.
- [17] MASELLA, M., BORGIS, D., AND CUNIASSE, P. Combining a polarizable force-field and a coarse-grained polarizable solvent model. ii. accounting for hydrophobic effects. *Journal of Computational Chemistry* **32**, 2011, pp. 2664–2678.
- [18] MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P., AND DE VRIES, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 2007, pp. 7812–7824.
- [19] CHIU, S.-W., SCOTT, H. L., AND JAKOBSSON, E. A coarse-grained model based on morse potential for water and n-alkanes. *Journal of Chemical Theory and Computation* **6**, 2010, pp. 851–863.
- [20] SHINODA, W., DEVANE, R., AND KLEIN, M. L. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation* **33**, 2007, pp. 27–36.
- [21] HADLEY, K. R. AND M<sup>c</sup>CABE, C. On the investigation of coarse-grained models for water: balancing computational efficiency and the retention of structural properties. *The Journal of Physical Chemistry B* **114**, 2010, pp. 4590–4599.
- [22] VAN HOOFF, B., MARKVOORT, A. J., VAN SANTEN, R. A., AND HILBERS, P. A. The cumulus coarse graining method: transferable potentials for water and solutes. *The Journal of Physical Chemistry B* **115**, 2011, pp. 10001–10012.
- [23] REITH, D., PÜTZ, M., AND MÜLLER-PLATHE, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24**, 2003, pp. 1624–1636.
- [24] MOORE, T. C., IACOVELLA, C. R., AND M<sup>c</sup>CABE, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140**, 2014, p. 224104.

- [25] BAYRAMOGLU, B. AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 2012, pp. 9205–9219.
- [26] QIAN, H.-J., CARBONE, P., CHEN, X., KARIMI-VARZANEH, H. A., LIEW, C. C., AND MÜLLER-PLATHE, F. Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures. *Macromolecules* **41**, 2008, pp. 9919–9929.
- [27] BAYRAMOGLU, B. AND FALLER, R. Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion. *Macromolecules* **46**, 2013, pp. 7957–7976.
- [28] CARBONE, P., VARZANEH, H. A. K., CHEN, X., AND MÜLLER-PLATHE, F. Transferability of coarse-grained force fields: The polymer case. *The Journal of Chemical Physics* **128**, 2008, p. 064904.
- [29] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 1983, pp. 926–935.
- [30] PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* **117**, 1995, pp. 1–19.
- [31] A Git repository for this package is hosted at <https://github.com/mosef-hub/msibi>.
- [32] ANDERSON, J. A., LORENZ, C. D., AND TRAVESSET, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 2008, pp. 5342–5359.
- [33] GLASER, J., NGUYEN, T. D., ANDERSON, J. A., LUI, P., SPIGA, F., MILLAN, J. A., MORSE, D. C., AND GLOTZER, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **192**, 2015, pp. 97–107.
- [34] MCGIBBON, R. T., BEAUCHAMP, K. A., HARRIGAN, M. P., KLEIN, C., SWAILS, J. M., HERNÁNDEZ, C. X., SCHWANTES, C. R., WANG, L.-P., LANE, T. J., AND PANDE, V. S. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal* **109**, 2015, pp. 1528 – 1532.
- [35] VEGA, C. AND DE MIGUEL, E. Surface tension of the most popular models of water by using the test-area simulation method. *The Journal of Chemical Physics* **126**, 2007, p. 154707.
- [36] WANG, H., JUNGHANS, C., AND KREMER, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *The European Physical Journal E* **28**, 2009, pp. 221–229.

- [37] YESYLEVSKYY, S. O., SCHÄFER, L. V., SENGUPTA, D., AND MARRINK, S. J. Polarizable water model for the coarse-grained martini force field. *PLoS Computational Biology* **6**, 2010, p. e1000810.
- [38] ZAVADLAV, J., MELO, M. N., MARRINK, S. J., AND PRAPROTNIK, M. Adaptive resolution simulation of polarizable supramolecular coarse-grained water models. *The Journal of Chemical Physics* **142**, 2015, p. 06B622\_1.
- [39] MUKHERJEE, B., DELLE SITE, L., KREMER, K., AND PETER, C. Derivation of coarse grained models for multiscale simulation of liquid crystalline phase transitions. *The Journal of Physical Chemistry B* **116**, 2012, pp. 8474–8484.

## CHAPTER 7

### DEVELOPMENT OF A COARSE-GRAINED MODEL OF CERAMIDE NS AND FFA

This Chapter is reproduced in part from Moore *et al.* 2016.<sup>1</sup>

#### 7.1 Introduction

The stratum corneum (SC) is the outermost layer of the skin, and serves as the barrier between the internal and external environments of the body, preventing, among other things, rapid dehydration and pathogen invasion.<sup>2</sup> In addition to cholesterol and free fatty acids (FFAs), several different ceramides (CERs) comprise the lipid matrix of the SC.<sup>3</sup> To date, 15 subclasses of CER have been identified in human skin that differ in headgroup and acyl tail structure.<sup>4</sup> Of these, nonhydroxy-sphingosine ceramide (CER NS, Figure 7.1) is the most abundant in healthy human skin.<sup>5</sup>

Despite knowledge of the lipid composition in human SC, the connection between lipid composition, structure, and barrier function remains unclear.<sup>6,7</sup> Understanding these connections is essential for developing effective models of skin for use in toxicology assessment and transdermal drug delivery development, as well as for developing effective treatments for skin diseases.<sup>7,8</sup> Molecular simulation studies of SC lipid systems could allow the connections between lipid composition, structure, and barrier function to be probed, as simulation enables direct visualization of and explicit control over the molecular species in the system. For example, molecular simulations have been used extensively to better understand the structure, phase behavior, thermodynamics, and transmembrane permeability of phospholipid-based bilayer systems.<sup>9–14</sup> However, CER-based bilayers have not been as widely studied as phospholipid-based systems, primarily because of the more complex nature of the lipid organization<sup>15–17</sup> and a

lack of accurate atomistic force fields, which was recently addressed.<sup>18</sup> The limited atomistic (*i.e.*, all-atom and united-atom) simulations that have been performed of CER bilayer structures have provided some insight into the lipid interactions,<sup>18–22</sup> *e.g.*, demonstrating strong connections between the phase transition temperature, CER headgroup structure, and lipid–lipid hydrogen bonding.<sup>18</sup> However, the dense gel and crystalline packings found in CER-rich phases<sup>23–27</sup> significantly reduce lipid mobility as compared to the liquid-crystalline phases typical of biological phospholipid-based bilayers under physiological conditions, resulting in the need for careful equilibration protocols and long equilibration times.<sup>20,28,29</sup> As such, the time and length scales on which certain phenomena occur (*e.g.*, phase-separation, self-assembly) are limiting for all-atom models. Furthermore, bilayer structures consisting of a small number of components are highly simplified models of SC lipid lamellae; to study mixtures relevant to the multicomponent SC that form multilamellar structures, large system sizes must be employed, further increasing the computational cost.

Coarse-grained (CG) molecular models, where groups of atoms are treated as single interaction sites (CG beads), provide an attractive alternative to atomistic models. CG models typically require several orders of magnitude less computational cost to access the same time scale as the corresponding atomistic model, due to the simplified representation of the system. This simplification results in fewer pair interactions, allows implicit treatment of long-ranged electrostatics, and generally softens the energy landscape, allowing a larger time step of integration. Combined, these properties allow CG models to access the long time scales needed to form equilibrium phases via self-assembly and probe the large system sizes needed to represent the SC lipid mixture. CG models have proven effective for the study of various lipid systems,<sup>30–32</sup> including studies that examine the self-assembly of other lipids important to the skin barrier.<sup>33</sup> Recently, Sovová *et al.* proposed a CG model of CER NS, which has been the most studied CER via all-atom molecular simulation.<sup>18–21,34</sup> This model

predicts stable lamellar structures as expected of CER conformations in the SC,<sup>35</sup> and is advantageous in that it is compatible with the popular MARTINI force field; however since it was not derived to match CER structure, the model fails to accurately capture key structural properties of pure CER NS bilayers, overestimating the area per lipid by  $\sim 20\%$  and substantially underestimating the tilt angle as compared to experimental and all-atom simulation results (*i.e.*, no tilt is observed with this model).<sup>18,21,35</sup> Furthermore, the CG mapping used in this model (*i.e.*, how the atoms are grouped into CG beads) lacks directional headgroup interactions that have been shown to influence the properties of CER systems,<sup>18,36,37</sup> overall making the CG model more representative of a generic lipid rather than CER NS, and thus of limited utility for accurately probing the behavior of systems representative of the SC.

In this chapter, a CG mapping and a force field for CER NS are derived to match structural properties calculated from all-atom simulations. The CER NS model proposed herein provides a more explicit treatment of hydroxyl groups than in the MARTINI-compatible CG CER model of Sovová *et al.*,<sup>35</sup> allowing the model to capture directional headgroup interactions that exist between lipids due to hydrogen bonding.<sup>18,38–40</sup> As CERs and FFAs contain identical tails (*i.e.*, long alkyl chains), a CG FFA model is developed in parallel, using the same tail beads for CER and FFA. The recently developed multistate iterative Boltzmann inversion (MS IBI)<sup>41</sup> scheme is used to optimize these CG force fields. While potentials derived with the original IBI scheme generally reproduce target RDFs with a high degree of accuracy, they often show limited transferability to other states,<sup>42,43</sup> due to the single state nature of the scheme. They may also demonstrate significant artifacts associated with structural correlations between beads (*e.g.*, potential wells and barriers associated with RDF peaks and valleys), and hence can be unrepresentative of the underlying potential landscape.<sup>41</sup> The MS IBI algorithm was developed to address these issues; specifically, the multistate nature of the method provides additional constraints to the

optimization, reducing structural artifacts, and yields potentials that are more generally applicable than those derived from the original (*i.e.*, single state) IBI method,<sup>41,44</sup> as is required for simulations that span multiple state points, such as simulations of lipid self-assembly. The structural properties and self-assembly of CER systems are studied using the derived CG model and compared to all-atom and experimental measurements, demonstrating close agreement and providing validation of the model.

## 7.2 Models and Methods

Deriving CG potentials via MS IBI involves several general steps. First, all-atom simulations are performed at various state points that are chosen to emphasize the different behaviors the CG model should reproduce. For example, for CERs, isotropic fluid and ordered lamellar states are included, which are both relevant for self-assembly. The all-atom trajectories are then used to create corresponding CG trajectories by mapping the all-atom trajectories to the CG level using a CG mapping operator that defines how the atoms are grouped into CG beads. Next, the CG trajectories mapped from the all-atom simulations are used to generate target data in the form of pair radial distribution functions (RDFs) for each pair type at each state. The MS IBI algorithm is then employed to iteratively derive the CG force field, such that the RDFs from the CG simulations match the target RDFs mapped from the all-atom simulations. As a last step, individual pair interactions in the force fields are refined to reproduce the correct balance between hydrophobic and hydrophilic interactions via simulated wetting experiments. The methodologies used in each of these steps are discussed in detail below.

### 7.2.1 Atomistic Simulations

All-atom simulations were performed with the fully atomistic CHARMM36 force field,<sup>45</sup> employing the TIP3P water model<sup>46</sup> and CHARMM-compatible headgroup

parameters for CERs, which were derived and validated in previous work.<sup>18</sup> All atomistic simulations were performed using the LAMMPS simulation engine,<sup>47,48</sup> with a van der Waals interaction cutoff of 12 Å; the PPPM solver with a real space cutoff of 12 Å was used for the calculation of the long-ranged electrostatic interactions.<sup>49</sup> A time step of 1.0 fs was used with the Nosé-Hoover thermostat in the canonical ensemble (*i.e.*, NVT, constant number of molecules, volume and temperature) and an additional barostat for simulations in the isothermal isobaric ensemble.<sup>50</sup> For NPT simulations, isotropic pressure control was used for bulk fluid states, while anisotropic pressure control was used for bilayer states, allowing the aspect ratio of the box to change. All-atom simulations used for gathering target data were run for 10 ns after sufficient equilibration with the trajectory written every 100 ps. A list of all the states used is given in Table 7.1, and the weights given to each state are listed in the Table 7.2; the rationale for the choice of the specific states is discussed in Section 7.3. Note that FFA C24:0 corresponds to a fully saturated FFA with a length of 24 carbons. CER NS C16 corresponds to CER NS with an acyl tail containing 16 carbons and is therefore of roughly equal length to the sphingosine tail. Similarly, CER NS C24 corresponds to a CER NS molecule with a longer acyl tail of 24 carbons (as in Figure 7.1a).

Bulk fluid simulations (both FFA and CER) were initialized by placing lipids on a  $6 \times 6 \times 6$  cubic lattice in a large simulation box to create a gas of 216 molecules. To randomize the lipids and distribute them evenly throughout the box, NVT simulations were performed for 5 ns at 1300 K. The box volume was then isotropically reduced to achieve the target density over 5 ns, which was found to be sufficient for the potential energy to converge, followed by 10 ns of simulation over which target data was collected; this was performed in either the NVT or NPT ensemble, depending on the state point of interest. The mixed FFA-water systems were initialized in the same manner; these systems contained 6480 water molecules and the appropriate number of lipids for the given concentration. The same approach was also used for the CER headgroups in

Table 7.1: Target states used for MS IBI optimizations. The states used to optimize each interaction are described in the text and given in Table 7.2. State condition specifies the ensemble—states with a pressure listed were simulated in the NPT ensemble; states with a density listed were simulated in the NVT ensemble. The dehydrated bilayer values are the area per lipid and the thickness (*i.e.*, the length of the simulation box normal to the bilayer plane since water is absent in this simulation).

Name	State condition	Temperature, K	Label for Table 7.2
<b>Pure FFA C24:0</b>			
Bulk fluid	1.0 atm	500	FFA1
Bulk fluid	0.811 g/mL	800	FFA2
Monolayer	30.0 Å <sup>2</sup> /chain	305	MONO
<b>FFA–water</b>			
0.4 mol % FFA	1.0 atm	305	W-FFA1
0.5 mol % FFA	1.0 atm	305	W-FFA2
0.85 mol % FFA	1.0 atm	305	W-FFA3
<b>Pure CER NS</b>			
Bulk fluid	1.0 atm	500	CER1
Bulk fluid	0.721 g/mL	500	CER2
Dehydrated bilayer	1.0 atm	305	DEH-BIL-NPT
Dehydrated bilayer	39.96 Å <sup>2</sup> , 57.18 Å	305	DEH-BIL-NVT
<b>CER NS–water</b>			
Bilayer	1.0 atm	305	BIL-NPT
Bilayer	40.0 Å <sup>2</sup>	305	BIL-NVT
CER headgroups	1.0 atm	400	W-CERHEADs

Table 7.2: States used to optimize each pair potential and the weight  $\alpha_{s,0}$  given to each state. Pairs are denoted by bold text, with the states used to optimize a given (set of) interaction(s) listed below the interaction name. The state condition names correspond to the label given in Table 7.1. To illustrate the dependence of the chosen weights on the force field, an additional optimization of the lipid–lipid self-interaction potentials was performed, where relatively higher weights were given to the bulk fluid states, given as  $\alpha_{s,0}(\text{high bulk})$ . Note that lipid–water interactions were not optimized with  $\alpha_{s,0}(\text{high bulk})$ .

States used	$\alpha_{s,0}$	$\alpha_{s,0}(\text{high bulk})$
<b>TER2–TER2, TER2–TAIL, TER2–FHEAD, TAIL–FHEAD, FHEAD–FHEAD</b>		
FFA1 + FFA2	0.1	0.8
<b>TAIL–TAIL</b>		
FFA1 + FFA2	0.1	0.8
MONO	0.4	0.1
DEH-BILs	0.8	0.1
<b>CERHEADs–CERHEADs</b>		
CER1 + CER2	0.1	0.8
DEH-BILs	0.8	0.1
<b>CERHEADs–TAILs</b>		
CER1 + CER2	0.1	0.8
<b>FHEAD–WATER + TAILs–WATER</b>		
W–FFA[1–3]	0.3	—
<b>CERHEADs–WATER</b>		
BILs + W–CERHEADs	0.3	—

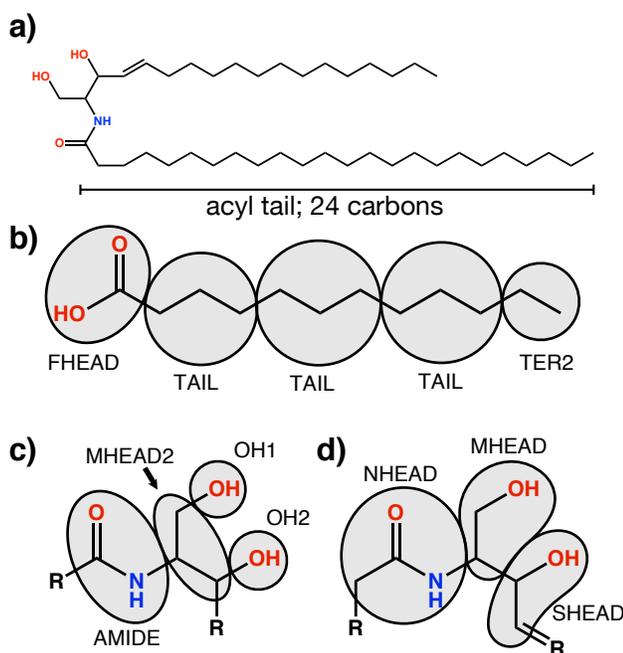


Figure 7.1: CER NS structure and CG mappings used in this work: (a) molecular structure of CER NS C24, where C24 gives the length of the CER acyl chain; (b) FFA mapping showing head (FHEAD), 3:1 tail beads (TAIL), and 2:1 terminal bead (TER2) as applied to C12:0; (c) four-site CER headgroup mapping proposed in this work; (d) three-site CER headgroup mapping. All beads interact through spherically symmetric pair potentials, even if drawn as irregular shapes. Note that beads are not drawn to scale but rather to emphasize the mapping and connectivity.

water, which contained 4000 water molecules and 728 CER headgroups.

Atomistic gel-phase CER NS bilayers were initialized and equilibrated in the same manner as in previous work.<sup>18</sup> Specifically, initial configurations were generated by placing lipids on a  $10 \times 10$  square lattice to create a leaflet. The leaflet was then replicated and inverted to create a bilayer and the bilayer solvated with 50 water molecules per lipid. A short energy minimization was performed to remove any high-energy atomic overlaps, followed by 30 ps simulation in the NVT ensemble at 305 K and 1.0 atm in the anisotropic NPT ensemble and the final 10 ns used to collect target data; we note that, since large oscillations in potential energy and area per lipid were not observed over the 100 ns of simulation, 10 ns is sufficient for collecting data on a property such as the RDF. Furthermore, previous work<sup>18</sup> has shown that

the structural properties of CER NS bilayer systems simulated for 100 ns at 305 K agree well with those from systems slowly cooled from just below the order–disorder transition, providing confidence that our systems are adequately equilibrated with this procedure. Dehydrated bilayer systems were initialized from the final configuration of the hydrated bilayer simulation, and simulated for an additional 40 ns to ensure convergence of the potential energy and area per lipid, with target data collected from the final 10 ns.

### 7.2.2 CG Simulations

All CG simulations utilized the HOOMD-Blue simulation engine.<sup>51,52</sup> Integration of the equations of motion follows Martyna–Tobias–Klein<sup>50</sup> for both NPT and NVT simulations. All CG simulations employed a nonbonded potential cutoff of 12 Å. Since CG beads are treated as electroneutral, no treatment of long-ranged electrostatics was needed. In all cases, a time step of 10 fs was used. Unless otherwise noted, the initial configuration of the CG simulation was the final configuration from the all-atom simulation, mapped to the CG level. For bulk fluid systems, this configuration was replicated twice in each direction so that the CG bulk fluid systems contained 8 times more molecules than the corresponding atomistic systems. For lamellar states, this configuration was replicated twice in each direction in the lamellar plane so that CG lamellar systems contained 4 times more molecules than the corresponding atomistic systems.

### 7.2.3 CG Mappings

The CG mappings used for the lipids studied in this work are shown in Figure 7.1. The position of each CG bead is taken to be the center of mass of the atoms represented by the bead, with the mass of the CG bead defined to be the mass of the atoms it represents. The mapping for FFA is shown in Figure 7.1b. The carboxylic acid headgroup atoms are naturally mapped to a single CG bead (FHEAD), with the lipid

tail described by “TAIL” beads that represent three methyl groups and a TER2 bead to describe the terminal two methyl groups when necessary, *e.g.*, in FFA C12:0, as shown in Figure 7.1b. Henceforth, the TAIL and TER2 beads in the FFA chain are collectively referred to as TAILS. Note that this mapping closely resembles that of Hadley and McCabe,<sup>53</sup> with the exception of using a 3:1 rather than a 4:1 mapping for the tail beads.

A four-site mapping for the CER headgroups is used, as shown in Figure 7.1c. An “AMIDE” bead maps the amide group; and “MHEAD2” bead maps the C1, C2, and C3 carbons (and their associated hydrogens) of the sphingosine tail (note that the name “MHEAD” is reserved for the middle headgroup bead in the three-site model, discussed below); an “OH1” bead maps the hydroxyl group on C1 of the sphingosine chain; and an “OH2” bead maps the hydroxyl group on C3 of the sphingosine chain. This CG mapping treats the hydroxyl groups explicitly to account for the strong directional interactions between the CER headgroups (*i.e.*, hydrogen bonding). Additionally, this mapping scheme provides a simple route for extension to CER species with different headgroups, since the AMIDE and MHEAD2 beads are common to all CER subclasses. We refer to the beads in the CER headgroup collectively as CERHEADs. This mapping uses the same tail mappings as the FFAs, and as such, TAIL beads in CER and FFA are described by the same interaction potentials. Note, while optimizations use CER NS C24 (see Figure 7.1a), the validation studies performed also include CER NS C16.

A three-site CER headgroup mapping was also investigated, as shown in Figure 7.1d to demonstrate the importance of the explicit treatment of hydroxyl groups. Here, the effect of the hydroxyl groups is treated implicitly within the backbone structure, rather than being explicitly modeled as separate groups, and as such is similar to the MARTINI CER mapping of Sovová *et al.*<sup>35</sup> The tails are treated the same as in the four-site CER headgroup model, using the same mappings and force field for the TAILS as in the FFA.

In CG simulations where water is present, a 4:1  $k$ -means mapped CG water model<sup>44</sup> was used that was developed by following the work of Hadley and M<sup>c</sup>Cabe in which waters are dynamically assigned to CG beads.<sup>54</sup> A dynamic mapping is necessary for water because a 4:1 mapping means atoms mapped to a common bead are necessarily parts of different molecules, and these molecules will not remain associated throughout the course of a simulation due to molecular diffusion. Thus, to map atomistic water to the CG level, a new mapping is effectively defined for each frame of the atomistic trajectory, with clusters of four molecules mapped to a single bead. The location of the center of mass of these beads is then used when calculating RDFs involving CG water. Since the original version of the water model was only optimized for bulk properties, it was recently reparameterized to also capture interfacial behavior.<sup>44</sup> As detailed in Chapter 6, the CG water model was optimized against TIP3P all-atom simulations (that is, the same water model used in the all-atom lipid simulations) using both bulk and interfacial properties of water as target states, with the resulting force field accurately capturing structure, density, and droplet stability at 305 K and 1 atm.<sup>44</sup>

#### 7.2.4 CG Force Field Derivation

The force field developed here consists of CG beads interacting through bonded and nonbonded potentials. These interactions are treated independently of one another, and are therefore derived separately.

The bonded interactions were derived from bond/angle distributions calculated from all-atom simulations. CG beads in a given molecule are considered bonded if they occur sequentially (*i.e.*, if they each contain an atom that shares a bond in the all-atom model). Additionally, any three consecutively bonded CG beads interact through an angle-bending potential. No torsional or dihedral interactions were included in the models. Instead, beads separated by three or more bonds interact

only through nonbonded pair potentials, whereas the nonbonded pair potential is excluded from beads separated by two or fewer bonds. As in previous work,<sup>53,55</sup> bonded interactions are determined from probability distributions following the approach of Milano *et al.*<sup>56</sup> Bond and angle interactions were modeled as harmonic springs, with equilibrium values and spring constants derived from the distributions sampled in the fluid-state atomistic simulations mapped to the CG level. After mapping the atomistic trajectories to the CG level, a normalized Gaussian distribution  $p(r)$  was fitted to the sampled distribution for each bond/angle type. A Boltzmann inversion of a Gaussian distribution yields a harmonic potential of the form

$$V(r) = -k_{\text{B}}T \ln(p) = K_{\text{r}}(r - r_0)^2,$$

where  $k_{\text{B}}$  is the Boltzmann constant,  $T$  is the absolute temperature,  $K_{\text{r}}$  is the spring constant determining the stiffness of the harmonic spring,  $r$  is the separation, and  $r_0$  is the equilibrium separation. This procedure was performed for each bond/angle type at each state, and average  $K_{\text{r}}$  and  $r_0$  for each bond/angle type are given in Table 7.3.

Nonbonded interactions were iteratively optimized via MS IBI to match target RDFs. In MS IBI, the nonbonded CG pair potential  $V_i(r)$  is iteratively updated on the basis of deviations between RDFs from the CG-mapped all-atom (AA) simulations,  $g_s^{\text{AA}}(r)$ , and the RDFs from a CG simulation,  $g_s^{(i)}(r)$ , using  $V_i(r)$  at state  $s$ . The pair potential is updated according to the sum of the differences between the Boltzmann-inverted target and CG RDFs at different state points according to

$$V_{\text{CG}}^{(i+1)}(r) = V_{\text{CG}}^{(i)}(r) - N_{\text{states}}^{-1} \sum_s \alpha_s(r) k_{\text{B}} T_s \ln \left[ \frac{g_s^{(\text{target})}(r)}{g_s^{(i)}(r)} \right],$$

where  $\alpha_s(r)$  is a scaling factor to prevent large updates, and to place emphasis on certain parts of the potential;  $k_{\text{B}}$  is the Boltzmann constant; and  $T_s$  is the temperature of state  $s$ ;  $N_{\text{states}}$  is the number of target states used in the optimization. Note that

Table 7.3: Bond and angle parameters for the 4-site CG model. For bonds,  $x_0$  is the equilibrium bond length in Å and  $k_x$  is the harmonic bond stretching constant in kcal/(mol Å<sup>2</sup>). For angles,  $x_0$  is the equilibrium bond angle in deg and  $k_x$  is the harmonic angle bending constant in kcal/(mol rad<sup>2</sup>). \*Angles treated the same as TAIL-TAIL-TAIL because the sampled distributions from the AA simulations were very similar to the TAIL-TAIL-TAIL distribution.

Name	$x_0$	$k_x$
<b>Bonds</b>		
TER2-TAIL	3.41	35.576
TAIL-TAIL	3.92	20.928
TAIL-FHEAD	3.28	24.296
TAIL-AMIDE	3.01	103.857
AMIDE-MHEAD2	2.92	190.045
MHEAD2-TAIL	3.62	27.575
MHEAD2-OH1	2.33	45.307
MHEAD2-OH2	2.07	151.987
<b>Angles</b>		
TER2-TAIL-TAIL	155.0	7.93
TAIL-TAIL-TAIL	158.0	6.33
TAIL-TAIL-FHEAD	148.0	6.33
TAIL-TAIL-AMIDE*	158.0	6.33
MHEAD2-AMIDE-TAIL	106.0	29.00
AMIDE-MHEAD2-TAIL	107.0	5.29
TAIL-TAIL-MHEAD2*	150.0	6.33
AMIDE-MHEAD2-OH1	109.0	8.39
AMIDE-MHEAD2-OH2	105.0	10.76
TAIL-MHEAD2-OH1	128.0	5.67
TAIL-MHEAD2-OH2	69.0	361.77
OH1-MHEAD2-OH2	115.0	12.22

this expression reduces to the original IBI algorithm for a single state point. Following previous work,<sup>41</sup>  $\alpha_s(r)$  was chosen to be a linear function of the form

$$\alpha_s(r) = \alpha_{s,0} \left( 1 - \frac{r}{r_{\text{cut}}} \right),$$

where  $\alpha_{s,0}$  is the value at  $r = 0$  and determines the maximum weight given to a particular state. This form places more emphasis at smaller separations, and ensures that the potential remains 0 at the cutoff. This allows short-range interactions to be optimized first and to see their effect on long-range behavior, before longer range updates to the potentials are made, which helps to suppress structural artifacts in the force field that may arise due to intermediate and long-range ordering in the system. We also note that, while it could be argued that making smaller updates to the potential at larger separations may lead to thermodynamic inconsistencies in the model, the potentials developed through the MS IBI approach are not only able to match the target RDFs but also satisfy the thermodynamic relationship between density and pressure (including related properties such as area per lipid). Rather than applying a pressure correction as in the original IBI method, CG models derived via MS IBI use a combination of NVT and NPT state points during the optimization. The RDFs are not likely to match if the pair potential predicts the wrong density, since the RDF is normalized by the density; thus, NPT states can be used to account for the pressure, as was shown in the derivation of the CG water force field.<sup>44</sup>

The pair potential is iteratively updated until some stopping criteria is met. Here, the value of the fitness function

$$f_{\text{fit}}(i) = 1 - \frac{\sum_{r=0}^{r_{\text{cut}}} |g^{(i)}(r) - g^{(\text{target})}(r)|}{\sum_{r=0}^{r_{\text{cut}}} [g^{(i)}(r) + g^{(\text{target})}(r)]}.$$

was used to determine the stopping criteria. The optimizations were terminated when  $|f_{\text{fit}}(i) - f_{\text{fit}}(i-1)| < 0.01$  for each pair being optimized (*i.e.*, when the fitness function

became stable for each pair in the optimization). It is worth noting that MS IBI, like the original IBI method, yields numerical force fields that are not constrained to a functional form, thus allowing increased flexibility that is not available when fitting to an analytical form.

Optimizations for particular pairs of interactions were performed successively. This was done for two reasons: (1) to reduce the number of interactions being simultaneously optimized during any single MS IBI optimization and (2) to simplify the force field development by allowing pairs to be optimized from simulations of pure systems (*e.g.*, pure CER) or single component hydrated systems where necessary (*e.g.*, CER–WATER). The general order of the optimizations was to first optimize the lipid–lipid self-interactions and then, using those self-interactions, derive the lipid–water interactions for each lipid species. The lipid–lipid cross-interactions (*e.g.*, FFA–CER) were not derived in this work, since the focus is on the study of self-assembled CER structures; the cross-interactions are presented in Chapter 8. Validation for each set of potentials was performed before using those potentials to optimize other interactions. For example, the TAIL–TAIL validation was performed before the TAIL–WATER potential was derived.

Seven distinct targets were used to optimize the FFA and CER nonbonded lipid self-interactions to capture a range of behaviors, including bulk fluid states (both FFA and CER), as well as an FFA monolayer and dehydrated CER bilayers. Target data from three different concentrations of FFA in water were used to optimize the TAILS–WATER interactions, and two distinct states were used to optimize the CERHEADS–WATER interactions, including hydrated CER NS C24 bilayers and CER NS headgroups solvated in water. Optimizations in both the NVT and NPT ensembles were carried out for select targets to ensure that the force field can reproduce the correct density in constant pressure simulations at 1 atm, as previously discussed. Note that not all pair potentials were optimized at all states. Instead, pair potentials

were optimized at select states to capture specific behaviors. A complete list of the states used to optimize each pair potential is given in Table 7.1. Additionally, the weight given to each state is listed in Table 7.2. The weights were chosen to place more emphasis on the lamellar states; that is, relatively higher weights were given to the monolayer and dehydrated bilayer states, since the ultimate goal of the force field is accurate modeling of lamellar structures present in stratum corneum. The relative values were chosen on the basis of heuristics developed in this Chapter and Chapter 5. MS IBI calculations were carried out using the open source `msibi` Python library,<sup>57</sup> which also makes use of the `MDTraj` Python library.<sup>58,59</sup> The optimized nonbonded interaction potentials are available via <http://github.com/iModels/sc-ff>.

### 7.2.5 Analysis

Several structural properties were used to characterize the ordered bilayer and multilamellar systems, including the area per lipid (APL), bilayer thickness, nematic order parameter ( $S_2$ ), and tilt angle ( $\theta$ ) of the lipid tails. The hydrophobicity of a surface was characterized by calculating the contact angle ( $\varphi$ ) of a droplet that forms on the surface. A more detailed explanation of these calculations is provided in Appendix A.

## 7.3 Results

As mentioned above, the lipid–lipid self-interactions were derived first, followed by the lipid–water interactions. Even though the primary goal here is to derive a CG force field for CER NS, we note that the force field for FFA, also relevant to the SC, is simultaneously derived. Both CER and FFA contain the same tails beads; however, the headgroups of FFA molecules interact less strongly than the headgroups of the CER molecules. As such, inclusion of FFA target states alongside CERs helps to lessen the influence of the strong CER headgroup interactions when optimizing

potentials involving the TAILS beads.

### 7.3.1 Pure CER

*TAIL-TAIL* Dehydrated CER bilayers (305 K, NPT and NVT) and FFA monolayer (305 K) were used in the TAIL bead optimizations to ensure the force field is able to capture the behavior of ordered lamellar structures. While an ordered structure is clearly an important target for capturing the SC bilayer structure, bulk FFA fluid states in both the NVT (800 K, 0.811 g/mL) and NPT (500 K, 1 atm) ensembles are also used to capture the behavior of disordered states expected prior to self-assembly. The inclusion of fluid states is also found to help decrease the level of structural artifacts in the force field that arise from the intermediate and long-ranged structural correlations found in the ordered phases, and thus ensures a better quality force field.<sup>41</sup> Note that elevated temperatures were required to induce fluid phases at the density or pressure of interest. Also, due to periodic boundary conditions, the dehydrated bilayer is effectively an infinitely periodic multilamellar system.

Figure 7.2 shows the *TAIL-TAIL* RDFs for the FFA monolayer state and bulk FFA NPT state, comparing the target and optimized RDFs for each; the optimized pair potential is also shown. Considering the monolayer state, the CG model captures the main RDF peak locations and heights, although some of the smaller-scale features are missed, likely due to the spherical symmetry of the CG potential. The in-plane density of this monolayer system closely compares with the density of the chains commonly found for CERs in a bilayer arrangement. Thus, this system provides a close approximation to the tail ordering found in the CER NS system. Note that this RDF does not decay to unity as the monolayer only fills a small region of the box, resulting in a system density that is lower than the local monolayer density, scaling the value of the RDF; however, this does not impact the optimization as the CG RDFs are normalized by the same factor. Satisfactory agreement is found for the

bulk NPT state, where the CG model correctly predicts the presence of a fluid phase and accurately predicts the density at 1 atm, with  $\rho_{\text{CG}} = 0.68 \pm 0.05 \text{ g/mL}$ , which compares well with  $\rho_{\text{AA}} = 0.69 \pm 0.01 \text{ g/mL}$ . However, we note that the peaks in the CG RDF appear to be shifted to smaller separations by  $\sim 0.5 \text{ \AA}$ . As the main focus of the force field development is to develop a model that accurately represents lamellar structures of the lipids in the SC, we consider such small deviations in the fluid state acceptable. We emphasize that, via MS IBI, a single CG pair potential is optimized that is capable of modeling both ordered and fluid-like states; prior efforts to use the standard IBI methodology to generate CG models of FFA required different force fields to capture different structural configurations, even though the CG mappings not change here.<sup>53</sup>

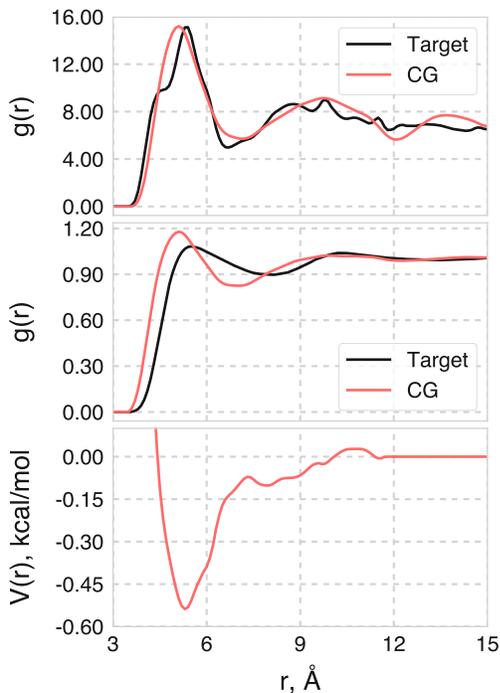


Figure 7.2: RDFs and pair potential from the potential optimization for the interactions between TAIL beads. Top, target (black) and CG (red) RDFs from the FFA monolayer state; middle, target (black) and CG (red) RDFs from the FFA bulk fluid NPT state (500 K, 1 atm); bottom, the final potential used to calculate the CG RDFs above.

As further evaluation of the TAIL–TAIL interaction, structural properties of

the CG FFA monolayer were also compared to results from atomistic simulations, since a matching TAIL-TAIL RDF does not guarantee that characteristic monolayer properties will also match. Well-ordered monolayers were observed in simulations of each model, with nematic order parameters  $S_{2,CG} = 0.97$  and  $S_{2,AA} = 0.90$  and tilt angles  $\theta_{CG} = 40 \pm 5^\circ$  and  $\theta_{AA} = 30 \pm 10^\circ$ . We note that the CG force field predicts a slightly increased level of ordering and increased tilt, although the differences in tilt angle are fully captured by the standard error of the two measurements, thus providing further validation of the pair potentials between TAILS beads.

*CER Headgroups* The CG pair potentials between the four beads of the CER headgroup (collectively referred to as CERHEADs) are derived from both NVT and NPT ensemble simulations of dehydrated bilayers (305 K) and isotropic fluids of CER NS (500 K, 1.0 atm; 500 K, 0.721 g/mL), where, similar to the fluid FFA states, elevated temperatures were required to induce a fluid state. We note that, while many prior simulations of CERs have focused on hydrated bilayers, the use of dehydrated lamellae removes the influence of the strong interactions with water and is more representative of the SC lamellae *in vivo*. Interactions with water are considered separately.

In the atomistic dehydrated CER NS bilayer, in-plane hydrogen bonding is most prevalent between atoms in the groups AMIDE-OH1, OH1-OH2, and AMIDE-OH2, as reported in Table 7.4. The RDFs for these pairs are shown in Figure 7.3. The target AMIDE-OH1 RDF from the dehydrated bilayer state, shown in Figure 7.3, shows a high, sharp peak at  $\sim 3.5 \text{ \AA}$  as a result of the significant hydrogen bonding capacity between atoms in these groups. The CG potential captures this peak and the weaker second peak in height and separation, indicating that the headgroup packing is correctly captured with the CG model. The fluid-phase CG AMIDE-OH1 RDF also agrees well with the target RDF, capturing the location of the first peak and the longer-ranged portion of the RDF ( $r > 7.5 \text{ \AA}$ ) but missing some features in the

range  $4.5 \text{ \AA} < r < 7.5 \text{ \AA}$ . Since the primary goal of this CG force field is an accurate structural representation of lamellar structures relevant to the SC lipid matrix, a higher weight was given to the dehydrated bilayer state, and thus again, this small deviation in the fluid-phase behavior is considered acceptable.

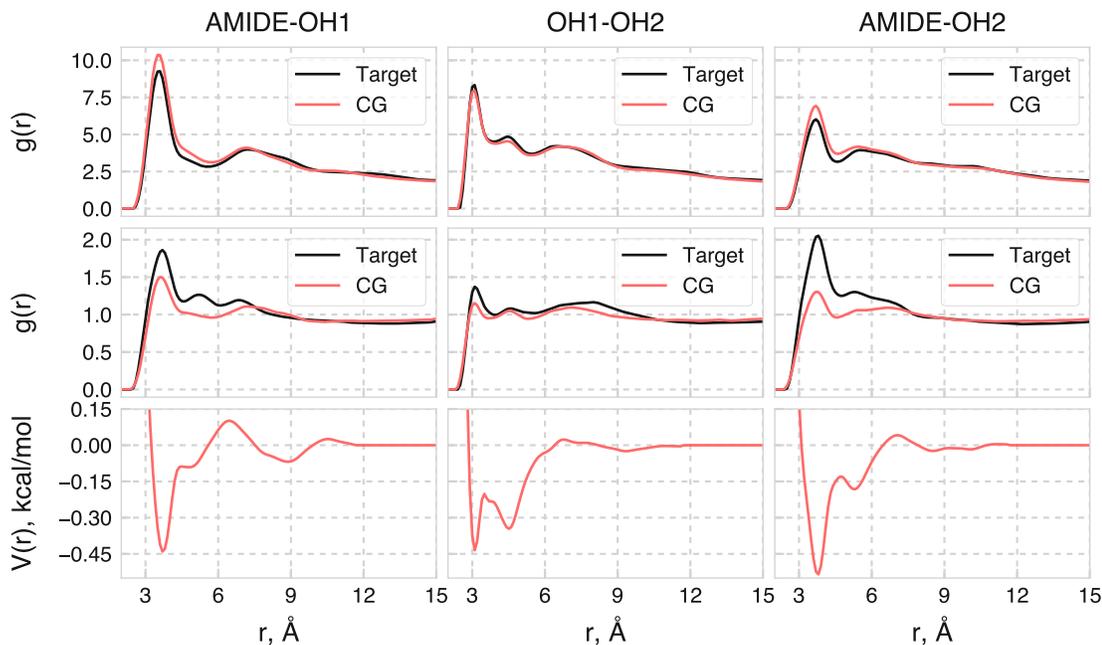


Figure 7.3: RDFs and final derived pair potential between various CERHEAD–CERHEAD beads. Each column corresponds to a particular pair interaction, given at the top of the column. Top row: target (black) and CG (red) RDFs for each pair in the dehydrated NPT bilayer state; middle row, target (black) and CG (red) RDFs for each pair in the bulk NPT fluid pure CER NS state (500 K, 1 atm); bottom row, final derived potential used to calculate the CG RDFs in that column.

As shown in Figure 7.3, the atomistic and CG OH1–OH2 dehydrated bilayer RDFs agree very well, with the three peaks captured in both height and location in the CG simulation. This result indicates that the CG model correctly reproduces the in-plane ordering of the CER headgroups in the dehydrated bilayer. The sharp first peak in the dehydrated bilayer target RDF is a result of the large number of hydrogen bonds that form between atoms in these groups, which we note is reproduced by the CG model. The CG force field also accurately reproduces the fluid RDF, which is expected given

Table 7.4: Summary of the in-plane hydrogen bonding in the atomistic dehydrated CER NS C24 bilayer mapped to the CG level. Note that the acceptor-donor pairs are the CG beads and that a hydrogen bond is considered to exist between CG beads if there is a hydrogen bond between atoms represented by the CG beads. Uncertainties are estimated based on the standard deviation.

Pair	Hydrogen bonds per lipid
OH1-AMIDE	$0.20 \pm 0.04$
OH1-OH2	$0.15 \pm 0.05$
OH2-AMIDE	$0.10 \pm 0.02$
OH1-OH1	$0.03 \pm 0.01$
OH2-OH2	$0.020 \pm 0.001$
AMIDE-AMIDE	$0.017 \pm 0.006$

the similarity between the fluid and dehydrated bilayer OH1-OH2 RDFs. Similarly, the target and CG dehydrated bilayer AMIDE-OH2 RDFs agree with a high level of accuracy, as shown in Figure 7.3. Several hydrogen bonds are also formed between atoms in these groups, which is manifested in the high peak in the dehydrated bilayer RDF at 3.5 Å. This peak is accurately captured in the CG simulation, again further indicating the in-plane packing of the CER headgroups in the dehydrated bilayer RDF is captured in the CG model. The CG model captures the general features of the bulk fluid phase, although quantitative agreement is not seen of the peak height; again, since the primary focus is on the ordered phase, this deviation is considered acceptable.

In general, a good fit is found for all remaining RDFs at all states (see Appendix C), although it is worth noting that interactions involving the AMIDE bead tend to not reproduce the first peak in the fluid phase RDFs as accurately as seen in the other interactions. Whereas the other CERHEADs beads contain one or fewer polar groups, the AMIDE bead contains two polar groups, resulting in a high degree of shape anisotropy in the underlying interactions that is not easily accounted for by a spherically symmetric potential. Although this problem is not unique to MS IBI,<sup>43,60</sup> it

becomes more evident when using a multistate coarse-graining methodology, in which the derived potential must find the best compromise between the fluid structure at the different target states, as dictated by the weight given to each target state. Although the weights clearly have an influence on the fit, when an additional optimization was performed in which higher weights were given to the bulk fluid states (listed as  $\alpha_{s,0}(\text{high bulk})$  in Table 7.2), as expected the fluid phase CG RDFs containing an AMIDE bead showed good agreement with the atomistic RDFs, at the expense of accurate fits in the bilayer states (*e.g.*, see Figure 7.4). Again, since the primary focus is quantitative agreement in the ordered phases, higher weight was given to those state points as compared to the bulk fluids, and thus, the small quantitative deviations in the bulk fluid properties are considered acceptable. Additionally, we note that the density of the pure CER NS bulk fluid NPT state calculated with the CG model,  $0.69 \pm 0.03$  g/mL, only deviates from the atomistic value of  $0.71 \pm 0.01$  g/mL by 4%.

Compared to the four-site model, relatively poor fits are observed for the ordered lamellar phases, as the CG models do not appear to capture the long-range in-plane structure, although we note bulk fluid properties are captured with similar accuracy to the four-site model. RDFs and potentials for the three-bead headgroup CER NS model are provided in Figures C.12 to C.17 of Appendix C.

*Pure CER Validation* As validation of the CG CER interactions beyond RDFs, the structural properties of dehydrated CER NS C24 bilayers simulated using the CG model are compared to those calculated from simulations of the all-atom (AA) model. Table 7.5 summarizes the properties of the dry bilayer calculated with each model in the anisotropic NPT ensemble. The use of this ensemble allows the lengths of the orthogonal simulation box to change independently, allowing for a less biased comparison of the different models, as systems start from preassembled bilayer configurations.

Overall, the CG force field provides accurate predictions of the key structural

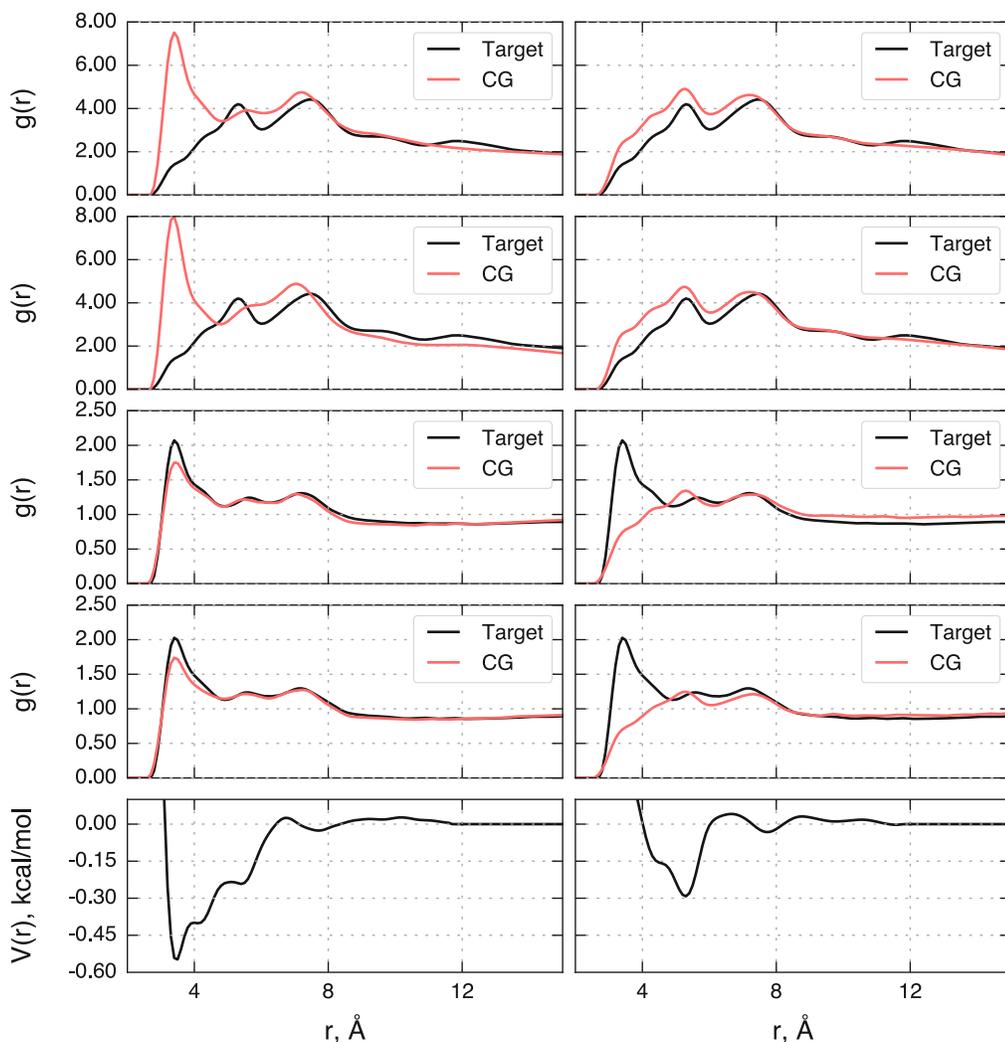


Figure 7.4: AMIDE-AMIDE RDFs and pair potentials from different optimizations to illustrate the effect of the relative weighting factors given to each state. Left: weights listed as  $\alpha_{s,0}$ (high bulk) in Table 7.2; right: weights listed as  $\alpha_{s,0}$  in Table 7.2. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

features of the dehydrated CER bilayer. Only minor variations in APL and  $S_2$  are found between the CG and atomistic simulations, with the CG simulations producing a slightly more compact bilayer (*i.e.*, smaller thickness) than the atomistic simulations. The tilt angle is slightly underpredicted in the CG simulations but is within the error of the atomistic system measurement. When comparing the volume per lipid (VPL) of the bilayers, only minor variations of the ratio of the CG to atomistic results is

Table 7.5: Structural properties of dehydrated CER NS C24 bilayers simulated with CHARMM36-CER all-atom (AA) and the CG models derived here. The four-site model corresponds to Figure 7.1c, while the three-site model corresponds to Figure 7.1d. Uncertainties are estimated based on the standard deviation.

Model	APL ( $\text{\AA}^2$ )	$S_2$	Thickness ( $\text{\AA}$ )	$\theta$ (deg)
AA	$40.0 \pm 0.1$	0.983	$57.1 \pm 0.1$	$12 \pm 5$
Four-site CG	$41.90 \pm 0.07$	0.976	$55.44 \pm 0.07$	$7 \pm 1$
Three-site CG	$48.99 \pm 0.07$	0.956	$58.31 \pm 0.08$	$9.5 \pm 0.5$

observed, with  $VPL_{CG}/VPL_{AA} = 1.02$ . As such, we can conclude that the CG CER force field accurately captures the structure and density of the dehydrated bilayer phase of CER NS. Table 7.5 also shows the properties of the dehydrated bilayer simulations of the three-site headgroup model depicted in Figure 7.1d. Note that the TAIL and TER2 beads are the same for both mappings and identical targets were used for optimizing both the three- and four-site mappings. Thus, differences in properties are a direct result of the chosen mapping. The simulation of the three-site model shows a significant overprediction of the APL as compared to the atomistic simulation. This inaccuracy is likely related to the lack of in-plane ordering in the three-site model, which is apparent in the dry bilayer RDFs, as shown in Figures C.12 to C.17 of Appendix C. The VPL of this model is also overpredicted compared to the atomistic simulations, with  $VPL_{CG}/VPL_{AA} = 1.25$ . It is worth noting that the APL for the three-site CG CER model is comparable to that obtained from the three-site MARTINI-compatible model of Sovová *et al.*,<sup>35</sup> although the lipids in the three-site CG model show appreciable tilt, which is not captured by the MARTINI model. As such, it is clear that the three-site CER headgroup model is an improvement over the MARTINI model, and that the four-site CER headgroup mapping better captures the in-plane packing of the CER headgroups, enabled by the more explicit description of the hydroxyl groups; thus, this treatment is essential to accurately capture the properties of CER NS. This is also observed for the hydrated bilayer systems, as

discussed below. Furthermore, when a higher weight is given to the bulk fluid states in the potential optimization, we find that the APL of the dehydrated bilayer increases significantly to a value of  $49.8 \pm 0.2 \text{ \AA}^2$ . This deviation is a result of capturing more of the bulk fluid behavior than the lamellar behavior in the potentials during the optimization. Thus, the choice of weights (*i.e.*, higher weights given to the lamellar states) appears to provide a good balance between accurately capturing the structure of the ordered lamellar phases and accurate prediction and accurate prediction of the bulk density, even though some of the fluid phase RDFs are slightly shifted compared to the atomistic target data.

### 7.3.2 Optimization of Lipid–Water Interactions

Here, the results of the lipid–water potential derivations are presented, focusing on beads within the CER molecule. Interactions are first derived using MSIBI. However, it is worth noting that system density alone can drive lipid–water structural correlations, even in systems that are hydrophobic; the MS IBI routine (and IBI in general) cannot necessarily differentiate between a correlation that is induced by density and one induced by underlying attraction. Thus, even though MS IBI produces a potential that captures the structural features of the target RDFs, the resulting potential may not provide adequate hydrophobic behavior. To address this issue, MS IBI is coupled with surface wetting simulations, which are used to validate, and, if necessary, tune specific interactions such that the atomistic wetting behavior is observed in the CG model. In this approach, targeted modifications are made to the CG interactions that were derived via MS IBI and that reproduce the correct excluded volume and RDFs, such that the appropriate hydrophobic/hydrophilic character is captured in the surface wetting simulations.

*TAILS–WATER* The TAILS–WATER interaction potentials were first derived using target data from simulations of FFA in water. As before, FFA was used instead

of CER to remove the effect of the strong interactions between the CER headgroups and water; CERHEADs–WATER interactions were optimized separately. The FFA–WATER optimization uses three distinct states, each with a different concentration of FFA in water, such that the different sized FFA aggregates are represented in the target RDFs. The CG simulations start from the final configuration of the atomistic simulations, in which the FFA is phase-separated from the water. To ensure that the CG potential can allow FFA to phase-separate, rather than just stabilize a phase-separated configuration, an additional state with the same concentration and target RDF as the most concentrated system is considered, except that the CG simulation starts with FFA dispersed in water; if FFA and water cannot phase-separate with the given CG potential, the CG RDF will not match the phase-separated target and the CG potential will be further refined by the MS IBI algorithm. Thus, inclusion of this state helps ensure that TAILs are sufficiently hydrophobic to drive phase-separation. While this only determines a minimum level of hydrophobicity needed for phase-separation, as a more hydrophobic system would also phase-separate, inclusion of this state greatly contributes to the realistic behavior of the CG model.

Figure 7.5 shows the RDFs for the most dilute and concentrated FFA–water targets, along with the optimized CG potential. Note that the highest concentration state corresponds to the system that starts from a dispersed state at each stage in the force field optimization. From the figure, we can see that both target RDFs are accurately reproduced by the CG pair potential. It is also apparent that the hydrophobic interactions are sufficient to drive phase-separation, given that the system starts dispersed and visually, the FFA phase-separates from the water, as shown in Figure 7.6a.

To ensure the interactions are sufficiently hydrophobic, a wetting simulation was performed in which a FFA monolayer was initialized with a water droplet on the tail side of the leaflet (as opposed to the headgroup side). The droplet forms a spherical

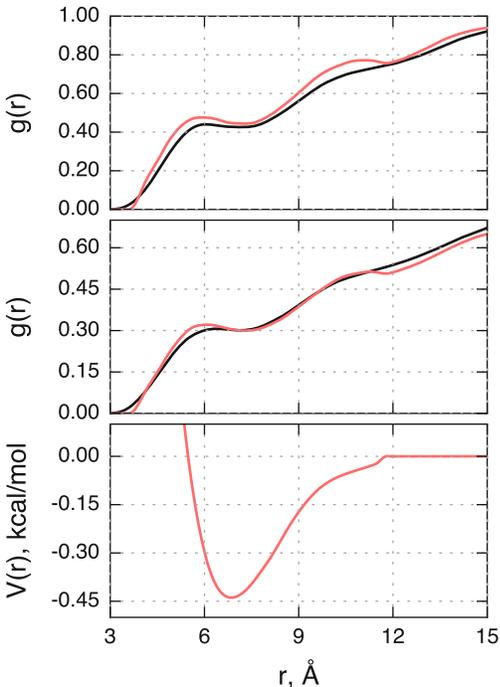


Figure 7.5: TAIL–WATER RDFs, target (black) and CG (red), from the most dilute (top) and most concentrated (middle) states. Bottom: final derived pair potential, before tuning to match the wetting behavior.

cap on the surface with a characteristic contact angle. Snapshots of the droplet from atomistic and CG simulations are shown in Figure 7.7a and b, respectively. From the figure, we can see that the atomistic model shows formation of a droplet with little surface wetting and thus a large contact angle  $\varphi_{AA} = 133^\circ$ , while the contact angle calculated using the MS IBI-derived TAILS–WATER potentials is  $\varphi_{CG} = 40^\circ$ . Note that the WATER–WATER CG potential was derived to capture the stability of a water droplet and provides a close approximation of the atomistic surface tension of a water droplet in a (near) vacuum.<sup>44</sup> As such, the deviation from atomistic behavior lies in the TAILS–WATER interactions, indicating the derived CG force field contains insufficient hydrophobic character, despite the ability to phase-separate, as discussed above. To adjust the hydrophobic character of the TAILS beads, the MS IBI-derived potentials were systematically scaled to be less attractive, with a factor of 0.25 found to provide good agreement with a contact angle of  $\varphi_{CG'} = 135^\circ$  (where the prime

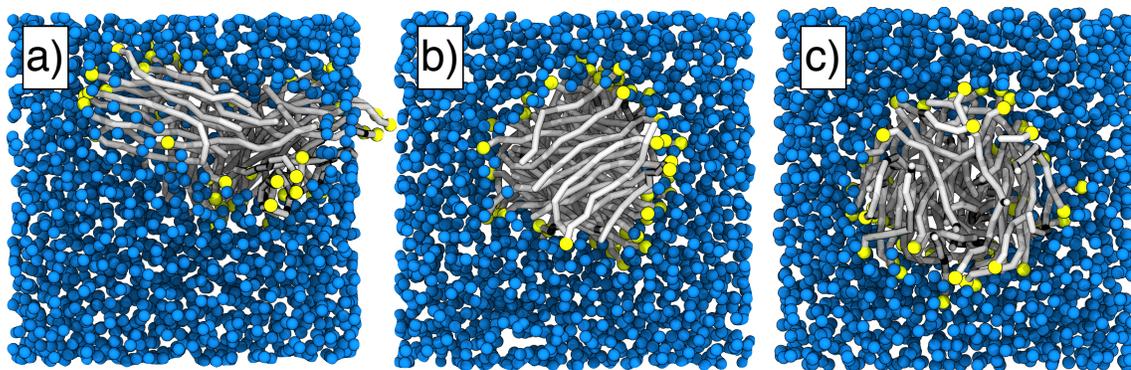


Figure 7.6: Snapshots of the FFA–water simulations using different force fields. (a) CG simulation using the derived force field before tuning the interactions. (b) CG system with potentials after scaling the FHEAD–WATER and TAILS–WATER interactions. (c) Atomistic system mapped to the CG level. CG beads colored as follows: FFA chains, silver with yellow FHEAD beads; water, blue spheres.

indicates use of the scaled potentials), as shown in Figure 7.7c. Examining the behavior of the hydrated monolayer, the atomistic simulations demonstrate a high level of nematic ordering with  $S_{2,AA} = 0.950$  at the droplet contact (where a lipid is considered in contact with the droplet if the terminal bead is within  $7 \text{ \AA}$  of a water bead in the droplet). The unscaled CG potential optimized via MS IBI, which does not capture the contact angle, appears to show a disruption in ordering near the droplet with  $S_{2,CG} = 0.622$ , as compared to  $S_{2,CG'} = 0.921$  with the scaled potential. This result suggests that the lipid hydrophobicity not only affects the lipid–water interface but also plays a role in the structure of the lipid phase, which is relevant to hydrated lipid systems. Simulating the bulk FFA–WATER system with the scaled potentials results in behavior that also better matches that atomistic model, as shown in Figure 7.6b and c, where the simulation using the scaled CG potentials produces a cylindrical FFA aggregate in better agreement with the atomistic model, rather than an elliptical aggregate produced with the unscaled CG potential. By scaling the TAILS–WATER potentials, the CG force field is able to phase-separate, reproduce the atomistic wetting behavior, and maintain the lipid ordering upon wetting. Therefore, the derived TAIL–WATER and TER2–WATER potentials scaled by a factor of 0.25

are transferred to the CER–water optimizations and used henceforth.

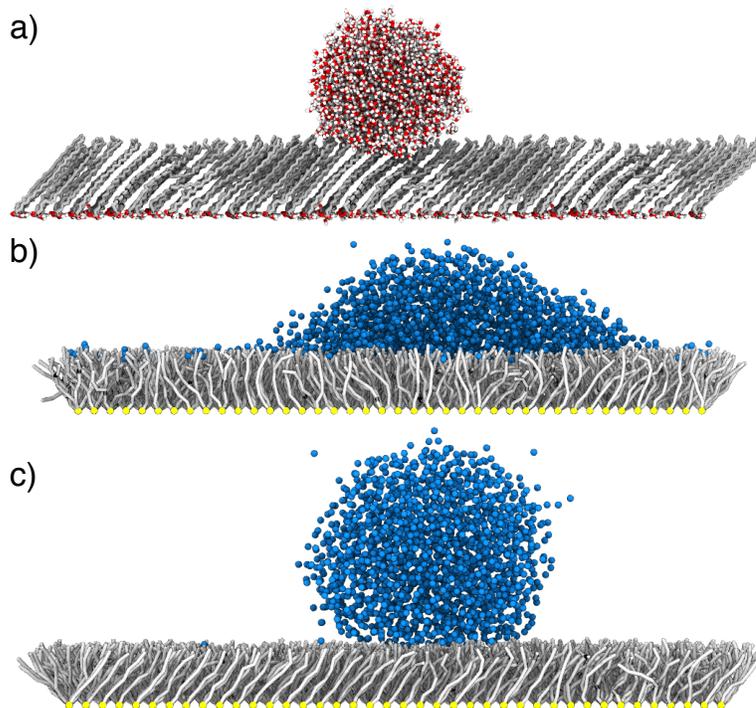


Figure 7.7: Snapshots from the droplet simulations of (a) the atomistic model, (b) the CG model before tuning TAILS–WATER potentials, and (c) the CG model with the TAIL–WATER potentials scaled by 0.25. Atoms are colored as follows: C, silver; O, red; H, white; methyl hydrogens are not shown for clarity. CG beads are colored as follows: TAILS, silver; FHEAD, yellow; WATER, blue.

*CERHEADs–WATER* While hydrophobicity-driven phase-separation is the first step in lipid self-assembly, hydrophilic interactions help drive the formation of organized structures.<sup>33,61</sup> Therefore, it is important that the lipid headgroups possess the correct hydrophilic character to drive self-assembly. The CER–WATER MS IBI optimization used three states to optimize the interactions between water and the CER headgroup beads: hydrated CER NS C24 bilayer (305 K, NVT and NPT) to capture the interfacial behavior and CER NS headgroups dispersed in water (1.0 atm, 400 K) to capture the solvation of the lipid headgroups necessary for bilayer self-assembly. We note that the headgroups in water were simulated at 400 K to ensure that the headgroups were well dispersed throughout the water. Close matches to the target RDFs are found,

as shown (along with the pair potentials) in Figures C.8 to C.11 of Appendix C. To determine if the CERHEADs beads have sufficient hydrophilic character, the wetting properties of a CER–WATER interface was examined. Atomistic wetting simulations of CER headgroups show complete wetting of the headgroups by water, that is,  $\varphi_{AA} \rightarrow 0^\circ$ , indicating a hydrophilic surface. This is expected, given the available hydrogen bonding sites in the CER headgroups. Wetting simulations using the derived CG force field show the same behavior, complete wetting of the surface, indicating the MS IBI optimizations captured the hydrophilic nature of the CER headgroups. As such, the CERHEADs–WATER potentials were not modified from their derived form. It is worth noting that the same procedure was performed for the three-site CER headgroup mapping model; sufficient hydrophilicity of these headgroups beads was observed and hence no scaling was performed.

### 7.3.3 Validation

As validation of the fully derived CG model (*i.e.*, where all interactions are represented), hydrated preassembled single-component bilayers composed of CER NS C16 and CER NS C24 were simulated, and the structural properties compared to corresponding atomistic simulations. Table 7.6 summarized the comparison between the CG and atomistic CHARMM-CER simulation results, where overall good agreement is seen, with the APL, nematic order, and bilayer thickness accurately reproduced. While neither model shows a strong correlation between CER length and tilt angle, appreciably more tilt is observed in the atomistic system than the CG systems. While this and other differences between the atomistic and CG simulations in Table 7.6 can be attributed, at least in part, to the loss of details in the CG model, the softened free energy landscape associated with the CG models should allow the systems to relax more efficiently than atomistic models, even over the same time scale. Furthermore, as discussed in the context of self-assembly below, small deviations in lipid packing

Table 7.6: Structural properties of CER NS C16/CER NS C24 bilayers at 305 K and 1.0 atm simulated using the CG models derived here and the CHARMM36-CER all-atom model (AA). Uncertainties are estimated based on the standard deviation.

System	Model	APL ( $\text{\AA}^2$ )	$S_2$	$d$ ( $\text{\AA}$ )	$\theta$ (deg)
CER NS C16	AA	$42.4 \pm 0.2$	0.998	$42.5 \pm 0.5$	$24.3 \pm 0.9$
	CG	$42.0 \pm 0.1$	0.982	$44.51 \pm 0.09$	$6.3 \pm 0.7$
CER NS C24	AA	$39.9 \pm 0.2$	0.979	$56.2 \pm 0.4$	$22 \pm 4$
	CG	$41.99 \pm 0.09$	0.981	$56.6 \pm 0.1$	$7 \pm 1$
	Three-site CG	$46.2 \pm 0.1$	0.914	$61.5 \pm 0.2$	$9 \pm 2$

in the CG simulations result in measurable differences in the tilt angle, which may also account for the perceived underprediction of tilt angles in the CG simulations. Note that only CER NS C24 was used in the force field derivation; therefore, these results also illustrate the predictive power of the CG model in the ability to accurately capture CER NS C16 behavior. Additionally, comparisons are made for CER NS C24 simulated with the three-site CG model derived here (see Table 7.6), where, again, the APL is overpredicted as compared to the atomistic simulation and is similar to the APL calculated with the MARTINI CER model ( $46 \text{\AA}^2$ ).<sup>35</sup>

As an additional validation, the properties calculated from simulations of the CG and atomistic models for pure systems in Table 7.6 compare favorably with the limited experimental data on CER-based systems. APL values calculated from experimental studies of CER monolayers range from  $37.8 \text{\AA}^2$  to  $45 \text{\AA}^2$  for CER NS C16<sup>62,63</sup> and  $39.0 \text{\AA}^2$  for CER NS C18.<sup>39</sup> Monolayers displaying tilt angles ranging from  $0^\circ$  to  $14^\circ$  for pure CER NS C16 and CER NS C18 have been reported from experimental measurements,<sup>39,63</sup> which are notably closer to the values predicted by the CG models than the atomistic model. Other simulation studies using different atomistic force fields also fall within similar ranges; *e.g.*, using the united atom GROMOS-based force field of Berger,<sup>64</sup> Notman *et al.* calculated an APL of  $37.4 \text{\AA}^2$  for CER NS C24 at 323 K,<sup>21</sup> while a value of  $39.8 \text{\AA}^2$  was reported for pure CER NS C16 at 305 K with the

same model in other work.<sup>18</sup> Additionally, Metcalf and Pandit calculated the APL of a pure CER NS C16 bilayer to be  $\sim 44 \text{ \AA}^2$  at 323 K using a different united atom model.<sup>65</sup> Tilt angles ranging from  $17^\circ$  for pure CER NS C24 at 323 K<sup>21</sup> to  $26^\circ$  for pure CER NS C16 at 323 K<sup>65</sup> have been reported from simulation studies. It is worth noting that the tilt angle in CG simulations significantly deviates from the atomistic simulations; this may be related to the loss of detail in the CG model, in which the local packing of the tails is altered compared to the atomistic model, resulting in different tilt behavior. However, as noted above, the tilt angles calculated in the CG simulations are actually closer to experimental values than the tilt angles from the atomistic simulations. Thus, overall, the four-site CER headgroup CG model developed here produces results for preassembled bilayers consistent with both experiment and various atomistic-level models.

In addition to the aforementioned structural metrics, the orientation of the OH groups between the CG and atomistic models for CER NS C24 systems were compared because of the importance of hydrogen bonding network formation in CER-based systems.<sup>38–40</sup> The orientation of the MHEAD2–OH1 vector relative to the outward bilayer normal (*i.e.*, pointing toward the water) in the pure CER NS C24 bilayer, shown in Figure 7.8, is preferentially aligned  $\sim 60^\circ$  off the bilayer normal in the atomistic simulation, suggesting the OH1 group interacts with the water phase. This behavior is captured with the CG model, where the MHEAD2–OH1 vector prefers an approximately  $50^\circ$  angle with respect to the bilayer normal. In contrast, the MHEAD2–OH2 vector prefers an orientation with the OH2 bead below the MHEAD2 bead (*i.e.*, not between the MHEAD2 and WATER beads), indicating it is more involved in the hydrogen bonding network formed between the CER headgroups compared to the OH1 bead. This behavior is qualitatively captured by the CG model, where the MHEAD2–OH2 vector tends to point away from the lipid–water interface, indicated by the MHEAD2–OH2 angle that is  $< 90^\circ$  in Figure 7.8. However, there is a quantitative

discrepancy between this property in the CG and atomistic simulations, where the CG system prefers an angle of  $\sim 95^\circ$ , compared to  $\sim 135^\circ$  in the atomistic system. This discrepancy does not appear to have any significant impact on the APL or in-plane structure of the lipids, and is likely a consequence of the loss of detail inherent to CG models. This comparison further suggests that the CG model, at least on a qualitative level, captures the hydrogen bond network formation between the CER headgroups.

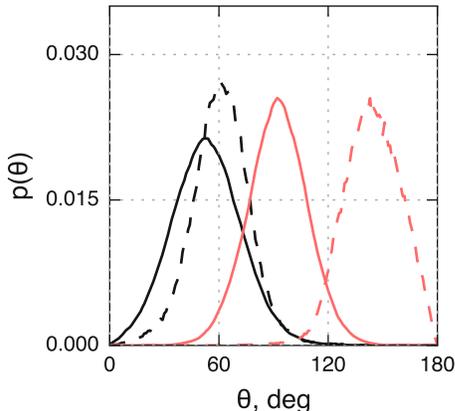


Figure 7.8: Orientation of the MHEAD2–OH vectors with respect to the bilayer normal in the CER NS C24 bilayer. A value of  $0^\circ$  indicates the OH bead lies directly between the MHEAD2 bead and the lipid–water interface; a value of  $90^\circ$  means the OH group is lying flat in the bilayer plane. Black lines, OH1; red lines, OH2; solid lines, CG; dashed lines, atomistic mapped to the CG level.

#### 7.4 CER NS C24 Self-Assembly

It is unlikely that preassembled atomistic systems would be able to transition from metastable states to lower energy states because of the time scales on which major morphological transitions occur compared to the time scales that are accessible with atomistic models. Thus, pre- and self-assembled structures studied via atomistic simulations are likely to be “stuck” in metastable configurations. For example, Das *et al.* studied SC lipid self-assembly using a united-atom force field, finding that the lipids self-assemble into a lamellar phase under confinement, but they form micellar phases

in bulk systems,<sup>29</sup> in contrast to experimental studies that suggest the formation of lamellar phases in the bulk,<sup>66,67</sup> this discrepancy is likely due to the need for much longer simulation times to form equilibrium phases, currently inaccessible to atomistic models. Given the reduced computational cost of CG models, they can be used to access the time scales required to examine the self-assembly of SC lipids.

Here, the self-assembly of CER NS C24 in water is examined, motivated by its abundance in the SC. Systems composed of 800 lipids and 10 000 water beads are studied, which represents a water to lipid ratio well above that of a fully hydrated bilayer. To remove dependence on the initial configuration, the CG-mapped atomistic bilayer (replicated twice in the  $x$  and  $y$  directions to 4 times the original size) system was heated to 1500 K for 20 ns, during which time the lipids and water became completely mixed and disordered, as shown in Figure 7.9a. Upon cooling to 305 K, a bilayer containing structural defects formed, as shown in Figure 7.9b. To increase the relaxation rate of the defects, the system was then expanded in the bilayer plane to double the original area over 10 ns while preserving the box volume, as shown in Figure 7.9c. Upon deforming the box to the original aspect ratio over 10 ns, a defect-free bilayer formed, as shown in Figure 7.9d. After switching to the anisotropic NPT ensemble, the self-assembled bilayer was found to be stable for >200 ns. This process was repeated three times with different expansion and compression rates to examine the influence of the relaxation rate on the final structure. The APLs of the bilayers over the final 50 ns of the NPT simulations are compared in Table 7.7, with little variation between the different systems observed. Interestingly, the self-assembled CG systems are found to pack more densely than the comparable preassembled bilayers, resulting in APLs that are closer to the value from the atomistic simulation (see Table 7.6, CER NS C24). The time scale associated with this relaxation influences the distribution of lipids between the leaflets in the bilayer structure, with shorter time scales associated with larger deviations from an ideal (*i.e.*, equal) distribution,

Table 7.7: Properties of self-assembled bilayers as a function of expansion and compression time (relaxation time).  $f_{\text{top}}/f_{\text{bottom}}$  represents the fraction of lipids in the top and bottom leaflets of the self-assembled bilayers;  $\theta_{\text{top}}/\theta_{\text{bottom}}$  represents the average tilt angle of the lipids in the top and bottom leaflets. For the stacked systems,  $f_{\text{top}}/f_{\text{bottom}}$  represents the fraction of lipids in each of the four layers, and  $\theta_{\text{top}}/\theta_{\text{bottom}}$  represents the average tilt angle of the lipid tails in each of the four layers. The APL of the stacked systems was calculated using the average number of lipids in each layer. The standard error in the tilt angle was less than  $3^\circ$  for each measurement.

Final Structure	$\tau_{\text{R}}$ , ns	APL, $\text{\AA}^2$	$f_{\text{top}}/f_{\text{bottom}}$	$\theta_{\text{top}}/\theta_{\text{bottom}}$
Bilayer	20	$40.00 \pm 0.02$	0.513/0.487	5.2/9.2
	40	$40.10 \pm 0.02$	0.511/0.489	5.0/7.2
	60	$39.69 \pm 0.03$	0.509/0.491	5.2/9.0
Stacked Bilayer	200	$42.76 \pm 0.02$	0.243/0.250/0.244/0.264	9.6/6.6/8.1/4.9
	400	$43.09 \pm 0.03$	0.240/0.255/0.253/0.253	9.5/5.5/6.3/6.1
	600	$42.83 \pm 0.02$	0.258/0.246/0.253/0.244	5.6/7.1/5.9/11.1

although all deviations are within 20% of ideal. These small variations, however, do create noticeable changes to the average tilt angle of the leaflets (although the average APL is relatively unaffected), with a larger tilt observed in the leaflets with fewer lipids, as shown in Table 7.7. This trend is expected—with a fewer number of lipids in a leaflet, the lipid tails tilt relative to the bilayer normal to decrease their spacing to an optimal level.

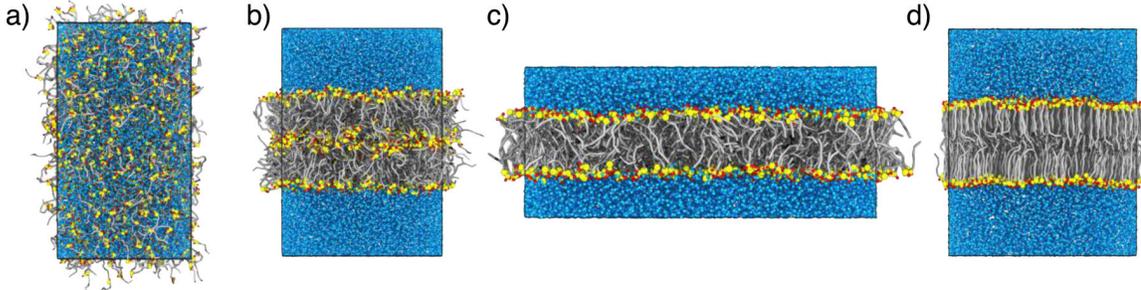


Figure 7.9: Self-assembly of CER NS C24 into an ordered bilayer: (a) initial heating above the mixing temperature; (b) cooling from high temperature; (c) full compression; (d) final configuration after compression and expansion. CG beads are colored as follows: TAILS, silver cylinders; AMIDE and MHEAD2, yellow spheres; OH1 and OH2, red spheres; WATER, blue sphere.

Since the simulated bilayer is a periodic structure (as periodic boundary conditions

are used), it is reasonable to assume that the self-assembled structure is dependent on the size and shape of the box. For example, one would expect two bilayers to form in the previously described CER NS C24 system if the box area parallel to the bilayer plane were reduced by a factor of 2. To explore this concept, the previous self-assembly protocol was repeated, with the exception that the aspect ratio of the simulation box at the end of the initial compression phase was set to half that of the initial self-assembly simulation (*i.e.*, half the cross-sectional area of the bilayer and twice the thickness). As with the bilayer self-assembly, a defected lamellar structure initially formed, where the lipids were well-ordered at the lipid–water interface and less ordered in the interior of the membrane, as shown in Figure 7.10a. After subsequent expansion and compression cycles, well-ordered, multilamellar structures (*i.e.*, stacked bilayers) formed, with a small amount of water between the middle leaflets (0.35 to 0.5 water *molecules* per interior lipid, compared to 50 in the bulk), as shown in Figure 7.10b. The relaxation rates of the stacked bilayer defects were much slower than that of the single bilayer systems, with 200 ns of both expansion and compression required to remove the defects, compared to 20 ns for the single bilayer assembly. Subsequent replication of this self-assembly protocol yielded similar structures, illustrating reproducibility. Without the expansion and compression to increase the relaxation rate of the defects, a defect-free stacked bilayer structure formed over the course of 1.3  $\mu\text{s}$  at 305 K. Thus, the expansion and compression protocol is used as a means to increase the rate of exploration of phase space of the CER–water system, rather than to influence the final state of the system.

Examining the multilayer systems, it is interesting to note that similar structural trends are observed in comparison to the self-assembled single bilayer systems. For example, the self-assembled multilayer systems also show deviations from an ideal distribution of lipids per leaflet, although the deviations are larger in the multilamellar systems (see Table 7.7), which may be expected given the differences between a

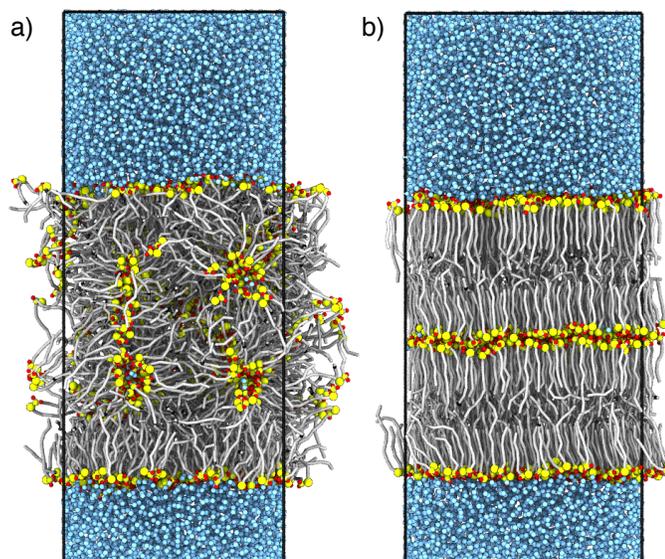


Figure 7.10: Self-assembly of CER NS C24 into a stacked bilayer: (a) after changing the aspect ratio of the simulation cell; (b) after subsequent area expansion and compression. Beads colored as follows: TAILS, silver cylinders; AMIDE and MHEAD2, yellow spheres, OH1 and OH2, red spheres; WATER, blue spheres.

lipid–water and a lipid–lipid interface. A similar relationship between the tilt angle of the tails in a leaflet and the number of lipids in the leaflet is also observed for the single bilayer in multilayer systems, where a higher tilt angle is observed for leaflets with fewer lipids. Compared to the self-assembled single bilayers, the APL of the self-assembled multilayer systems is slightly larger and closer to the preassembled CG bilayers, which again may be related to differences manifesting from the multilamellar organization.

The lipid–lipid interface in multilayer structures makes it possible for lipids to adopt different configurations, which may alter the properties of the system. For example, several models of SC lipid organization have been proposed, with a key difference between the models being whether the CER tails adopt an extended conformation (with the two tails on opposite sides of the headgroup), a hairpin conformation (with both tails on the same side of the headgroup), or some combination of the two. For example, the stacked monolayer model assumes an extended conformation for the CERs,<sup>68</sup> while the sandwich model assumes a hairpin conformation of the CERs.<sup>6,69</sup>

In contrast, the single gel phase model proposes a mixture of extended and hairpin CER conformations,<sup>70</sup> and the armature reinforcement model proposes a transition from extended to hairpin CER conformations upon hydration.<sup>67,71</sup>

To explore the CER conformations in the stacked bilayer configurations, the angle  $\theta_{\text{tails}}$  between the two tails of each CER molecule was examined.  $\theta_{\text{tails}}$  was defined as the angle between the two vectors from the MHEAD2 bead to the terminal bead of each tail and thus distinguishes between the lipid being in an extended ( $\theta_{\text{tails}} > 90^\circ$ ) or hairpin ( $\theta_{\text{tails}} < 90^\circ$ ) conformation. From the simulations, we find that the vast majority of the lipids in the multilamellar system exist in a hairpin conformation, with no correlation between conformation and expansion/compression time (for all systems, 91% to 94% of the lipids exist in a hairpin conformation). As might be expected, the hairpin configuration is strongly preferred at the lipid–water interface due to the hydrophobic nature of the lipid tails; however, when considering the lipids in the inner leaflets, the number of lipids adopting a hairpin conformation decreases to 82% to 87%. Thus, while the hairpin conformation appears to be the preferred CER conformation in the pure, multilayer CER NS C24 system, roughly 15% of the lipids in the inner leaflets adopt an extended conformation. Thus, the pure CER NS system is most consistent with the hairpin conformation assumed in the sandwich model. However, it should be noted that the CER conformation in the complex lipid mixtures of human SC, which include both cholesterol and FFAs, could be different than those observed for pure CER NS. It is reasonable to assume that the intramolecular bonded forces dictate the CER conformation to an extent. Therefore, to validate the intramolecular potentials and ensure they are not biasing the extended *versus* hairpin conformations in the stacked bilayer system, a single CER molecule in a vacuum was simulated with the CG and atomistic force fields, and the distributions of the angle between the two tails ( $\theta_{\text{tails}}$ ) and end-to-end distance of the tails calculated with each model. Good agreement between the distributions calculated with the two models was found,

with each predicting, on average, a hairpin conformation of the two tails, as shown in Figure 7.11. Therefore, the observed hairpin *versus* extended distribution is physically relevant and not an artifact of the intramolecular bonded forces in the CG force field.

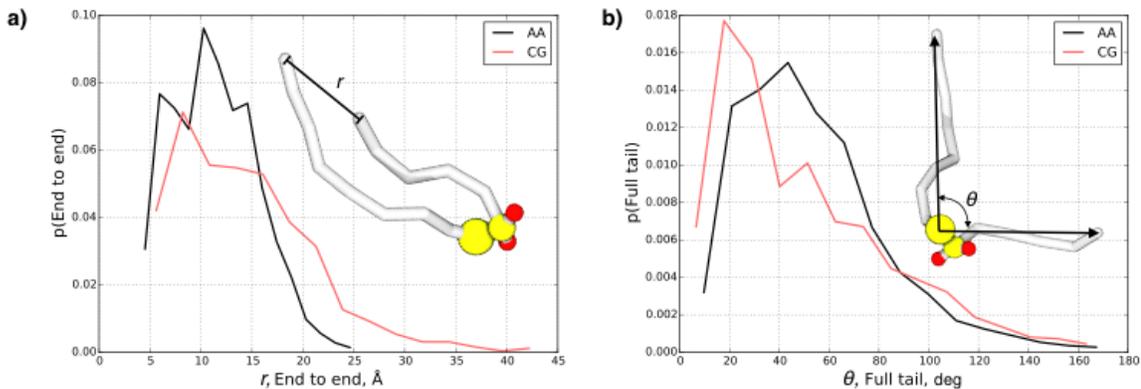


Figure 7.11: CER NS conformations from simulations of a single molecule in vacuum. (a) end-to-end distance between the two tails; (b) angle between each of the two ceramide tails.

## 7.5 Conclusion

This chapter shows that MS IBI is an appropriate method for deriving CG models of SC lipids, specifically CER NS, which is the most abundant CER in human SC. The transferable nature of the MS IBI-derived force field is illustrated by the fact that the TAIL beads of CER NS and FFA are treated as the same type, and accurately model the tail behavior of the two different types of molecule. This transferability is promising for the development of models for other SC lipids, as the TAILS beads can confidently be transferred to describe the acyl tails, leaving only the headgroup interactions to be derived. Further, the inclusion of explicit hydroxyl groups on the CER headgroups allows the model to accurately capture the in-plane packing of the molecules. This property is important for any CER model, especially, for example, when studying permeability, where the primary barrier is at the lipid–water interface.<sup>72</sup> In contrast, an implicit treatment of the hydroxyl groups does not allow the model to

capture this behavior, which results in less accurate structural properties of lamellar phases with the CG model. This CG mapping also allows for easy extension to other CER species, since the AMIDE and MHEAD2 beads are common to all CERs.

The derived CG model was first validated through simulations of preassembled CER NS bilayers. Good agreement on several important bilayer properties was found between the CG and fully atomistic simulations and experimental data where available. Through preassembled bilayer simulations, we further showed that the CG mapping involving explicit hydroxyl groups more accurately reproduces the bilayer properties, and that this increased integrity of the model incurs a minimal increase in computational expense. After validation on preassembled structures, self-assembly of CER NS into bi- and multilayer structures was studied with the CG model, with the self-assembled structures exhibiting similar properties to the preassembled systems. Simulations of the CG model show a predominance of CERs adopting a hairpin configuration in the multilayer systems.

While this chapter presents the derivation of a model specific to CER NS, it also provides the necessary framework for developing force fields to accurately model a realistic SC lipid mixture. Recent experimental work has focused on developing lipid mixtures that adequately reproduce the barrier properties of human skin.<sup>37,67,73-75</sup> CG SC lipid models can offer molecular-level insight into phenomena that cannot be observed through experiment and atomistic molecular simulation *e.g.*, the self-assembly into and transitions between well-ordered lamellar phases. CG models are thus a practical tool for complementing experimental work in this field, *e.g.*, aiding in the discovery of promising lipid mixtures that reproduce the properties of human SC. CG models for the other lipid classes of SC (FFA and cholesterol) are presented in the following chapters, allowing simulation studies of representative SC lipid mixtures.

## 7.6 Bibliography

- [1] MOORE, T. C., IACOVELLA, C. R., HARTKAMP, R., BUNGE, A. L., AND M<sup>c</sup>CABE, C. A coarse-grained model of stratum corneum lipids: free fatty acids and ceramide NS. *The Journal of Physical Chemistry B* **120**, 2016, pp. 9944–9958.
- [2] YARDLEY, H. J. AND SUMMERLY, R. Lipid composition and metabolism in normal and diseased epidermis. *Pharmacology & Therapeutics* **13**, 1981, pp. 357–383.
- [3] GRAY, G. M. AND YARDLEY, H. J. Different populations of pig epidermal cells: isolation and lipid composition. *Journal of Lipid Research* **16**, 1975, pp. 441–447.
- [4] RABIONET, M., GORGAS, K., AND SANDHOFF, R. Ceramide synthesis in the epidermis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 2014, pp. 422–434.
- [5] PONEC, M., WEERHEIM, A., LANKHORST, P., AND WERTZ, P. W. New acylceramide in native and reconstructed epidermis. *Journal of Investigative Dermatology* **120**, 2003, pp. 581–588.
- [6] BOUWSTRA, J. A., DUBBELAAR, F. E. R., GOORIS, G. S., AND PONEC, M. The lipid organisation in the skin barrier. *Acta Dermato-Venereologica; Supplement* **208**, 2000, pp. 23–30.
- [7] VAN SMEDEN, J., JANSSENS, M., GOORIS, G. S., AND BOUWSTRA, J. A. The important role of stratum corneum lipids for the cutaneous barrier function. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* **1841**, 2014, pp. 295–313.
- [8] VAN SMEDEN, J., JANSSENS, M., KAYE, E. C. J., CASPERS, P. J., LAVRIJSEN, A. P., VREEKEN, R. J., AND BOUWSTRA, J. A. The importance of free fatty acid chain length for the skin barrier function in atopic eczema patients. *Experimental Dermatology* **23**, 2014, pp. 45–52.
- [9] FELLER, S. E. Molecular dynamics simulations of lipid bilayers. *Current Opinion in Colloid and Interface Science* **5**, 2000, pp. 217–223.
- [10] LINDAHL, E. AND SANSOM, M. S. P. Membrane proteins: molecular dynamics simulations. *Current Opinion in Structural Biology* **18**, 2008, pp. 425–431.
- [11] LYUBARTSEV, A. P. AND RABINOVICH, A. L. Recent development in computer simulations of lipid bilayers. *Soft Matter* **7**, 2011, pp. 25–39.
- [12] BENNETT, W. F. D. AND TIELEMAN, D. P. Computer simulations of lipid membrane domains. *Biochimica et Biophysica Acta - Biomembranes* **1828**, 2013, pp. 1765–1776.

- [13] RABINOVICH, A. L. AND LYUBARTSEV, A. P. Computer simulation of lipid membranes: methodology and achievements. *Polymer Science Series C* **55**, 2013, pp. 162–180.
- [14] BEMPORAD, D., LUTTMANN, C., AND ESSEX, J. W. Computer simulation of small molecule permeation across a lipid bilayer: dependence on bilayer properties and solute volume, size, and cross-sectional area. *Biophysical Journal* **87**, 2004, pp. 1–13.
- [15] GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., VAN MECHELEN, J. B., DEMÉ, B., AND BOUWSTRA, J. A. Disposition of ceramide in model lipid membranes determined by neutron diffraction. *Biophysical Journal* **100**, 2011, pp. 1481–1489.
- [16] ŠKOLOVÁ, B., HUDSKÁ, K., PULLMANOVÁ, P., KOVÁČIK, A., PALÁT, K., ROH, J., FLEDDERMAN, J., ESTRELA-LOPIS, I., AND VÁVROVÁ, K. Different phase behavior and packing of ceramides with long (C16) and very long (C24) acyls in model membranes: infrared spectroscopy using deuterated lipids. *The Journal of Physical Chemistry B* **118**, 2014, pp. 10460–10470.
- [17] MOJUMDAR, E. H., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., AND BOUWSTRA, J. A. Skin lipids: localization of ceramide and fatty acid in the unit cell of the long periodicity phase. *Biophysical Journal* **108**, 2015, pp. 2670–2679.
- [18] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>c</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [19] PANDIT, S. A. AND SCOTT, H. L. Molecular-dynamics simulation of a ceramide bilayer. *The Journal of Chemical Physics* **124**, 2006, p. 014708.
- [20] DAS, C., NORO, M. G., AND OLMSTED, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **97**, 2009, pp. 1941–1951.
- [21] NOTMAN, R., DEN OTTER, W. K., NORO, M. G., BRIELS, W. J., AND ANWAR, J. The permeability enhancing mechanism of DMSO in ceramide bilayers simulated by molecular dynamics. *Biophysical Journal* **93**, 2007, pp. 2056–2068.
- [22] GUPTA, R. AND RAI, B. Molecular dynamics simulation study of skin lipids: effects of the molar ratio of individual components over a wide temperature range. *The Journal of Physical Chemistry B* **119**, 2015, pp. 11643–11655.
- [23] SHAH, J., ATIENZA, J. M., DUCLOS, R. I., RAWLINGS, A. V., DONG, Z., AND SHIPLEY, G. G. Structural and thermotropic properties of synthetic C16:0 (palmitoyl) ceramide: effect of hydration. *Journal of Lipid Research* **36**, 1995, pp. 1936–1944.

- [24] BOUWSTRA, J. A., GOORIS, G. S., CHENG, K., WEERHEIM, A., BRAS, W., AND PONEC, M. Phase behavior of isolated skin lipids. *Journal of Lipid Research* **37**, 1996, pp. 999–1011.
- [25] MOORE, D. J., REREK, M. E., AND MENDELSON, R. FTIR spectroscopy studies of the conformational order and phase behavior of ceramides. *The Journal of Physical Chemistry B* **101**, 1997, pp. 8933–8940.
- [26] BRIEF, E., KWAK, S., CHENG, J. T. J., KITSON, N., THEWALT, J., AND LAFLEUR, M. Phase behavior of an equimolar mixture of N-palmitoyl-D-erythro-sphingosine, cholesterol, and palmitic acid, a mixture with optimized hydrophobic matching. *Langmuir* **25**, 2009, pp. 7523–7532.
- [27] PAZ RAMOS, A. AND LAFLEUR, M. Chain length of free fatty acids influences the phase behavior of stratum corneum model membranes. *Langmuir* **31**, 2015, pp. 11621–11629.
- [28] HOOPE, M. I., NORO, M. G., LONGO, M. L., AND FALLER, R. Bilayer structure and lipid dynamics in a model stratum corneum with oleic acid. *The Journal of Physical Chemistry B* **115**, 2011, pp. 3164–3171.
- [29] DAS, C., NORO, M. G., AND OLMSTED, P. D. Lamellar and inverse micellar structures of skin lipids: Effect of templating. *Physical Review Letters* **111**, 2013, p. 148101.
- [30] MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P., AND DE VRIES, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **111**, 2007, pp. 7812–7824.
- [31] SHINODA, W., DEVANE, R., AND KLEIN, M. L. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation* **33**, 2007, pp. 27–36.
- [32] IZVEKOV, S. AND VOTH, G. A. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B* **109**, 2005, pp. 2469–2473.
- [33] HADLEY, K. R. AND McCABE, C. A simulation study of the self-assembly of coarse-grained skin lipids. *Soft Matter* **8**, 2012, pp. 4802–4814.
- [34] NOTMAN, R., ANWAR, J., BRIELS, W. J., NORO, M. G., AND DEN OTTER, W. K. Simulations of skin barrier function: free energies of hydrophobic and hydrophilic transmembrane pores in ceramide bilayers. *Biophysical Journal* **95**, 2008, pp. 4763–4771.
- [35] SOVOVÁ, Ž., BERKA, K., OTYEPKA, M., AND JUREČKA, P. Coarse-grain simulations of skin ceramide NS with newly derived parameters clarify structure of melted phase. *The Journal of Physical Chemistry B* **119**, 2015, pp. 3988–3998.

- [36] MOORE, D. J., REREK, M. E., AND MENDELSON, R. Role of ceramides 2 and 5 in the structure of the stratum corneum lipid barrier. *International Journal of Cosmetic Science* **21**, 1999, pp. 353–368.
- [37] GARIDEL, P., FÖLTING, B., SCHALLER, I., AND KERH, A. The microstructure of the stratum corneum lipid barrier: mid-infrared spectroscopic studies of hydrated ceramide:palmitic acid:cholesterol model systems. *Biophysical Chemistry* **150**, 2010, pp. 144–156.
- [38] PASCHER, I. Molecular arrangements in sphingolipids conformation and hydrogen bonding of ceramide and their implication on membrane stability and permeability. *Biochimica et Biophysica Acta - Biomembranes* **455**, 1976, pp. 433–451.
- [39] LÖFGREN, H. AND PASCHER, I. Molecular arrangements of sphingolipids. the monolayer behaviour of ceramides. *Chemistry and Physics of Lipids* **20**, 1977, pp. 273–284.
- [40] REREK, M. E., CHEN, H.-C., MARKOVIC, B., VANWYCK, D., GARIDEL, P., MENDELSON, R., AND MOORE, D. J. Phytosphingosine and sphingosine ceramide headgroup hydrogen bonding: structural insights through thermotropic hydrogen/deuterium exchange. *The Journal of Physical Chemistry B* **105**, 2001, pp. 9355–9362.
- [41] MOORE, T. C., IACOVELLA, C. R., AND M<sup>C</sup>CABE, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140**, 2014, p. 224104.
- [42] BAYRAMOGLU, B. AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 2012, pp. 9205–9219.
- [43] BAYRAMOGLU, B. AND FALLER, R. Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion. *Macromolecules* **46**, 2013, pp. 7957–7976.
- [44] MOORE, T. C., IACOVELLA, C. R., AND M<sup>C</sup>CABE, C. Development of a coarse-grained water forcefield via multistate iterative Boltzmann inversion. In *Foundations of Molecular Modeling and Simulation*. Springer (2016), pp. 37–52.
- [45] KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O’CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D., AND PASTOR, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 2010, pp. 7830–7843.
- [46] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 1983, pp. 926–935.

- [47] PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* **117**, 1995, pp. 1–19.
- [48] URL <https://www.lammps.org>.
- [49] PLIMPTON, S. J., POLLOCK, R., AND STEVENS, M. Particle-mesh ewald and rRESPA for parallel molecular dynamics simulations. In *PPSC* (1997).
- [50] MARTYNA, G. J., TOBIAS, D. J., AND KLEIN, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 1994, pp. 4177–4189.
- [51] ANDERSON, J. A., LORENZ, C. D., AND TRAVESSET, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 2008, pp. 5342–5359.
- [52] GLASER, J., NGUYEN, T. D., ANDERSON, J. A., LUI, P., SPIGA, F., MILLAN, J. A., MORSE, D. C., AND GLOTZER, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **192**, 2015, pp. 97–107.
- [53] HADLEY, K. R. AND M<sup>c</sup>CABE, C. A coarse-grained model for amorphous and crystalline fatty acids. *The Journal of Chemical Physics* **132**, 2010, p. 134505.
- [54] HADLEY, K. R. AND M<sup>c</sup>CABE, C. On the investigation of coarse-grained models for water: balancing computational efficiency and the retention of structural properties. *The Journal of Physical Chemistry B* **114**, 2010, pp. 4590–4599.
- [55] HADLEY, K. R. AND M<sup>c</sup>CABE, C. A structurally relevant coarse-grained model for cholesterol. *Biophysical Journal* **99**, 2010, pp. 2896–2905.
- [56] MILANO, G., GOUDEAU, S., AND MÜLLER-PLATHE, F. Multicentered Gaussian-based potentials for coarse-grained polymer simulations: linking atomistic and mesoscopic scales. *Journal of Polymer Science Part B: Polymer Physics* **43**, 2005, pp. 871–885.
- [57] A Git repository for this package is hosted at <https://github.com/mosef-hub/msibi>.
- [58] MCGIBBON, R. T., BEAUCHAMP, K. A., HARRIGAN, M. P., KLEIN, C., SWAILS, J. M., HERNÁNDEZ, C. X., SCHWANTES, C. R., WANG, L.-P., LANE, T. J., AND PANDE, V. S. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal* **109**, 2015, pp. 1528 – 1532.
- [59] A Git repository for this package is hosted at <http://mdtraj.org/latest>.
- [60] FALLER, R. AND REITH, D. Properties of poly(isoprene): model building in the melt and in solution. *Macromolecules* **36**, 2003, pp. 5406–5414.

- [61] MARRINK, S. J., LINDAHL, E., EDHOLM, O., AND MARK, A. E. Simulation of the spontaneous aggregation of phospholipids into bilayers. *Journal of the American Chemical Society* **123**, 2001, pp. 8638–8639.
- [62] BROCKMAN, H. L., MOMSEN, M. M., BROWN, R. E., HE, L., CHUN, J., BYUN, H.-S., AND BITTMAN, R. The 4, 5-double bond of ceramide regulates its dipole potential, elastic properties, and packing behavior. *Biophysical Journal* **87**, 2004, pp. 1722–1731.
- [63] SCHEFFER, L., SOLOMONOV, I., WEYGAND, M. J., KJAER, K., LEISEROWITZ, L., AND ADDADI, L. Structure of cholesterol/ceramide monolayer mixtures: implications to the molecular organization of lipid rafts. *Biophysical Journal* **88**, 2005, pp. 3381–3391.
- [64] BERGER, O., EDHOLM, O., AND JÄHNIG, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **72**, 1997, pp. 2002–2013.
- [65] METCALF, R. AND PANDIT, S. A. Mixing properties of sphingomyelin ceramide bilayers: a simulation study. *The Journal of Physical Chemistry B* **116**, 2012, pp. 4500–4509.
- [66] WARTEWIG, S. AND NEUBERT, R. H. H. Properties of ceramides and their impact on the stratum corneum structure: a review. Part 1: ceramides. *Skin Pharmacology and Physiology* **20**, 2007, pp. 220–229.
- [67] KESSNER, D., RUETTINGER, A., KISELEV, M. A., WARTEWIG, S., AND NEUBERT, R. H. H. Properties of ceramides and their impact on the stratum corneum structure. Part 2: stratum corneum lipid model systems. *Skin Pharmacology and Physiology* **21**, 2008, pp. 58–74.
- [68] SWARTZENDRUBER, D. C., WERTZ, P. W., KITKO, D. J., MADISON, K. C., AND DOWNING, D. T. Molecular models of the intercellular lipid lamellae in mammalian stratum corneum. *Journal of Investigative Dermatology* **92**, 1989, pp. 251–257.
- [69] KUEMPEL, D., SWARTZENDRUBER, D. C., SQUIER, C. A., AND WERTZ, P. W. In vitro reconstitution of stratum corneum lipid lamellae. *Biochimica et Biophysica Acta - Biomembranes* **1372**, 1998, pp. 135–140.
- [70] NORLÉN, L. Skin barrier structure and function: the single gel phase model. *Journal of Investigative Dermatology* **117**, 2001, pp. 830–836.
- [71] KISELEV, M. A. Conformation of ceramide 6 molecules and chain-flip transitions in the lipid matrix of the outermost layer of mammalian skin, the stratum corneum. *Crystallography Reports* **52**, 2007, pp. 525–528.

- [72] DAS, C., OLMSTED, P. D., AND NORO, M. G. Water permeation through stratum corneum lipid bilayers from atomistic simulations. *Soft Matter* **5**, 2009, pp. 4549–4555.
- [73] DE JAGER, M. W., GOORIS, G. S., DOLBANYA, I. P., BRAS, W., PONEC, M., AND BOUWSTRA, J. A. Novel lipid mixtures based on synthetic ceramides reproduce the unique stratum corneum lipid organization. *Journal of Lipid Research* **45**, 2004, pp. 923–932.
- [74] REREK, M. E., VANWYCK, D., MENDELSON, R., AND MOORE, D. J. FTIR spectroscopic studies of lipid dynamics in phytosphingosine ceramide models of the stratum corneum lipid matrix. *Chemistry and Physics of Lipids* **134**, 2005, pp. 51–58.
- [75] GOORIS, G. S. AND BOUWSTRA, J. A. Infrared spectroscopic study of stratum corneum model membranes prepared from human ceramides, cholesterol, and fatty acids. *Biophysical Journal* **92**, 2007, pp. 2785–2795.

## CHAPTER 8

### DEVELOPMENT OF THE CROSS-INTERACTIONS FOR COARSE-GRAINED MODELS OF CER NS AND FFA

This chapter is reproduced in part from Moore *et al.* 2017.<sup>1</sup>

#### 8.1 Introduction

Skin plays the essential role in human physiology of preventing water loss and protecting the body from external physical, chemical, and biological attack. This barrier function is known to be localized to the stratum corneum (SC), the outer most layer of skin.<sup>2</sup> The SC is composed of dead skin cells surrounded by a dense lamellar-structured lipid matrix containing an approximately equimolar mixture of ceramides (CERs), free fatty acids (FFAs), and cholesterol.<sup>3</sup> This lipid mixture notably lacks phospholipids, which are present in most biological membranes. Since the lipid matrix is the only continuous path through the SC, it is generally recognized that the lipids play an integral role in the barrier properties of skin. Anomalies in lipid composition are known to be associated with impaired barrier function, (see, for example, Refs. 4–7); however, while many global properties of SC lipid systems can be gleaned from experimental measurements, molecular-level details often remain elusive. Hence, atomistic molecular dynamics simulations have begun to play an important role in examining the molecular-level structure and composition-related changes to membrane structure and properties<sup>8–16</sup> typically focusing on systems based on CER NS (C18 sphingosine base linked to a non-alpha hydroxy acyl chain). Simulation studies have thus far provided considerable insight into the role of CER acyl chain length,<sup>13,17</sup> CER headgroup chemistry,<sup>12</sup> and the addition of other lipids on bilayer properties,<sup>10,14,16–18</sup> as well as examining the permeation of small molecules through

CER-based bilayers.<sup>13,19,20</sup> However, most studies have relied on preassembled bilayer configurations, which may unduly influence the resulting structure and observed properties; the dense, gel phase structures formed by the SC lipids result in very low lipid mobility,<sup>15</sup> making them effectively stuck in their initial configurations over the typical simulation timescales considered. The use of bilayer, rather than multilayer structures, may also limit the insight that can be gained, as they restrict the conformations that can be adopted for the two-tailed CERs, *e.g.*, hairpin versus extended, which is a controversial topic in the experimental literature.<sup>21</sup> Self-assembled structures, where lipids form the desired structure with minimal external input, would be preferable, but the timescales required to reach equilibrium on what is likely a rough free energy landscape make this approach impractical for atomistic models.

To address these issues, computationally efficient coarse-grained (CG) models can be used. CG models allow the long times and large system sizes required to observe significant lipid rearrangements, including self-assembly, to be studied.<sup>22</sup> While numerous methods are available for deriving CG force fields,<sup>23-26</sup> the structure-based iterative Boltzmann inversion (IBI)<sup>27</sup> method is commonly used. The IBI method iteratively updates pair potentials between successive CG simulations to match target data in the form of site-site radial distribution functions (RDFs), and has been successfully applied to systems ranging from small molecules<sup>28</sup> to lipids<sup>29,30</sup> and polymers.<sup>31-33</sup> While IBI yields force fields that generally reproduce the target fluid structure with high accuracy, these force fields often fail to reproduce behavior at other state points,<sup>28,32-34</sup> as IBI only provides a means to match the target RDFs and cannot determine if the derived pair potentials accurately represent the underlying energy landscape. This problem is further intensified by the fact that there may be any number of pair potentials that sufficiently reproduce the target RDF,<sup>34</sup> and determination of which one is most representative is nontrivial. Hence, the utility of IBI-based force fields may be limited for studying processes that span multiple states,

*e.g.*, self-assembly.

In Chapter 5, I presented the development of a multistate extension to IBI (MS IBI) that helps to address the transferability problem of IBI-derived force fields, whereby pair potentials are optimized to match target RDFs at multiple state points simultaneously.<sup>34</sup> This extension effectively adds additional constraints to the optimization process, with the goal of finding a pair potential that simultaneously matches each target RDF, and thus adequately represents the underlying energy landscape. In this Chapter, I use MS IBI to expand the force field presented in Chapter 7 to include the cross-interactions between CER NS and FFA headgroups, and use the extended force field to investigate the self-assembly of large membranes composed of mixtures of CER and FFA.

## 8.2 Methods and Model

### 8.3 Simulations

Atomistic simulations used to gather target data were run with the CHARMM36 all-atom force fields.<sup>35</sup> Parameters for the CER NS headgroups were taken from previous work;<sup>12</sup> these have been shown to accurately reproduce the phase behavior of a pure CER NS bilayer. Water, when present, was modeled with the TIP3P water model.<sup>36</sup> For the atomistic simulations, van der Waals interactions were neglected beyond 12 Å, and were set to smoothly decay to zero between 10 Å to 12 Å. Long-ranged electrostatic interactions were treated with the particle mesh Ewald method<sup>37</sup> with a real-space cutoff of 12 Å. All atomistic simulations were run with GROMACS 5.1<sup>38</sup> with a 1 fs timestep. Constant pressure simulations of bulk fluid states employed an isotropic barostat, while bilayer states employed a semi-isotropic barostat with the pressure controlled independently in the bilayer normal and lateral directions. In both cases, the barostat used a pressure coupling constant of 10 ps and a compressibility of

$4.5 \times 10^{-5} \text{ bar}^{-1}$ .

Three distinct atomistic simulations were performed. First, a bulk fluid equimolar mixture of CER NS C24 and FFA C24:0, with 100 molecules of each, at 1 atm and 500 K. 500 K was used to induce a fluid phase, as the lipids were found to be frozen at the skin temperature of 305 K. This system was initialized by placing the lipids on a lattice in a very large simulation box and simulating for 5 ns at 800 K, followed by slowly reducing the temperature to 500 K while compressing the system down to a realistic density over 5 ns. The barostat was then turned on and the system was allowed to equilibrate for 5 ns, followed by 10 ns of data collection. A corresponding system was run at constant volume, with the density set to the average density from the constant pressure simulation. The remaining atomistic simulation was a bilayer containing an equimolar mixture of CER NS C24 and FFA C24:0. The initialization and equilibration of this system follows the protocol described in Chapter 4.

Two classes of CG simulations were performed: those during the MS IBI optimizations and those using the optimized force field. Initial configurations for the MS IBI optimization simulations were the final configurations from the corresponding atomistic simulations, mapped to the CG level. To increase sampling, bulk fluid CG systems were replicated  $2\times$  in each direction, and the bilayers were replicated  $3\times$  in each lateral direction. The bulk systems were run for 5 ns of equilibration, followed by 5 ns of data collection; the barostat was turned on in the final 1 ns of equilibration for the NPT state. The bilayer systems were equilibrated for 10 ns (with the barostat turned on for the final 5 ns for the NPT system), followed by 30 ns of data collection. The pressure coupling scheme for the CG simulations was identical to the corresponding atomistic simulations, *i.e.*, isotropic for the bulk systems and semi-isotropic for the bilayer systems. For force field validation, the initial configuration of the preassembled bilayer was the same as for the MS IBI CG simulations. The self-assembly simulations used several unique protocols, which are described in the corresponding sections below.

All CG simulations were run using the GPU-accelerated HOOMD-Blue toolkit<sup>39,40</sup> with a 10 fs timestep.

### 8.3.1 Model Description and Force Field Optimization

The coarse-grained mappings for the lipids used in this Chapter are shown in Figure 8.1 and were defined and used Chapter 7. Briefly, a 3-to-1 mapping is used, where each CG bead generally represents 3 heavy atoms. A notable exception is the hydroxyl groups in the CER headgroup, which are each mapped to a single bead; this was found to be necessary to capture the in-plane packing of the CER headgroups in a bilayer configuration, as discussed in Chapter 7. Both molecules share identical TAIL beads, which represent 3 successive carbons and their associated hydrogens in the lipid tails.

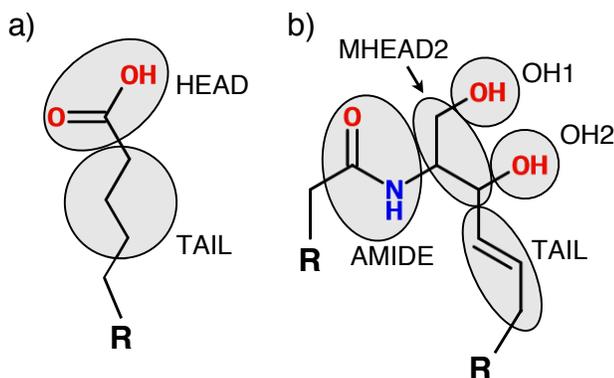


Figure 8.1: CG mappings for the lipid headgroups considered in this Chapter; a) FFA and b) CER NS. “R” denotes a set of TAIL beads bonded in series, with the number of tail beads determined by the length of the lipid tails. Note that the CG beads interact through spherically symmetric potentials despite being drawn as ellipses.

The specific terms in the force field were described in Chapter 7, but are also provided here for reference. Two CG beads are considered bonded if any two atoms represented by the two CG beads share a bond in the atomistic model. CG beads that are bonded interact through a harmonic bond-stretching potential, and any consecutively bonded triplet of CG beads interact through a harmonic angle-bending potential; the derivation of the parameters for these interactions follows the method of

Milano *et al.*,<sup>41</sup> where the (assumed Gaussian) distributions sampled in the atomistic simulations are Boltzmann-inverted to produce a harmonic potential. In addition to the bonded interactions, CG beads also interact through nonbonded pair potentials, which are optimized with MS IBI. The nonbonded pair interactions are neglected for beads that interact through bond-stretching or angle-bending potentials. The nonbonded interactions derived in this (and other) Chapters are made available at <https://github.com/iModels/sc-ff>.

The bonded interactions and nonbonded pair potentials for beads on the same type of molecule, along with the lipid–water pair potentials, were optimized with MS IBI and presented in Chapter 7. In this chapter, the lipid–lipid cross-interactions for CER NS and FFA, *i.e.*, the AMIDE–HEAD, MHEAD2–HEAD, OH1–HEAD, and OH2–HEAD interactions, are optimized using MS IBI. Bilayer and bulk fluid states, as described above, were to gather target RDFs. A higher weight was given to the bilayer states as compared to the bulk fluid states, with  $\alpha_{0,\text{bilayer}} = 0.5$  and  $\alpha_{0,\text{bulk}} = 0.1$ , where  $\alpha_0$  corresponds to  $\alpha$  value at  $r = 0$  for the linear  $a_s(r)$  function in Equation (5.4). The initial pair potentials,  $V^0(r)$ , were the Boltzmann-inverted RDFs from the bilayer state, *i.e.*,

$$V_{\gamma\beta}^0(r) = -k_B T \ln g_{\gamma\beta,\text{bilayer}}^{(\text{AA})}(r).$$

The optimizations were terminated when the fitness function, defined as

$$f_{\text{fit}} = 1 - \frac{\sum_{r=0}^{r_{\text{cut}}} |g^{(i)}(r) - g^{(\text{target})}(r)|}{\sum_{r=0}^{r_{\text{cut}}} [g^{(i)}(r) + g^{(\text{target})}(r)]},$$

increased by less than 0.01 for each pair being optimized at each state.

### 8.3.2 Analysis

To quantify the structure of the bilayers, the area per lipid (APL), nematic order ( $S_2$ ), tilt angle of the lipid tails ( $\theta$ ), and the bilayer thickness were calculated. These calculations were performed in the same manner as in Chapter 7, and are outlined in Appendix A. Important to note, however, is that the atomistic systems were first mapped to the CG level for analysis for a more direct comparison with the CG systems. For the self-assembled systems, the asymmetry of the number of lipids in each leaflet was quantified. A lipid was considered in the top leaflet if the  $z$  component of its center of mass was greater than the  $z$  component of the total lipid center of mass, and in the bottom leaflet otherwise. This quantity was found to not fluctuate once the bilayers had formed.

## 8.4 Results and Discussion

### 8.4.1 Force Field Derivation and Validation

The results of the CER NS–FFA headgroup optimizations are shown in Figures 8.2 and 8.3 for the NPT and NVT states, respectively. The target RDFs are reproduced with a high degree of accuracy at the bilayer states, as indicated by the  $f_{\text{fit}}$  values  $>0.95$ . The first peak and the general shape of the RDFs are accurately reproduced for each pair, which indicates that the CG model accurately captures the structural ordering of the headgroups within the bilayer plane. The RDFs at the bulk states are also accurately captured, with  $0.891 \leq f_{\text{fit}} \leq 0.915$  for all pairs. The CG model generally underpredicts the height of the first RDF peak and slightly overpredicts the RDF beyond the first peak, which indicates a higher degree of headgroup clustering but a lower degree of headgroup ordering in the first solvation shell in the CG model compared to the atomistic one. Note that the AMIDE–HEAD interactions for the bulk show the largest deviation, due to the anisotropic interactions, as was observed in Chapter 7. The CG model has slightly longer-ranged ordering than the atomistic,

which is a necessary compromise in order to capture the ordering in the bilayer states. The lower degree of RDF matching in the bulk phases is expected, as the bilayer states were given higher weight (via  $\alpha_s(r)$ ), since they are ultimately of more interest. Despite the lower fitness in the bulk states, the bulk densities calculated from NPT simulations with the CG ( $0.713 \pm 0.003$  g/mL) and atomistic ( $0.703 \pm 0.006$  g/mL) models compare favorably.

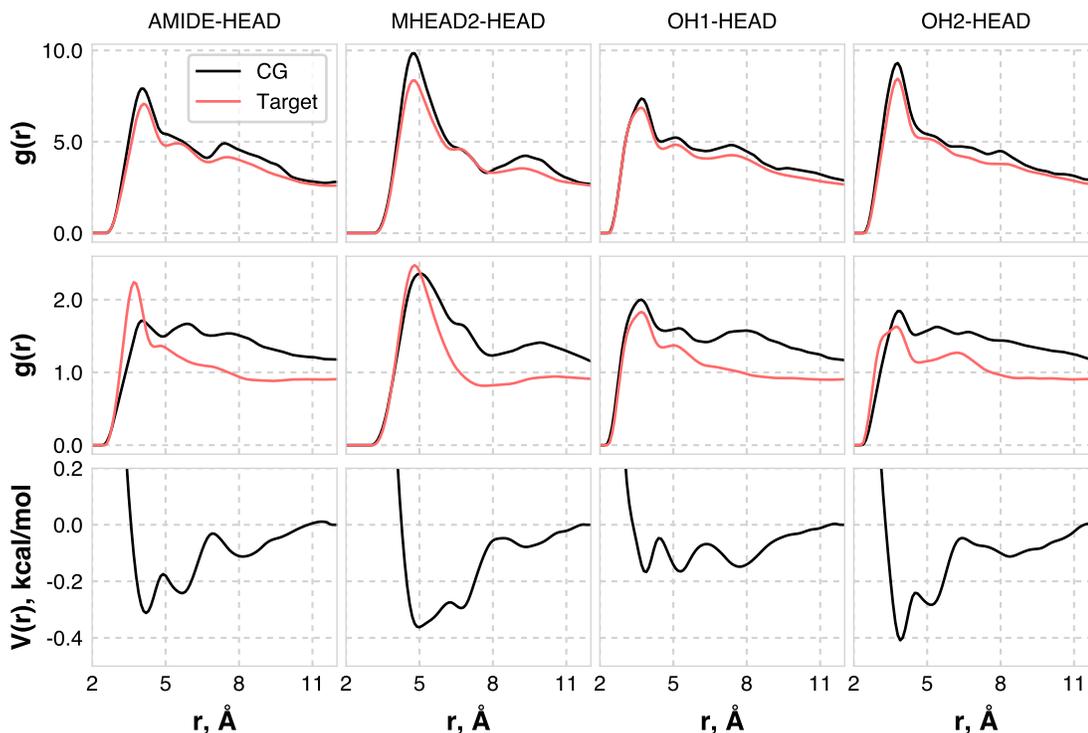


Figure 8.2: Representative RDFs and final optimized pair potentials for the pair interactions considered in this Chapter. Each column corresponds to a single pair, given at the top of the column. The first row shows the target and CG RDFs at the NPT bilayer state, the middle row shows the target and CG RDFs at the NPT bulk fluid state, and the bottom row shows the optimized pair potentials.

The pair potentials were further validated with a 200 ns simulation of the equimolar CER NS C24-FFA C24:0 bilayer to examine structural properties beyond the RDF, which are listed in Table 8.1. The high nematic order parameter ( $S_2$ ) shows that both models form bilayers with a high degree of orientational order in the lipid tails.

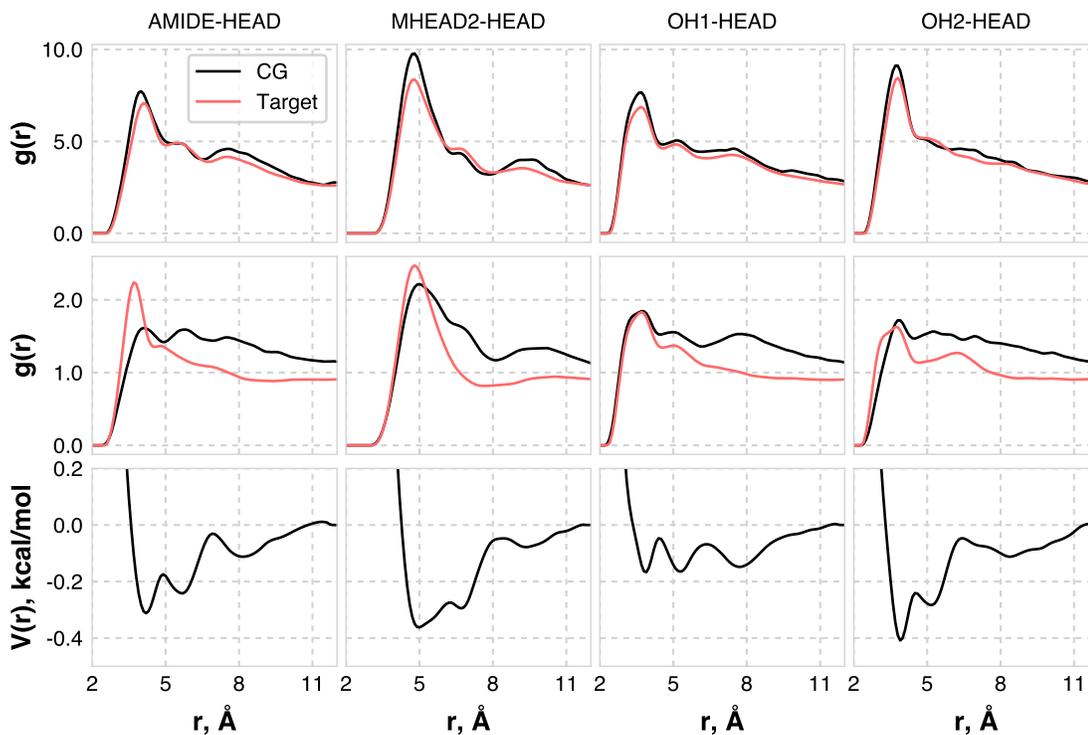


Figure 8.3: Representative RDFs and final optimized pair potentials for the pair interactions considered in this Chapter. Each column corresponds to a single pair, given at the top of the column. The first row shows the target and CG RDFs at the NVT bilayer state, the middle row shows the target and CG RDFs at the NVT bulk fluid state, and the bottom row shows the optimized pair potentials.

The CG model overpredicts the area per lipid (APL) by  $\sim 5\%$ , which is on the same order as the deviation for a pure CER NS C24 bilayer as observed in Chapter 7. The atomistic model predicts a higher tilt angle of the lipid tails with respect to the bilayer normal, although the difference is quite small ( $\sim 7\%$  when considering the relevant scale of  $0^\circ$  to  $90^\circ$ ). The bilayer thickness agrees very well with the atomistic model, differing by  $1 \text{ \AA}$ , illustrating the structural accuracy of the model with regard to both the lipid tails and the lipid–water interface. Overall, there is excellent agreement in the structural properties of multicomponent bilayers simulated with the atomistic and CG models.

Table 8.1: Comparison of structural properties of equimolar CER NS C24–FFA C24:0 preassembled (PA) and self-assembled (SA) bilayers simulated with the atomistic and CG models. The two numbers listed for the self-assembled systems (CG, SA-X) are for each of the two leaflets, as the asymmetric distribution of lipids between the leaflets, which is listed in the rightmost column, gives rise to different properties of the individual leaflets. Uncertainties represent the standard error in the calculations; for the self-assembled system, the uncertainties are on the same order as the preassembled CG system.

System	$S_2$	APL, $\text{\AA}^2$	Tilt, deg	Thickness, $\text{\AA}$	$n_{\text{top}}:n_{\text{bottom}}$
Atomistic, PA	$0.991 \pm 0.001$	$29.2 \pm 0.4$	$13 \pm 2$	$59.7 \pm 0.1$	-
CG, PA	$0.9787 \pm 0.0008$	$30.87 \pm 0.07$	$6.3 \pm 0.5$	$60.72 \pm 0.02$	-
CG, SA-1	0.964/0.975	31.18/30.68	9.6/6.1	$60.22 \pm 0.01$	0.964
CG, SA-2	0.976/0.976	31.36/30.70	6.4/6.2	$60.790 \pm 0.002$	0.978
CG, SA-3	0.975/0.957	30.43/32.37	6.0/11.7	$60.04 \pm 0.02$	1.064

#### 8.4.2 Self-assembly Into Lamellar Structures

After validating the model on preassembled bilayers, the self-assembly of the CG model into lamellar structures was examined. In general, bilayer self-assembly occurs in two major steps.<sup>42</sup> First, the lipid aggregate in the water to form separate lipid and aqueous domains; next, the lipids organize themselves such that the hydrophilic headgroups shield the hydrophobic tails from the aqueous phase, thus forming a bilayer structure. In this Chapter, the lipids were initialized in a high temperature, disordered, but phase-separated, configuration, shown in Figure 8.4a; the systems were set up such that the lipid phase spans the box in the  $x$  and  $y$  directions, but not along  $z$ , to ensure that the lipids can form a 2D periodic structure. From the high temperature state, the system temperature was first reduced to 305 K over 100 ns, followed by an isochoric expansion in the  $xy$  plane to  $1.7\times$  the APL of the preassembled, equilibrated bilayer. The system was then isochorically compressed in the  $xy$  plane to the equilibrium bilayer APL, after which the semi-isotropic barostat was activated, and the simulation continued for another 200 ns at 305 K and 1 atm. This approach of starting from a phase-separated state, followed by expansion and compression, was found in Chapter 7

to not affect the self-assembled structures, but rather makes the self-assembly more efficient and reliable, as major structural defects are more rapidly worked out compared to simply simulating the system at constant temperature and pressure. Three different bilayer self-assembly simulations were performed and analyzed to show reproducibility, with total compression/expansion times of 200 ns, 400 ns, and 600 ns (see Table 8.1).

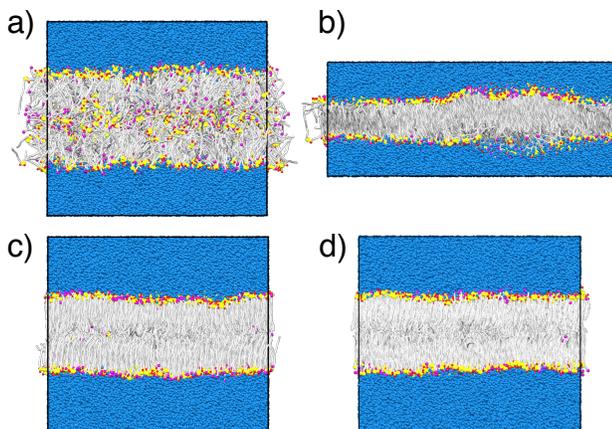


Figure 8.4: System snapshots during the bilayer self-assembly process of an equimolar mixture of CER NS C24 and FFA C24:0: a) initial, disordered configuration, b) at maximum expansion, c) at the end of the compression phase, and d) after 200 ns of simulation at 305 K and 1 atm. CG beads are colored as follows: blue = water, silver = lipid tails, yellow = AMIDE and MHEAD2, red = OH1 and OH2, and magenta = FFA headgroups.

Stable, well-ordered bilayers formed in each of the three self-assembly simulations. Snapshots of one of the systems during the self-assembly process are shown in Figure 8.4. Due to the stochastic nature of lipid organization during the cooling phase, the bilayers demonstrate an asymmetric distribution of lipids between the two leaflets and this asymmetry leads to different leaflets having different structural properties. This information is summarized in Table 8.1 and shows that when a particular leaflet has a lower APL (*i.e.*, more lipids, since the area is taken as the area of the simulation box), the lipid tails in that leaflet are more ordered and demonstrate a smaller tilt angle. In general, the structural properties of the self-assembled bilayers tend to agree well with those of the preassembled bilayer.

In the SC, the lipids are organized into repeating lamellar structures (*i.e.*, multilayers), and thus understanding the lipid structure in multilayers is of considerable interest. Here, using the validated CG force fields, simulations of multilayer self-assembly were performed in a similar manner to the bilayer self-assembly simulations, with two changes. First, instead of initially cooling to 305 K from the initial configuration, the system is isochorically compressed in the  $xy$  plane to 1/2 the equilibrium APL, so that, in principle, two bilayers can fit within the simulation box. This compression was followed by the expansion/compression process used for the bilayer self-assembly, except that the temperature was held at 550 K during the expansion, and then lowered to 305 K during the compression. The temperature was kept elevated to give the lipids more energy to reorganize during the expansion. To show reproducibility, three different total expansion/compression times were used: 200 ns, 400 ns and 600 ns. Stable multilayer structures formed in each simulation. A representative self-assembly process is shown in Figure 8.5. Of particular interest in these systems are the fractions of CER molecules that are in hairpin versus extended conformations, as this may have implications for the lamellar arrangements of CERs in the SC.<sup>21</sup> Here, an extended conformation was defined as the angle between the two tails of CER having an angle  $>90^\circ$ , and a hairpin conformation otherwise. In each of the self-assembled structures, between 23% to 28% of the CERs in the middle layers adopt a hairpin conformation, compared to  $\sim 15\%$  for pure CER systems, as reported in Chapter 7. This result suggests a mixture of hairpin and extended conformations is favorable. All systems obtain a thin layer of water mixed with the headgroups in the middle of the lipid phase, ranging from 0.127 to 0.221 water *beads* per lipid, or 0.508 to 0.884 water *molecules* per lipid, which is consistent with experimental observations of low lipid headgroup hydration in model SC membranes.<sup>43,44</sup>

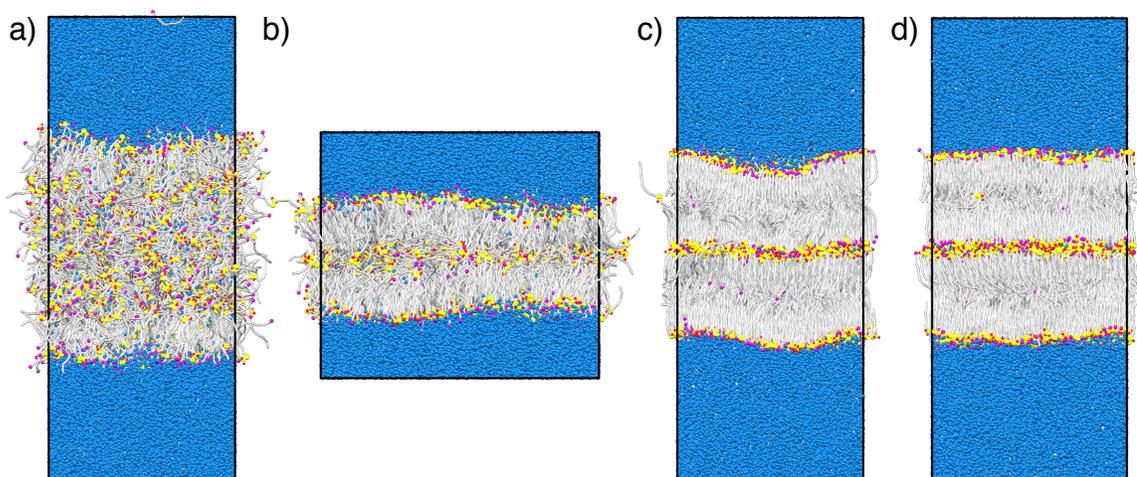


Figure 8.5: Multilayer self-assembly process of an equimolar mixture of CER NS C24 and FFA C24:0: a) disordered configuration after resizing to fit two bilayers, b) at maximum expansion, c) the final configuration of the compression phase, and d) final configuration after 200 ns of simulation at 305 K and 1 atm. CG beads colored the same as in Figure 8.4.

## 8.5 Conclusion

The work described in this Chapter highlights the use of multiscale modeling in studying the SC layer of skin. While atomistic simulations can be used to study properties at the atomistic level, *e.g.*, hydrogen bonding,<sup>12</sup> CG models are needed to study systems on a larger scale. Building on prior work, the work presented in this Chapter uses MS IBI to derive transferable CG models for skin lipids from atomistic simulations. These CG models can be used to study native lipid structures in the skin through self-assembly, which removes the influence of the initial configurations and hence the structures represent preferred lipid morphologies. As such, skin lipid systems can be computationally studied at multiple levels of resolution to provide a complete picture of how atomic-level features of lipids give rise to mesoscale properties. Information gleaned from these studies will prove useful for a wide range of applications, from strategies for restoring an impaired skin barrier to selectively penetrating the skin barrier, *e.g.*, for transdermal drug delivery.<sup>45,46</sup>

## 8.6 Bibliography

- [1] MOORE, T. C., IACOVELLA, C. R., LEONHARD, A. C., BUNGE, A. L., AND M<sup>C</sup>CABE, C. Molecular dynamics simulations of stratum corneum lipid mixtures: a multiscale perspective. *Biochemical and Biophysical Research Communications* **in press**, 2017, pp. 1–6.
- [2] WERTZ, P. W. AND VAN DEN BERGH, B. The physical, chemical and functional properties of lipids in the skin and other biological barriers. *Chemistry and Physics of Lipids* **91**, 1998, pp. 85–96.
- [3] MADISON, K. C. Barrier function of the skin: "la raison d'être" of the epidermis. *Journal of Investigative Dermatology* **121**, 2003, pp. 231–241.
- [4] JANSSENS, M., VAN SMEDEN, J., GOORIS, G. S., BRAS, W., PORTALE, G., CASPERS, P. J., VREEKEN, R. J., HANKEMEIER, T., KEZIC, S., WOLTERBEEK, R., LAVRIJSEN, A. P., AND BOUWSTRA, J. A. Increase in short-chain ceramides correlates with an altered lipid organization and decreased barrier function in atopic eczema patients. *Journal of Lipid Research* **53**, 2012, pp. 2755–2766.
- [5] LAVRIJSEN, A. P. M., BOUWSTRA, J. A., GOORIS, G. S., WEERHEIM, A., BODDÉ, H. E., AND PONEC, M. Reduced skin barrier function parallels abnormal stratum corneum lipid organization in patients with lamellar ichthyosis. *Journal of Investigative Dermatology* **105**, 1995, pp. 619–624.
- [6] PILGRAM, G. S. K., VISSERS, D. C. J., VAN DER MEULEN, H., KOERTEN, H. K., PAVEL, S., LAVRIJSEN, S. P. M., AND BOUWSTRA, J. A. Aberrant lipid organization in stratum corneum of patients with atopic dermatitis and lamellar ichthyosis. *Journal of Investigative Dermatology* **117**, 2001, pp. 710–717.
- [7] VAN SMEDEN, J., JANSSENS, M., BOITEN, W. A., VAN DRONGELEN, V., FURIO, L., VREEKEN, R. J., HOVNANIAN, A., AND BOUWSTRA, J. A. Intercellular skin barrier lipid composition and organization in netherton syndrome patients. *Journal of Investigative Dermatology* **134**, 2014, pp. 1238–1245.
- [8] PANDIT, S. A. AND SCOTT, H. L. Molecular-dynamics simulation of a ceramide bilayer. *The Journal of Chemical Physics* **124**, 2006, p. 014708.
- [9] DAS, C., NORO, M. G., AND OLMSTED, P. D. Fast cholesterol flip-flop and lack of swelling in skin lipid multilayers. *Soft Matter* **10**, 2014, pp. 7346–7352.
- [10] DAS, C., NORO, M. G., AND OLMSTED, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **97**, 2009, pp. 1941–1951.
- [11] DAS, C., NORO, M. G., AND OLMSTED, P. D. Lamellar and inverse micellar structures of skin lipids: Effect of templating. *Physical Review Letters* **111**, 2013, p. 148101.

- [12] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>c</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [13] GUPTA, R., DWADASI, B. S., AND RAI, B. Molecular dynamics simulation of skin lipids: effect of ceramide chain lengths on bilayer properties. *The Journal of Physical Chemistry B* **120**, 2016, pp. 12536–12546.
- [14] GUPTA, R. AND RAI, B. Molecular dynamics simulation study of skin lipids: effects of the molar ratio of individual components over a wide temperature range. *The Journal of Physical Chemistry B* **119**, 2015, pp. 11643–11655.
- [15] HOOPEs, M. I., NORO, M. G., LONGO, M. L., AND FALLER, R. Bilayer structure and lipid dynamics in a model stratum corneum with oleic acid. *The Journal of Physical Chemistry B* **115**, 2011, pp. 3164–3171.
- [16] NOTMAN, R., DEN OTTER, W. K., NORO, M. G., BRIELS, W. J., AND ANWAR, J. The permeability enhancing mechanism of DMSO in ceramide bilayers simulated by molecular dynamics. *Biophysical Journal* **93**, 2007, pp. 2056–2068.
- [17] DAS, C. AND OLMSTED, P. D. The physics of stratum corneum lipid membranes. *Philosophical Transactions of the Royal Society A* **374**, 2016, p. 20150126.
- [18] AKINSHINA, A., DAS, C., AND NORO, M. G. Effect of monoglycerides and fatty acids on a ceramide bilayer. *Physical Chemistry Chemical Physics* **18**, 2016, pp. 17446–17460.
- [19] DAS, C., OLMSTED, P. D., AND NORO, M. G. Water permeation through stratum corneum lipid bilayers from atomistic simulations. *Soft Matter* **5**, 2009, pp. 4549–4555.
- [20] GUPTA, R., SRIDHAR, D. B., AND RAI, B. Molecular dynamics simulation study of permeation of molecules through skin lipid bilayer. *The Journal of Physical Chemistry B* **120**, 2016, pp. 8987–8996.
- [21] KESSNER, D., RUETTINGER, A., KISELEV, M. A., WARTEWIG, S., AND NEUBERT, R. H. H. Properties of ceramides and their impact on the stratum corneum structure. Part 2: stratum corneum lipid model systems. *Skin Pharmacology and Physiology* **21**, 2008, pp. 58–74.
- [22] INGÓLFSSON, H. I., LOPEZ, C. A., UUSITALO, J. J., DE JONG, D. H., GOPAL, S. M., PERIOLE, X., AND MARRINK, S. J. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 2014, pp. 225–248.
- [23] ERCOLESSI, F. AND ADAMS, J. B. Interatomic potentials from first-principles calculations: the force-matching method. *Europhysics Letters* **26**, 1994, p. 583.

- [24] IZVEKOV, S. AND VOTH, G. A. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B* **109**, 2005, pp. 2469–2473.
- [25] LYUBARTSEV, A. P. AND LAAKSONEN, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52**, 1995, p. 3730.
- [26] CHAIMOVICH, A. AND SHELL, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *The Journal of Chemical Physics* **134**, 2011, p. 094112.
- [27] REITH, D., PÜTZ, M., AND MÜLLER-PLATHE, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24**, 2003, pp. 1624–1636.
- [28] QIAN, H.-J., CARBONE, P., CHEN, X., KARIMI-VARZANEH, H. A., LIEW, C. C., AND MÜLLER-PLATHE, F. Temperature-transferable coarse-grained potentials for ethylbenzene, polystyrene, and their mixtures. *Macromolecules* **41**, 2008, pp. 9919–9929.
- [29] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A structurally relevant coarse-grained model for cholesterol. *Biophysical Journal* **99**, 2010, pp. 2896–2905.
- [30] HADLEY, K. R. AND M<sup>C</sup>CABE, C. A coarse-grained model for amorphous and crystalline fatty acids. *The Journal of Chemical Physics* **132**, 2010, p. 134505.
- [31] FALLER, R. AND REITH, D. Properties of poly(isoprene): model building in the melt and in solution. *Macromolecules* **36**, 2003, pp. 5406–5414.
- [32] BAYRAMOGLU, B. AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 2012, pp. 9205–9219.
- [33] BAYRAMOGLU, B. AND FALLER, R. Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion. *Macromolecules* **46**, 2013, pp. 7957–7976.
- [34] MOORE, T. C., IACOVELLA, C. R., AND M<sup>C</sup>CABE, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140**, 2014, p. 224104.
- [35] KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O’CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D., AND PASTOR, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 2010, pp. 7830–7843.

- [36] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 1983, pp. 926–935.
- [37] ESSMANN, U., PERERA, L., BERKOWITZ, M. L., DARDEN, T., LEE, H., AND PEDERSEN, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 1995, pp. 8577–8593.
- [38] ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B., AND LINDAHL, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 2015, pp. 19–25.
- [39] ANDERSON, J. A., LORENZ, C. D., AND TRAVESSET, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 2008, pp. 5342–5359.
- [40] GLASER, J., NGUYEN, T. D., ANDERSON, J. A., LUI, P., SPIGA, F., MILLAN, J. A., MORSE, D. C., AND GLOTZER, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **192**, 2015, pp. 97–107.
- [41] MILANO, G., GOUDEAU, S., AND MÜLLER-PLATHE, F. Multicentered Gaussian-based potentials for coarse-grained polymer simulations: linking atomistic and mesoscopic scales. *Journal of Polymer Science Part B: Polymer Physics* **43**, 2005, pp. 871–885.
- [42] MARRINK, S. J., LINDAHL, E., EDHOLM, O., AND MARK, A. E. Simulation of the spontaneous aggregation of phospholipids into bilayers. *Journal of the American Chemical Society* **123**, 2001, pp. 8638–8639.
- [43] KISELEV, M. A., RYABOVA, N. Y., BALAGUROV, A. M., DANTE, S., HAUSS, T., ZBYTOVSKA, J., WARTEWIG, S., AND NEUBERT, R. H. H. New insights into the structure and hydration of a stratum corneum lipid model membrane by neutron diffraction. *European Biophysics Journal* **34**, 2005, pp. 1030–1040.
- [44] ENGELBRECHT, T. N., SCHRÖTER, A., HAUSS, T., DEMÉ, B., SCHEIDT, H. A., HUSTER, D., AND NEUBERT, R. H. H. The impact of ceramides np and ap on the nanostructure of stratum corneum lipid bilayer. part i: neutron diffraction and <sup>2</sup>H NMR studies on multilamellar models based on ceramides with symmetric alkyl chain length distribution. *Soft Matter* **8**, 2012, pp. 6599–6607.
- [45] GUY, R. H. Transdermal drug delivery. In *Drug Delivery*. Springer (2010), pp. 399–410.
- [46] PRAUSNITZ, M. R. AND LANGER, R. Transdermal drug delivery. *Nature Biotechnology* **26**, 2008, pp. 1261–1268.

## CHAPTER 9

### DEVELOPMENT OF COARSE-GRAINED MODELS FOR A MINIMAL STRATUM CORNEUM LIPID MIXTURE

The work described in this chapter is based on a manuscript in preparation.

#### 9.1 Introduction

In previous Chapters, I laid out the need for CG models of SC lipids. In Chapters 6 to 8, I presented the development of CG force fields for water, CER NS, and FFA. However, this force field lacks the necessary interactions with CHOL to simulate a realistic mixture of SC lipids. In this Chapter, the missing interactions involving CHOL are optimized, such that the updated CG force field contains all of the interactions necessary to simulate a mixture of CER NS, CHOL, and FFA.

Previous work in the M<sup>c</sup>Cabe group focused on developing a CG model of CHOL using IBI.<sup>1</sup> This model was unique in that it was the first structurally-derived CG model of CHOL. The model was accurate in the sense that the fluid structure, density, and crystal structure were accurately reproduced. However, as a result of the single-state nature of the force field optimization, two separate force fields were required to capture the structure in ordered and disordered phases.<sup>1</sup> As a result, each of the different force fields were used in separate portions of a self-assembly simulation, the “fluid” one during the initial stages, and the “crystal” one once a self-assembled structure had formed. Here, a single, transferable force field for CHOL is developed to be used in self-assembly simulations of SC lipid mixtures. Therefore, the work described herein can be viewed as a refinement of the previous CHOL model of Hadley and M<sup>c</sup>Cabe.

## 9.2 Models and Methods

Three distinct sets of simulations were performed. Atomistic simulations were performed to gather target data that was used to derive the CG force field. CG simulations were performed to optimize the CG force field, including the simulations performed during MS IBI and surface wetting simulations. Finally, simulations using the optimized CG force fields were used to first validate the force field and then to study the self-assembly of lipid mixtures containing CER NS, CHOL, and FFA.

### 9.2.1 Atomistic Simulations

All atomistic simulations were performed with the GROMACS simulation engine<sup>2</sup> with a 1 fs timestep. van der Waals interactions were cut off at 12 Å, and set to smoothly decay to zero between 10 Å to 12 Å. The particle-mesh Ewald summation method<sup>3</sup> was used to handle long-ranged electrostatic interactions, with a real-space cutoff of 12 Å. The Nosé–Hoover thermostat<sup>4</sup> with a time coupling constant of 1 ps was used to control the temperature. The Parinello–Rahman barostat<sup>5</sup> was used for pressure control, with a coupling constant of 10 ps and a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ . For constant-pressure bulk fluid simulations, pressure was controlled isotropically. For constant-pressure bilayer simulations, pressure was controlled separately in the bilayer normal and lateral directions. All lipids were modeled with the atomistic CHARMM36 force field,<sup>6</sup> supplemented by parameters for the CER NS headgroup where necessary.<sup>7</sup> Water was modeled with the TIP3P water model.<sup>8</sup>

A large number of different atomistic simulations were performed to gather target data. Pure, bulk fluid CHOL simulations contained 64 CHOL molecules, and were run at four different statepoints, as listed in Table 9.1. CHOL–water systems contained 1250 water molecules, and the corresponding number of CHOL molecules to give the concentrations listed in Table 9.1. Bulk fluid lipid mixtures contained 64 copies of each lipid. These systems were initialized with lipids randomly distributed throughout

a large simulation box, and simulating for 5 ns at 800 K to randomize the positions of the molecules. The system temperature was reduced to the target temperature while condensing down to a realistic liquid density over 5 ns. For constant-pressure, bulk fluid systems, the barostat was turned on, and the systems were equilibrated for 5 ns, before collecting data over 10 ns. For constant-volume, bulk fluid systems, the systems were equilibrated at the target density for 5 ns before collecting target data over the final 10 ns. Mixed-lipid bilayers were taken from simulations described in Chapter 4. The dehydrated bilayers were constructed by removing the water from the corresponding bilayer structures, replicating the bilayer in the  $z$  direction, and equilibrating with the RWMD protocol described in Chapter 4. The CHOL-FFA C16:0 monolayer was constructed by placing an equal number of CHOL and FFA molecules on a hexagonal lattice with an area per lipid (APL) of  $33.1 \text{ \AA}^2$ . The atoms corresponding to the headgroups of each lipid were fixed, and the system was equilibrated for 10 ns at 405 K. The system temperature was then lowered to 305 K over the course of 20 ns, and allowed to equilibrate for 10 ns. Target data was then collected for 50 ns.

Simulations of droplets on surfaces were performed for tuning various interactions in the CG force field. These include a water droplet on the tail side of a CHOL monolayer surface, and CER headgroups on the headgroup side of a CHOL monolayer. These systems were initialized by carving a sphere out of a bulk fluid of the droplet species (*i.e.*, either water or CER headgroups), and placing it on the specified surface. The droplet was allowed to relax, and the simulation was run for 5 ns after the contact angle had reached a steady-state value.

## 9.2.2 CG Simulations

All CG simulations were performed using the GPU-accelerated HOOMD-Blue simulation engine with a 10 fs timestep.<sup>9,10</sup> Constant-pressure CG simulations used the same pressure coupling schemes as the corresponding atomistic simulations. Three

Table 9.1: Target states used for MS IBI optimizations. The states used to optimize each interaction are described in the text. State condition specifies the ensemble—states with a pressure listed were simulated in the NPT ensemble; states with a density listed were simulated in the NVT ensemble.

Name	State condition	Temperature, K	Label for Table 7.2
<b>Pure CHOL</b>			
Bulk fluid	1.0 atm	550	CHOL1
Bulk fluid	0.85 g/mL	400	CHOL2
Bulk fluid	0.74 g/mL	550	CHOL3
Bulk fluid	0.65 g/mL	700	CHOL4
<b>CHOL–water</b>			
1.0 mol % CHOL	1.0 atm	305	W-CHOL1
2.0 mol % CHOL	1.0 atm	305	W-CHOL2
<b>CER NS–CHOL</b>			
Bulk fluid	1 atm	500	CRCHL-NPT
Bulk fluid	0.74 g/mL	500	CRCHL-NVT
Bilayer	1 atm	305	CRCHL-BIL-NPT
Bilayer	$38.8 \text{ \AA}^2$	305	CRCHL-BIL-NVT
Dehydrated multilayer	1 atm	305	CRCHL-MULT-NPT
Dehydrated multilayer	$38.8 \text{ \AA}^2$	305	CRCHL-MULT-NVT
<b>CHOL–FFA C24:0</b>			
Bulk fluid	1 atm	400	CHLFFA-NPT
Bulk fluid	0.79 g/mL	400	CHLFFA-NVT
<b>CHOL–FFA C16:0</b>			
Monolayer	$33.1 \text{ \AA}^2/\text{chain}$	305	CHLFFA-MONO
Bilayer	1 atm	305	CHLFFA-BIL-NPT
Bilayer	$29.9 \text{ \AA}^2$	305	CHLFFA-BIL-NVT
Dehydrated multilayer	1 atm	305	CHLFFA-MULT-NPT
Dehydrated multilayer	$29.9 \text{ \AA}^2$	305	CHLFFA-MULT-NVT

Table 9.2: States used to optimize each pair potential and the weight  $\alpha_{s,0}$  given to each state. Pairs are denoted by bold text, with the states used to optimize a given (set of) interaction(s) listed below the interaction name. The state condition names correspond to the label given in Table 9.1.

States used	$\alpha_{s,0}$
<b>CHOL-CHOL</b>	
CHOL1 + CHOL2 + CHOL3 + CHOL4	0.7
<b>CHOL-WATER</b>	
W-CHOL1 + W-CHOL2	0.4
<b>CERHEADS-CHEAD</b>	
CRCHL-(NPT + NVT)	0.05
CRCHL-BIL-(NPT+NVT)	0.5
CRCHL-MULT-(NPT+NVT)	0.5
<b>CERHEADS-CBODY</b>	
CRCHL-(NPT + NVT)	0.05
<b>FHEAD-CHEAD</b>	
CHLFFA-(NPT+NVT)	0.05
CHLFFA-BIL-(NPT+NVT)	0.5
CHLFFA-MULT-(NPT+NVT)	0.5
<b>FHEAD-CBODY</b>	
CHLFFA-(NPT+NVT)	0.05
<b>TAILS-CHEAD</b>	
CHLFFA-(NPT+NVT)	0.05
<b>TAILS-CBODY</b>	
CHLFFA-(NPT+NVT)	0.05
CHLFFA-MONO	0.5
CHLFFA-BIL-(NPT+NVT)	0.5
CHLFFA-MULT-(NPT+NVT)	0.5

distinct sets of CG simulations were performed. The first set was run during the MS IBI optimizations to generate the RDF for each iteration. These use the final configuration from the corresponding atomistic simulation for the initial configuration, except that configurations were replicated  $2\times$  in each direction for bulk systems and  $3\times$  in each lamellar lateral direction for the lamellar systems. These simulations use the force field derived in Chapter 8, except for interactions involving CHOL beads, which were optimized during this stage. The next set of simulations used the newly derived CHOL interactions to validate the force field and tune, where necessary, specific interactions. These include the surface wetting simulations used to tune specific interactions. These simulations were performed analogously to their atomistic counterparts, except that systems were simulated for longer times to ensure stability of the structures. If it was found that the force field needed to be tuned based on the surface wetting behavior, these simulations were repeated with the modified force field. This cycle was repeated until no further modifications were required (*i.e.*, the CG wetting behavior matched the atomistic wetting behavior). Finally, after the CG force field was modified and validated, the self-assembly of mixtures of the various lipids was simulated. Systems used for self-assembly were initialized in disordered configurations, with the lipids and water in separate domains, and the lipid phase spanning the box in the  $x$  and  $y$  directions, but not in  $z$ ; this setup was used to ensure that the lipid phase only forms a 2D periodic structure. The size and shape of the simulation box was initialized with a target density and aspect ratio amenable to the formation of a periodic structure. In all systems, the density of the water phase was set to 1.0 g/mL, and the lipid density was set to 0.8 g/mL. Given the lipid density and composition, the lipid volume is specified, and the cross-sectional area of the box was set based on the estimated APL and number of leaflets. For example, for the equimolar CER NS-CHOL-FFA bilayer, the APL was estimated at  $35 \text{ \AA}^2$ , so the

cross-sectional area  $A_c$  was set to

$$A_c = 35 \text{ \AA}^2 * \frac{n_{\text{lipids}}}{2},$$

where the denominator would be 4 for a stacked bilayer. To remove high-energy overlaps, very brief NVE simulations were first performed, with a limit on the displacement of each bead at each timestep. Next, a brief simulation at 305 K and 1 atm with an isotropic barostat was performed to allow the total density of the system to equilibrate while keeping the aspect ratio of the box constant. After density equilibration, the barostat was turned off, and the systems underwent an expansion/compression process, as described in Chapter 8. Following the expansion/compression, the semi-isotropic barostat was turned on, and the system was annealed from 450 K to 305 K over 100 ns to remove any defects formed during the expansion/compression phase. After annealing, the simulation was continued for at least another 200 ns, and data for analysis was collected for at least 100 ns.

### 9.2.3 CG Mappings

The CG mapping for CHOL is the same as in the work of Hadley and McCabe,<sup>1</sup> and is shown in Figure 9.1. In this mapping, the CHOL molecule is modeled as 9 separate interaction sites, of which there are 5 unique types for the nonbonded interactions. The CHEAD (*i.e.*, CHOL head) bead represents the hydroxyl group and two of the carbon atoms in the A ring. The remaining carbon atoms that make up the 4-ring structure are divided into 4 separate RING beads, as shown in Figure 9.1. The CHOL tail attached to the D ring is divided into two distinct beads, CTAIL and CTERM (*i.e.*, CHOL tail and CHOL terminal). The methyl groups attached to C10 and C13 are each mapped to a CHME bead (*i.e.*, chiral methyl), as this was shown to help capture the molecular roughness of CHOL.<sup>1</sup> For the remainder of this Chapter, the RING, CTAIL, CTERM, and CHME beads are collectively referred to as CBODY.

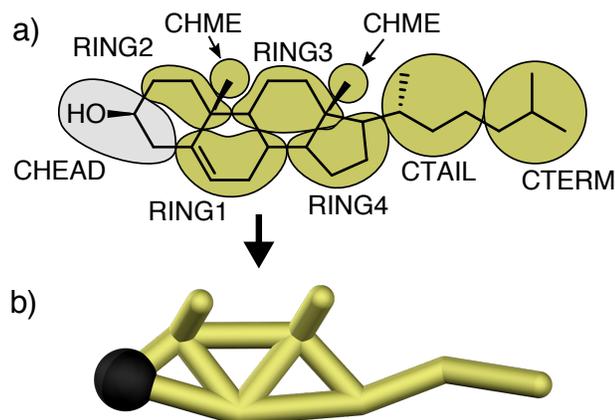


Figure 9.1: a) CG mapping used for CHOL. b) stick representation of the CG CHOL model. The four interconnected RING beads are treated as the same type (RING) for nonbonded interactions, but are distinguished here since they are treated as separate types for bonded interactions; the two CHME beads are treated as the same type for both bonded and nonbonded interactions. Despite being drawn as irregular shapes in (a), the CG beads interact through spherically symmetric nonbonded potentials. Note that the CG beads are not drawn to scale, but are drawn to emphasize connectivity.

#### 9.2.4 CG Force Field Description and Optimization

The CG force field consists of 3 unique types of interactions: 2-body bond-stretching interactions between bonded CG beads, 3-body angle-bending interactions between 3 consecutively bonded CG beads, and pairwise nonbonded interactions between CG beads separated by 3 or more bonds (or on different molecules). The bond-stretching and angle-bending interactions take the form of harmonic potentials, whereas the nonbonded interactions take the form of numerical pair potentials (*i.e.*, they are not constrained to a functional form). The bonded interactions for CER NS and FFA were derived and reported in Chapter 7. The nonbonded interactions for CER NS, FFA, and water were reported in Chapters 7 and 8.

In this Chapter, as in Chapter 7, bonded interactions for CHOL were derived using the distributions sampled in the atomistic simulations, following the method of

Table 9.3: Bond parameters for the CG CHOL model.

Name	$r_0$ , Å	$K_r$ , kcal/(mol Å <sup>2</sup> )
CHEAD-RING1	4.27	125.5
CHEAD-RING2	2.58	245.9
RING1-RING2	3.57	324.1
RING2-RING3	4.20	132.2
RING3-RING4	3.21	362.6
RING1-RING4	4.48	152.1
RING1-RING3	3.60	352.3
RING4-CTAIL	3.69	50.3
CTAIL-CTERM	4.22	13.2
RING2-CHME	2.37	65.4
RING3-CHME	2.66	116.2

Milano *et al.*<sup>11</sup> 11 bond and 23 angle types are present in the CG model, and each have unique stiffnesses and equilibrium separations. The parameters used for each bond/angle type are listed in Tables 9.3 and 9.4.

MS IBI was used to optimize a number of sets of nonbonded pair interactions. This was done because sets of pair potentials were derived sequentially, in that the CHOL-CHOL interactions were first optimized, and then used in the optimization of the CHOL-WATER interactions. Finally, with the CHOL-CHOL and CHOL-WATER interactions optimized and validated, they were used during the CG simulations in MS IBI for optimizing the CER-CHOL and CHOL-FFA cross-interactions.

The optimization of the pure CHOL interactions involved 4 target states, each a bulk fluid of CHOL molecules, as listed in Table 9.2. The CHOL-WATER interactions were optimized with target data from just two states, as listed in Table 9.2. The optimization CER-CHOL and CHOL-FFA cross-interactions were carried out simultaneously using target data from the states listed in Table 9.2. Following the work in Chapters 7 and 8, interactions were optimized against target data from representative states, that is, states where the interactions are thought to play a role in the structure. Following

Table 9.4: Angle parameters for the CG CHOL model.

Name	$\theta_0$ , deg	$K_\theta$ , kcal/(mol rad <sup>2</sup> )
RING1-CHEAD-RING2	56.3	1132.6
RING2-RING1-RING4	115.1	663.1
RING3-RING1-RING4	44.8	3617.7
RING2-RING1-RING3	71.0	1323.0
CHEAD-RING1-RING4	151.2	351.3
CHEAD-RING1-RING3	144.3	180.2
CHEAD-RING1-RING2	36.7	2941.4
RING1-RING2-RING3	54.0	2678.8
CHEAD-RING2-RING3	139.4	597.5
CHEAD-RING2-RING1	85.4	1011.1
CHEAD-RING2-CHME	92.8	75.5
RING3-RING2-CHME	60.9	436.6
RING1-RING2-CHME	49.7	643.4
RING2-RING3-CHME	133.2	77.5
RING1-RING3-RING4	81.1	1329.5
RING1-RING3-RING2	53.4	2787.1
RING4-RING3-CHME	53.6	671.7
RING2-RING3-RING4	133.4	446.5
RING1-RING3-CHME	95.5	354.5
RING1-RING4-RING3	52.5	2675.5
RING3-RING4-CTAIL	94.4	276.1
RING1-RING4-CTAIL	144.3	180.2
RING4-CTAIL-CTERM	125.0	7.6

this line of reasoning, the specific pair interactions were optimized at the states listed in Table 9.2. In this manner, the lipid–lipid headgroup (*i.e.*, hydrophilic–hydrophilic) interactions were all optimized at bulk states to capture the interactions in disordered configurations, and bilayer and stacked bilayer states to capture the headgroup packing in the lamellar phases. The hydrophobic–hydrophobic (*i.e.*, TAILS–CBODY) interactions were optimized at bulk states to capture the behavior in disordered phases, and also at the bilayer, stacked bilayer, and monolayer states to capture the in-plane packing relevant to the lamellar phases. The mixed hydrophobic–hydrophilic interactions were only optimized at the bulk states, as there are much fewer of those interactions in the lamellar phases and hence are not expected to drive any structural behavior at those states.

Optimizations were carried out using the open-source MS IBI Python library.<sup>12</sup> The optimizations were terminated when the value of the fitness function defined in Equation (5.5) changed by less than 0.01 for all pairs at all states between successive iterations.

Analysis of the CG systems was performed the same as described in Chapters 7 and 8 and outlined in Appendix A.

## 9.3 Results

### 9.3.1 Forcefield Derivation

*Pure CHOL and CHOL–WATER* Overall, excellent agreement was found between the target and CG RDFs. In any given CHOL system, there will be far more RING–RING interactions, so this pair is used as a representative pair for the force field; the optimized pair potential, along with the target and CG RDFs at two of the states used in the optimization, are shown in Figure 9.2. As expected based on the excellent agreement in the RDFs, the CG and atomistic models show good agreement in the density of the NPT target state, with values of  $0.732 \pm 0.005$  g/mL (CG) and

$0.73 \pm 0.01$  g/mL (atomistic).

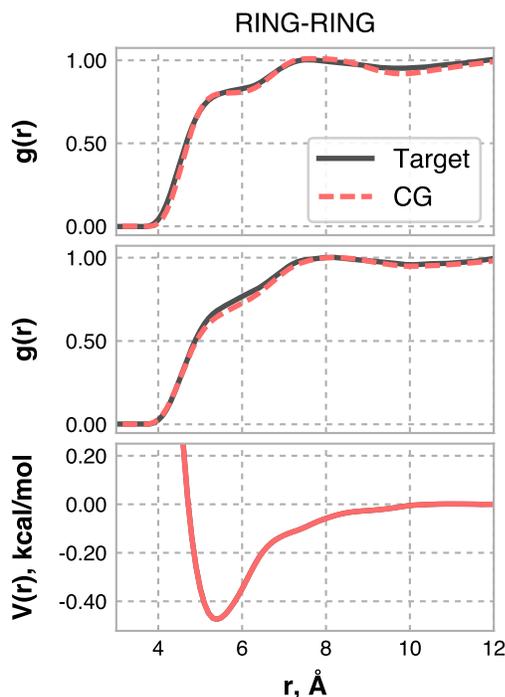


Figure 9.2: RDFs and pair potential from the pure CHOL force field optimization. Top: target(black) and CG (ref) RDFs from the 400 K NVT state; middle: target and CG RDFs at the 550 K NPT state; bottom: pair potential that yields the CG RDFs above.

Overall, good agreement was found between the CG and target RDFs for interactions between CHOL and water. However, as discussed in Chapter 7, RDF matching does not ensure that the partitioning behavior will be replicated. Therefore, to test the hydrophobicity of the CHOL beads, the wetting behavior of water on a CHOL monolayer was investigated. Using the optimized force field, the water tended to spread on the CBODY surface, forming a contact angle of  $82^\circ$ . This result indicates that the CBODY surface is more hydrophilic than the corresponding atomistic surface, which gave a contact angle of  $127^\circ$ . The CBODY–WATER interactions were scaled to be less attractive, and a scaling value of 0.4 was found to give good agreement with the atomistic contact angle. In contrast, water was found to readily wet the CHEAD surface, and therefore the CHEAD–WATER interaction was not modified from its

optimized form.

After the CHOL–CHOL and CHOL–WATER pair potentials were optimized and modified, they were used in the optimization of the mixed-lipid interactions. Two sets of pair potentials were optimized: CHOL–FFA and CER–CHOL. Although CER and FFA tails both consist of TAIL beads, the interactions between the TAIL and CHOL beads were only optimized with data from CHOL–FFA mixtures for simplicity. As in Chapter 7, interactions were optimized using target data from states where specific interactions are thought to play an important structural role, *e.g.*, target data from lamellar states is not used in the interactions between lipid headgroups and tails. In all cases, more weight (through the  $\alpha$  term of the MS IBI pair potential update formula) was given to the lamellar states ( $\alpha = 0.5$ ) than the bulk fluid states ( $\alpha = 0.05$ ). The results of these sets of optimizations are discussed separately below.

*CHOL–FFA* The TAIL–CBODY RDFs are generally reproduced with a fairly high degree of accuracy, which is important since there will be more of these interactions in any given system containing CHOL. The TAIL–RING interactions are likely dominant in these system because of the number of each of these beads, therefore, this pair is a good representation of the TAIL–CBODY interactions. The TAIL–RING RDFs are shown in Figure 9.3, where it can be seen that the target RDFs are well-reproduced with the CG model. The CHEAD–FHEAD RDFs are decently captured at the lamellar states, as shown in Figure 9.4. Interestingly, the best fit is found at the stacked bilayer state, even compared to the bilayer state. This is likely a result of the relative locations of the first peaks: at the bilayer state, the first peak in the RDF is at a larger separation compared to the stacked bilayer and bulk states. To compensate, the pair potential shifts the first RDF peak to smaller separation values. Like in previous optimizations, the fits at the lamellar states are better than at the bulk states as a result of the lamellar states being given a larger weight in the optimization process. The RDFs between the mixed hydrophilic–hydrophobic beads, *e.g.*, TAIL–CHEAD (shown in

Figure 9.5), are captured with less accuracy than in previous Chapters. This may be a result of the shape of CHOL, which is more complex than the simpler chain-like CER and FFA molecules. Despite the discrepancies, the density of the bulk fluid NPT state was found to be  $0.81 \pm 0.03$  g/mL, which is in good agreement with the atomistic value of  $0.79 \pm 0.01$  g/mL. Additionally, in a lamellar configuration, there are likely to be very few of these types of interactions. Therefore, these deviations in the RDF are deemed acceptable considering the target objective of this force field.

*CER NS-CHOL* Only the CERHEADs-CHEAD and CERHEADs-CBODY interactions are considered here, since the TAILS-CHOL interactions were optimized using target data from CHOL-FFA systems. In general, as shown in Figure 9.6, the CERHEADs-CHEAD interactions are accurately reproduced at the lamellar states. As seen previously, the fit is much better at these states compared to the bulk fluid states, as a result of the higher weight given to the lamellar states in the optimization process. Despite the accurate RDFs and stable lamellar phases, preliminary self-assembly simulations found that CER-CHOL mixtures did not form bilayer or multilayer structures, in contrast to the stable bilayers simulated with the atomistic model. To investigate, the wetting behavior of CER headgroups on a CHOL headgroup surface was examined. Here, the surface was a CHOL monolayer where the CTERM beads were fixed, and the droplet was composed of CER headgroups (*i.e.*, no tails) and interacted with the CHEAD side of the monolayer. The CER headgroups fully wetted the CHEAD surface with the atomistic model. However, with the CG model, the CER headgroups tended to form a droplet on the surface, indicating an imbalance in the CERHEADs-CERHEADs and CERHEADs-CHEAD interactions. This imbalance was corrected by systematically making the CERHEADs-CHEAD interactions more attractive until the CER headgroups wetted the surface with the atomistic model. It was found that a scaling factor of 1.5 was required to give the correct wetting behavior; therefore, the derived interactions scaled by 1.5 were used for the rest of this work.

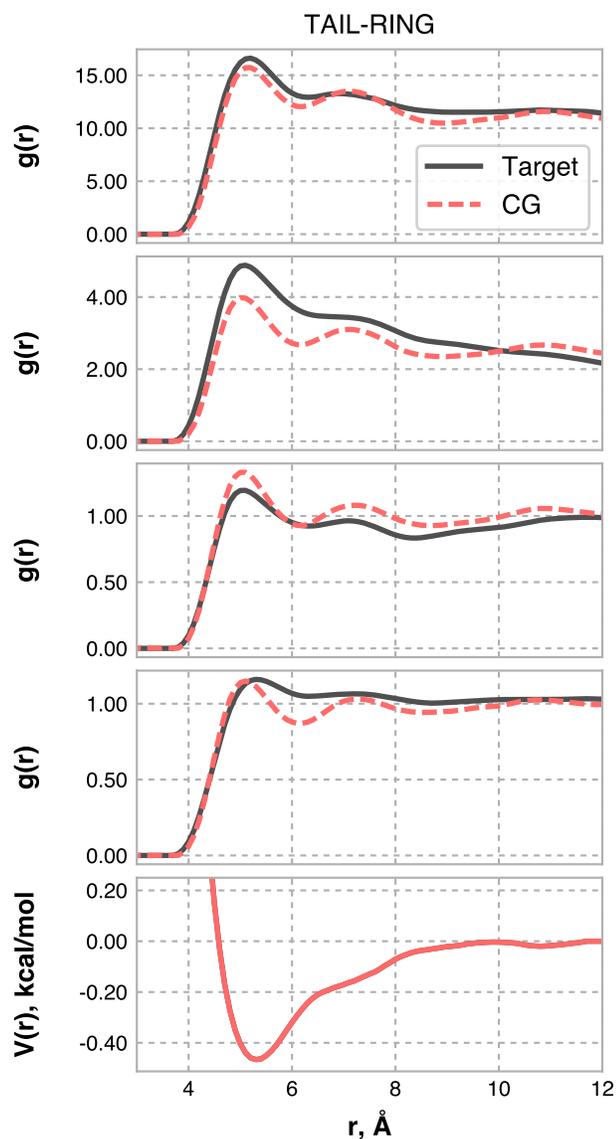


Figure 9.3: TAIL-RING RDFs and pair potential. From top to bottom: RDFs from the monolayer state; RDFs at the NPT bilayer state; RDFs at the stacked bilayer state; RDFs at the bulk fluid NPT state; and the pair potential that yields the CG RDFs above.

The CERHEADs-CBODY interactions, which were optimized using target data from only the bulk fluid states, generally show a good match with the target RDFs.

### 9.3.2 Force Field Validation

After optimizing and modifying (*e.g.*, based on wetting simulations) all of the lipid-lipid cross-interactions, a set of validation simulations using the final force field

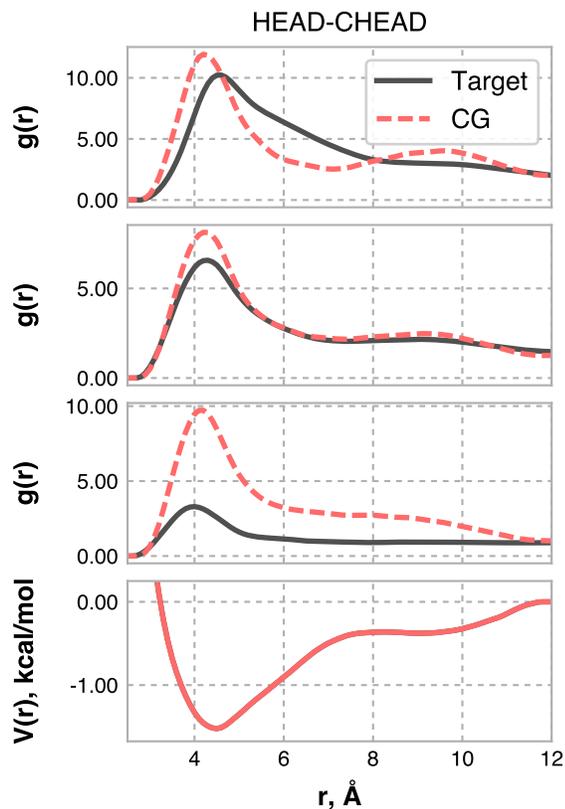


Figure 9.4: CHEAD–FHEAD RDFs and pair potential. From top to bottom: RDFs from the bilayer state; RDFs from the stacked bilayer state; RDFs from the bulk fluid NPT state; and the pair potential that yields the CG RDFs above.

was performed. The aim of these simulations was to validate the force field beyond the individual pair interactions by isolating binary lipid–lipid cross-interactions (*i.e.*, CER NS–CHOL and CHOL–FFA). To this end, two-component bilayers were simulated and compared to the corresponding atomistic simulations. Note that only CER NS–CHOL and CHOL–FFA bilayers are considered here, as the CER NS–FFA interactions were validated in Chapter 8.

As shown in Table 9.5, the CG force field accurately reproduces the structural properties of the bilayers simulated with the atomistic model. For both bilayers, the CG APL is within 2.5% of the atomistic APL, showing that the in-plane packing of the lipids is accurately described with the CG model. The thickness of the CG CER

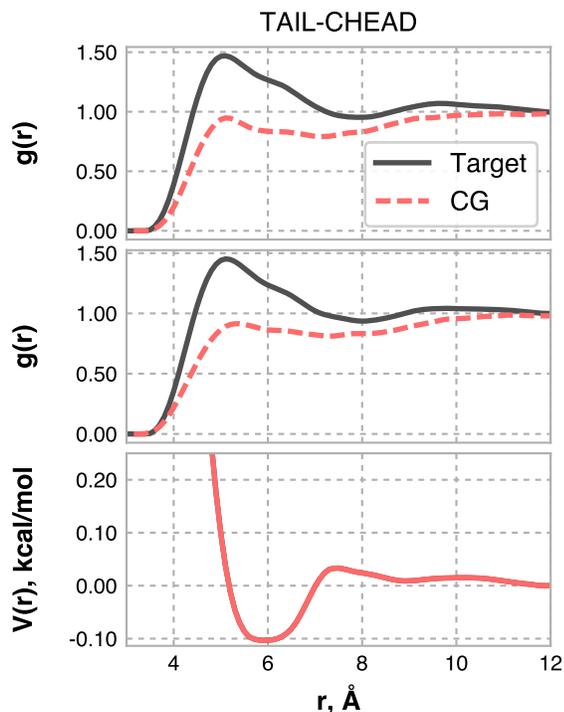


Figure 9.5: TAIL-CHEAD RDFs and pair potential. From top to bottom: RDFs from the bulk fluid NPT state; RDFs from the bulk fluid NVT state; and the pair potential that yields the CG RDFs above.

NS-CHOL bilayer shows almost perfect agreement with the atomistic bilayer, while the CG model underpredicts the thickness by  $\sim 7.5\%$ . In both cases, the nematic order of the lipid tails is underpredicted with the CG model; however, the relatively high values calculated for each bilayer with each model indicates that both models predict highly ordered bilayers. In both cases, the CG model overpredicts the tilt angle, although only slightly. These results illustrate the structural accuracy of the CG model derived in this work.

### 9.3.3 Self-Assembly Into Lamellar Structures

Using the validated force fields, the self-assembly of SC lipid mixtures into multilayer structures was investigated. The purpose of these studies was twofold: first, we ensure that an equimolar mixture of CER NS C24, CHOL, and FFA C24:0 will form a

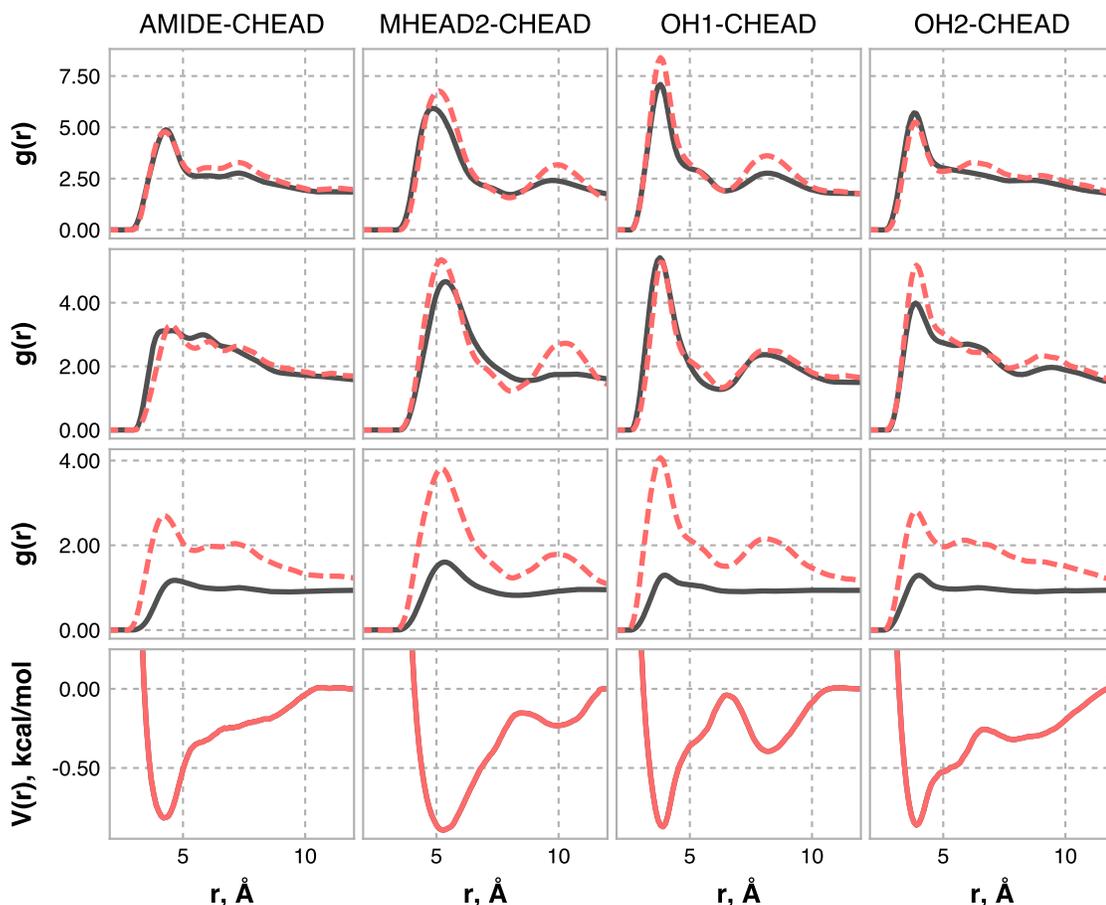


Figure 9.6: CERHEADS-CHEAD RDFs and pair potential. From top to bottom: RDFs at the stacked bilayer state; RDFs at the bilayer state; RDFs at the bulk fluid NPT state; and the pair potentials that yield the RDFs above. Target RDFs are shown by solid black lines, and CG RDFs are shown as the dashed red lines. Each column corresponds to a given pair, listed at the top of that column.

multilayer structure, as similar systems have been shown to form lamellar structures experimentally.<sup>14</sup> Next, we examine the effect of the CHOL composition on the self-assembled structures. To this end, following the self-assembly protocol described in Section 9.2.2, the self-assembly of these mixtures with varying compositions was investigated. First, an equimolar mixture was simulated. Next, systems with decreasing amounts of CHOL were simulated, with 20%, 14%, and 9% CHOL. Finally, mixtures with increasing amounts of CHOL were simulated, with 50%, 62.5%, and 75% CHOL.

Focusing first on the equimolar CER NS C24, CHOL, FFA C24:0 system, we find

Table 9.5: Properties of the atomistic (AA) and CG bilayers used for the validation of the CG force field. Note that the CHOL–FFA bilayer is composed of FFA C16:0, as this system contains the required hydrophobic mismatch to form a stable bilayer structure.<sup>13</sup>

System	APL, Å <sup>2</sup>	$d$ , Å	$S_2$ , Tilt, deg	
CER NS C24–CHOL				
AA	$38.8 \pm 0.3$	$47.8 \pm 0.1$	$0.958 \pm 0.007$	$8.7 \pm 0.8$
CG	$39.74 \pm 0.06$	$48.0 \pm 0.1$	$0.931 \pm 0.003$	$10.8 \pm 0.2$
CHOL–FFA C16:0				
AA	$29.92 \pm 0.09$	$35.9 \pm 0.2$	$0.951 \pm 0.004$	$10.8 \pm 0.5$
CG	$30.58 \pm 0.08$	$33.22 \pm 0.05$	$0.908 \pm 0.004$	$13.0 \pm 0.3$

that this mixture reproducibly forms multilayer structures. A representative multilayer structure is shown in Figure 9.7. Analysis of the CER conformations of the lipids in the middle leaflets reveals that  $33 \pm 3\%$  are in extended conformations, compared to  $\sim 15\%$  for pure CER NS C24 systems and  $\sim 25\%$  for equimolar CER NS–FFA C24:0 mixtures. Therefore, it appears that the presence of CHOL increases the tendency of the lipids to form extended configurations. As summarized in Table 9.6, when the amount of CHOL in the system is reduced, the fraction of CERs adopting extended conformations remains mostly constant initially, but drops to 28% with 9 mol % CHOL, providing more evidence that the presence of CHOL increases the fraction of extended CERs.

When the fraction of CHOL was increased to 50% and higher, multilayer structures did not form. Like the other systems, the lipids remain phase-separated from the water, and the hydrophobicity of the lipid tails allow well-ordered monolayer structures to form at the lipid–water interface. However, the lipids away from the lipid–water interface did not form lamellar structures, but rather more inverse-micellar structures with the lipid headgroups forming cylindrical tubes, as shown in Figure 9.8. This structure may be a result of the shape of CHOL, which is bulkier and more cone-shaped than the more cylindrical CER and FFA. This cone shape of CHOL leads to more

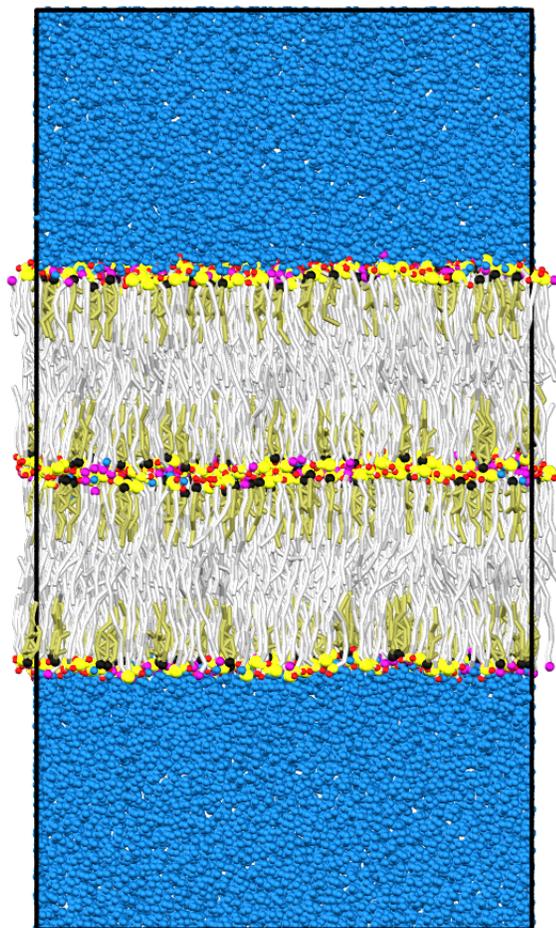


Figure 9.7: A self-assembled multilayer composed of equimolar amounts of CER NS C24, CHOL, and FFA C24:0. Lipids colored as follows: blue, water; silver, TAILS beads; yellow, AMIDE and MHEAD2 beads; purple, FHEAD beads; black, CHEAD beads; tan, CBODY beads; and the simulation cell is drawn in black.

empty space around the lipid headgroups, disrupting the planar packing needed to form a lamellar structure. When there is less CHOL in the system, this effect is not as pronounced. However, when the fraction of CHOL increases above a certain amount (50 mol % in this case), it may lead to lamellar instability. Further investigation into this effect is required to explain this phenomenon.

Table 9.6: Percentage of CER NS molecules within the middle leaflets that adopt extended conformations as a function of system composition. †Taken from Chapter 7. ‡Taken from Chapter 8.

CER:CHOL:FFA	% extended
1:0:0†	15
1:0:1‡	25
5:1:5	$28 \pm 2$
3:1:3	$33 \pm 3$
2:1:2	$34 \pm 2$
1:1:1	$33 \pm 3$

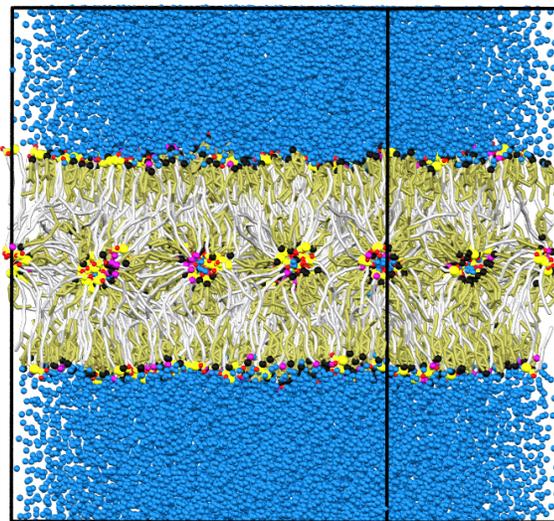


Figure 9.8: A snapshot of the inverse-micellar structures that formed within the lipid phase of the system with 62.5 mol% CHOL. Lipids colored as follows: blue, water; silver, TAILS beads; yellow, AMIDE and MHEAD2 beads; purple, FHEAD beads; black, CHEAD beads; tan, CBODY beads; and the simulation cell is drawn in black.

## 9.4 Bibliography

- [1] HADLEY, K. R. AND M<sup>c</sup>CABE, C. A structurally relevant coarse-grained model for cholesterol. *Biophysical Journal* **99**, 2010, pp. 2896–2905.
- [2] ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B., AND LINDAHL, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 2015, pp. 19–25.
- [3] ESSMANN, U., PERERA, L., BERKOWITZ, M. L., DARDEN, T., LEE, H., AND PEDERSEN, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 1995, pp. 8577–8593.
- [4] HOOVER, W. G. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A* **31**, 1985, pp. 1695–1967.
- [5] PARRINELLO, M. AND RAHMAN, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *Journal of Applied Physics* **52**, 1981, pp. 7182–7190.
- [6] KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O’CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D., AND PASTOR, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B* **114**, 2010, pp. 7830–7843.
- [7] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>c</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [8] JORGENSEN, W. L., CHANDRASEKHAR, J., MADURA, J. D., IMPEY, R. W., AND KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 1983, pp. 926–935.
- [9] ANDERSON, J. A., LORENZ, C. D., AND TRAVESSET, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 2008, pp. 5342–5359.
- [10] GLASER, J., NGUYEN, T. D., ANDERSON, J. A., LUI, P., SPIGA, F., MILLAN, J. A., MORSE, D. C., AND GLOTZER, S. C. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications* **192**, 2015, pp. 97–107.
- [11] MILANO, G., GOUDEAU, S., AND MÜLLER-PLATHE, F. Multicentered Gaussian-based potentials for coarse-grained polymer simulations: linking atomistic and mesoscopic scales. *Journal of Polymer Science Part B: Polymer Physics* **43**, 2005, pp. 871–885.

- [12] A Git repository for this package is hosted at <https://github.com/mosefhub/msibi>.
- [13] PARÉ, C. AND LAFLEUR, M. Formation of liquid ordered lamellar phases in the palmitic acid/cholesterol system. *Langmuir* **17**, 2001, pp. 5587–5594.
- [14] UCHIYAMA, M., OGURI, M., MOJUMDAR, E. H., GOORIS, G. S., AND BOUWSTRA, J. A. Free fatty acids chain length distribution affects the permeability of skin lipid model membranes. *Biochimica et Biophysica Acta - Biomembranes* **1858**, 2016, pp. 2050–2059.

## CHAPTER 10

### CONCLUSIONS AND OUTLOOK

The prevalence of using simulation for studying SC lipid systems is growing, and future advances in computing power will only make this trend continue. However, there are certain challenges that must be overcome if SC lipid simulations are to continue aiding researchers. The work described in this Thesis is an important step forward in addressing these challenges.

In Chapter 4, a common challenge in simulating SC lipid systems, namely, difficult equilibration, is addressed by proposing and validating a simulated tempering-inspired equilibration methodology for gel-phase bilayers. We show that this method efficiently equilibrates initial bilayer structures, allowing systems starting from different configurations to relax to the same structure. We also show that running at an elevated temperature, as is often done to avoid equilibration issues,<sup>1-3</sup> is less efficient than the proposed method for equilibrating systems. Therefore, the proposed method, or similar methods, should be used to equilibrate dense, gel-phase bilayer systems to ensure that the results are not overly influenced by the initial configuration. Also in Chapter 4, we investigated a series of SC lipid bilayers, showing that the main effect of a shorter CER fatty acid tail length is on the interior of the bilayer, where less tail length asymmetry leads to a smaller interdigitation region, yielding a thinner bilayer. The equimolar CER NS, CHOL, FFA C24:0 system was shown to accurately reproduce the lamellar structure of the SPP from a similar experimental system,<sup>4,5</sup> and therefore conclusions drawn from studies on this system also likely apply to the SPP of the SC lipid matrix. While preassembled bilayer structures are useful to study, they are not likely to undergo any significant lamellar transitions in practical simulation timescales; therefore, self-assembled structures are needed to probe the molecular organization of

lipids in the SC, requiring CG models.

In Chapter 5, a multistate extension to iterative Boltzmann inversion, MS IBI, is developed, showing that state-transferable CG force fields can be developed. Specifically, we show that a CG force field optimized with MS IBI more accurately reproduces the structural properties of both gel and fluid phase alkane monolayers than a CG force field optimized with the single-state method. As many chemical processes studied with CG molecular models are inherently multistate, it is important that the chosen force field is accurate across different states. MS IBI presents a suitable method to address this transferability issue of CG models.

In Chapters 6 to 9, MS IBI was used to develop a CG force field that includes the interactions between water, CER NS, FFA, and CHOL. Chapter 6 describes the development of the CG water model, where MS IBI allowed a model to be developed that captures the fluid structure, density, and surface tension of water. This result suggests that MS IBI-derived force fields can, at least to some degree, reproduce both structural and thermodynamic properties. In Chapter 7, a structurally-accurate CG model of CER NS was developed. This work in particular highlights a specific issue related to structurally-derived CG force fields, that is, that pair RDFs may not contain enough information to account for partitioning behavior. To address this issue, a simulated wetting procedure to tune specific interactions was proposed and used to fix the apparent hydrophobicity of the lipid tails. This wetting procedure is not specific to CG models, and hence could be used when developing new atomistic force fields. The self-assembly of SC lipid mixtures is investigated in Chapters 7 to 9, where it is observed that mixtures with specific compositions can form stable multilayer structures. The CERs are found to mainly adopt hairpin configurations, but a significant fraction adopt extended conformations. These results lend support to the proposed “sandwich” models of lipid organization in the SC,<sup>6,7</sup> but no definite claims can be made from simple mixtures that lack the acylceramides. However, this

work is a solid foundation upon which to build future work.

## 10.1 Recommended Future Work

To elucidate how the CER tail length influences the permeability of membranes composed of SC lipid mixtures, permeability calculations on these systems should be performed to allow relationships between the lipid structure and barrier properties to be uncovered. While some effort has been made on this front,<sup>8</sup> mixtures of CERs with different tail lengths have not been investigated, despite their connection to the SC lipid matrix, especially in diseased skin.

MS IBI should be tested for its range of validity. For example, IBI was shown to be inadequate to yield force fields that describe polymers in both bulk and confined environments.<sup>9,10</sup> A general set of principles that describes the range of transferability of MS IBI-derived force fields can yield would be useful. This information would aid researchers in choosing a force field development method, and ultimately save resources for more valuable endeavours.

To study the molecular organization of the SC lipid matrix with simulation, CG models of other ceramides need to be developed, especially the acylceramides. Preliminary tests of transferring the OH to make CER NP and CER AP have been promising thus far, but the validity of this approach should be confirmed. It has been shown that a LPP can form in a simple mixture of CERs (including EOS), CHOL, and FFA,<sup>11</sup> and thus this system is amenable to simulation. Examining the self-assembly of the LPP could provide a wealth of information into the molecular organization of native SC, so effort should be focused on developing a CG model of CER EOS. However, this task requires atomistic simulations of CER EOS, but it is currently unknown how to initialize simulations of CER EOS because a detailed description of its structure in lamellar systems is lacking. Therefore, collaboration with experimentalists should be pursued so that physically relevant atomistic simulations

can be performed to gather meaningful atomistic data with which to optimize a CG force field.

Finally, self-assembled structures of CG models should not be the end goal, but rather a means to an end. Since we are ultimately interested in connecting simulation results to experimental results, atomistic structures are necessary to compare properties on the atomistic level, *e.g.*, hydrogen bonding. Therefore, reverse-mapping back to the atomistic level from self-assembled CG structures should be pursued, allowing a more direct comparison to experimental systems (*e.g.*, through simulated neutron scattering length density profiles, as in Chapter 4). This would also allow the stability of the CG structures to be probed with more detailed models, providing more confidence that the self-assembled structures accurately represent reality. Reverse-mapped structures should also be used for permeability studies, since the multilayer structures more accurately represent the native SC lipid matrix. This would allow comparisons with existing permeability studies,<sup>8,12</sup> and could be used to validate the bilayer approximation currently used in simulation studies of SC lipids.

## 10.2 Bibliography

- [1] DAS, C., NORO, M. G., AND OLMSTED, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **97**, 2009, pp. 1941–1951.
- [2] DAS, C., NORO, M. G., AND OLMSTED, P. D. Lamellar and inverse micellar structures of skin lipids: Effect of templating. *Physical Review Letters* **111**, 2013, p. 148101.
- [3] DAS, C., NORO, M. G., AND OLMSTED, P. D. Fast cholesterol flip-flop and lack of swelling in skin lipid multilayers. *Soft Matter* **10**, 2014, pp. 7346–7352.
- [4] GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., VAN MECHELEN, J. B., DEMÉ, B., AND BOUWSTRA, J. A. Disposition of ceramide in model lipid membranes determined by neutron diffraction. *Biophysical Journal* **100**, 2011, pp. 1481–1489.
- [5] MOJUMDAR, E. H., GROEN, D., GOORIS, G. S., BARLOW, D. J., LAWRENCE, M. J., DEMÉ, B., AND BOUWSTRA, J. A. Localization of cholesterol and fatty acid in a model lipid membrane: a neutron diffraction approach. *Biophysical Journal* **105**, 2013, pp. 911–918.
- [6] BOUWSTRA, J. A., DUBBELAAR, F. E. R., GOORIS, G. S., AND PONEC, M. The lipid organisation in the skin barrier. *Acta Dermato-Venereologica; Supplement* **208**, 2000, pp. 23–30.
- [7] HILL, J. R. AND WERTZ, P. W. Molecular models of the intercellular lipid lamellae from epidermal stratum corneum. *Biochimica et Biophysica Acta - Biomembranes* **1616**, 2003, pp. 121–126.
- [8] GUPTA, R., SRIDHAR, D. B., AND RAI, B. Molecular dynamics simulation study of permeation of molecules through skin lipid bilayer. *The Journal of Physical Chemistry B* **120**, 2016, pp. 8987–8996.
- [9] BAYRAMOGLU, B. AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative Boltzmann inversion. *Macromolecules* **45**, 2012, pp. 9205–9219.
- [10] BAYRAMOGLU, B. AND FALLER, R. Modeling of polystyrene under confinement: Exploring the limits of iterative Boltzmann inversion. *Macromolecules* **46**, 2013, pp. 7957–7976.
- [11] DE JAGER, M. W., GOORIS, G. S., DOLBENYA, I. P., BRAS, W., PONEC, M., AND BOUWSTRA, J. A. The phase behaviour of skin lipid mixtures based on synthetic ceramides. *Chemistry and Physics of Lipids* **124**, 2003, pp. 123–134.
- [12] DAS, C., OLMSTED, P. D., AND NORO, M. G. Water permeation through stratum corneum lipid bilayers from atomistic simulations. *Soft Matter* **5**, 2009, pp. 4549–4555.

## APPENDIX A

### Analysis

#### A.1 General Structural Properties

Unless otherwise stated, the area per lipid (APL) is always taken to be the cross-sectional area of the simulation box divided by the number of lipids in each leaflet. Note that for self-assembly simulations, this method means the different monolayers (*i.e.*, bilayer/multilayer leaflets) will have different APL values.

Several quantities depend on the “direction” of the lipid tails, which is defined as the long axis of the lipid tail. The long axis of a lipid tail is taken to be the eigenvector corresponding to the minimum eigenvalue of the inertia tensor, which is calculated from the positions and masses of the atoms in the lipid tail. Two tails are defined for CER NS, one for each of the fatty acid (FA) and sphingosine (SPH) tails. For atomistic simulations, the FA tail is defined as the alpha carbon of the CER FA plus the next 15 carbons and their associated hydrogens. Note that this definition corresponds to the entire CER NS C16 tail, but neglects the terminal 8 methyl groups of the CER NS C24 tail so as to remove the influence of the low tail density region on the inertia tensor. The SPH tail is defined as the terminal 13 methyl groups of the SPH base. CHOL is treated as a single tail, and as such all atoms are included in the inertia tensor calculation. For FFAs, only the first 16 carbons (counting from the headgroup) are included in the calculation, again to remove the influence of the low density tail region on the inertia tensor. For CG simulations, the same definitions are used, in that the terminal 3 CG beads of the CER NS C24 FA and FFA tails are neglected in the inertia calculation and all CG beads are used to calculate the inertia tensor of CHOL.

The tilt angle of the lipid tails is defined as the average angle between the long

axis of each lipid tail and the bilayer/multilayer normal, *i.e.*, the  $z$ -axis. The area per tail (APT) is defined as the cross-sectional area of the simulation box divided by the number of tails in each leaflet (*i.e.*, 2 for CER, 1 each for CHOL and FFA), multiplied by the cosine of the tilt angle to map the area into a plane perpendicular to the lipid tails.

The nematic order parameter  $S_2$ , which provides a measure of the global order of the lipid tails, was calculated as follows.<sup>1</sup> The long axes of the lipid tails in each leaflet are used to construct a nematic tensor  $Q_{\alpha\beta}$  of the system, defined as

$$Q_{\alpha\beta} = N_T^{-1} \sum_{i=1}^{N_T} \frac{3}{2} \mathbf{u}_{i\alpha} \mathbf{u}_{i\beta} - \frac{1}{2} \delta_{\alpha\beta},$$

where  $N_T$  is the number of tails,  $u$  is the long vector of the lipid tail,  $\alpha$  and  $\beta$  are looping variable that correspond the coordinate axes (*i.e.*,  $x$ ,  $y$ , and  $z$ ) and  $\delta$  is the Kronecker delta. The largest eigenvalue of  $Q$  gives the nematic order parameter  $S_2$  of the components, and the corresponding eigenvector is the “director” of the system. Note, for systems that have a small population of lipids lying flat in between the individual leaflets, the lipids that are lying flat are not included in the APL, APT, tilt angle, or  $S_2$  calculations such that these metrics quantify the structure of the leaflets and not defects in the system. Also note that for bilayer and multilayer systems, the tilt angle and nematic order parameter of the lipid tails was calculated for each leaflet individually, and reported values are the average of the values for each leaflet.

Density profiles were constructed by histogramming the positions of particles (atoms in atomistic simulations or CG beads in CG simulations) in specific groups, weighted by the particle mass, along the relevant axis. The histogram was then normalized by the number of frames and the bin width to get average density profile per frame. Unless otherwise noted, a bin width of 1 Å was always used.

Bilayer thicknesses were calculated from the water density profiles. The water

density profiles were used to calculate the lipid–water interfaces, defined as the  $z$  value where the water density drops from the bulk value to  $1/e$  of the bulk value. The bilayer thickness was calculated as the distance between the lipid–water interface on either side of the bilayer. A schematic of this calculation is shown in Figure A.1.

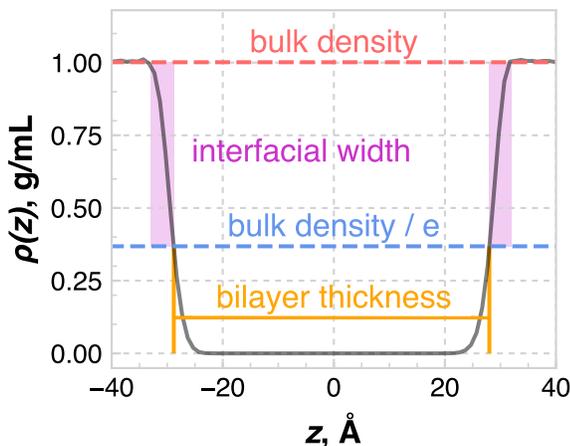


Figure A.1: Schematic of the bilayer thickness calculation. The gray line is the water density profile across the bilayer, from which the bulk value is calculated. The intersection of the water density profile and the  $\rho = \rho_{bulk}/e$  line defines the edges of the bilayer. The width of the lipid–water interface is shown as the width of the purple rectangles.

## A.2 Contact Angle

To tune the hydrophobic/hydrophilic interactions in the CG force fields, simulations were performed of droplets on lipid surfaces. The hydrophobicity/hydrophilicity of particular beads was tuned by scaling the CG pair interaction of interest (*e.g.*, TAIL–WATER) until the CG contact angle agreed with the contact angle from atomistic simulations. To calculate the contact angle, the density profile of an ideal spherical cap was fit to the density profile of the droplet fluid (*e.g.*, water) on the surface (*e.g.*, FFA TAILS), with the assumption that the fluid forms a spherical cap on the surface. The density profile fit gives the height  $h$  of the spherical cap, along with the radius of

the sphere  $R$ , from which the contact angle  $\varphi$  can be calculated as

$$\varphi = 90^\circ - \sin^{-1} \left( \frac{R - h}{R} \right).$$

### A.3 Hydrogen Bonding

Hydrogen bonding analysis was performed in the same manner in all Chapters. Hydrogen bond donors were identified as nitrogen or oxygen atoms bonded to a hydrogen, and hydrogen bond acceptors were defined as any nitrogen or oxygen atom. Thus, a hydrogen bond is represented as D—H $\cdots$ A, where D denotes the donor atom, H is the hydrogen involved in the hydrogen bond (and is covalently bonded to the donor atom), and A denotes the acceptor atom. A set of geometric criteria was used to determine if a hydrogen bond exists between a donor/acceptor pair, following the definition used by Ferrario *et al.*:<sup>2</sup> a hydrogen bond is considered to exist if 1) the donor–acceptor distance  $r_{AD} \leq 3.5 \text{ \AA}$ , 2) the hydrogen–acceptor distance  $r_{AH} \leq 2.6 \text{ \AA}$ , and 3) the hydrogen–donor–acceptor angle  $\theta_{HDA} \leq 30^\circ$ .

#### A.4 Bibliography

- [1] GUO, S., MOORE, T. C., IACOVELLA, C. R., STRICKLAND, L. A., AND M<sup>c</sup>CABE, C. Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *Journal of Chemical Theory and Computation* **9**, 2013, pp. 5116–5126.
- [2] FERRARIO, M., HAUGHNEY M., McDONALD, I. R., AND KLEIN, M. L. Molecular-dynamics simulation of aqueous mixtures: methanol, acetone, and ammonia. *The Journal of Chemical Physics* **93**, 1990, pp. 5156–5166.

## APPENDIX B

### Appendix to Chapter 4

#### B.1 Structural Properties

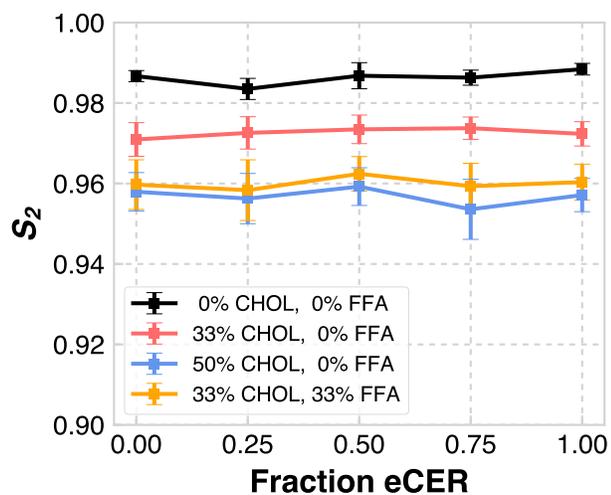


Figure B.1: Nematic order parameter  $S_2$  of the systems considered in this Chapter. Note that the abscissa gives the fraction of CER NS that is eCER, not the total fraction of eCER in the system.

Table B.1: Hydrogen bonds between lipid components for the ternary system with a 1:1 eCER:uCER ratio.

	eCER	uCER	CHOL	FFA
<b>eCER</b>	2.2	2.9	4.3	1.4
<b>uCER</b>	2.9	1.5	3.7	1.1
<b>CHOL</b>	4.3	3.7	0.0	0.7
<b>FFA</b>	1.4	1.1	0.7	1.0

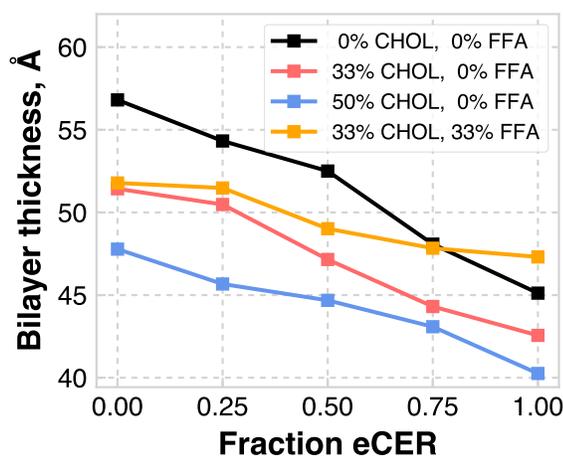


Figure B.2: Bilayer thickness, as calculated from the water density profile, as a function of system composition. Note that the abscissa gives the fraction of CER NS that is eCER, not the total fraction of eCER in the system.

Table B.2: Hydrogen bonds between lipid components for the 2:1 CER NS:CHOL system with a 1:1 eCER:uCER ratio.

	eCER	uCER	CHOL
eCER	7.3	12.1	6.4
uCER	12.1	6.9	5.0
CHOL	6.4	5.0	0.0

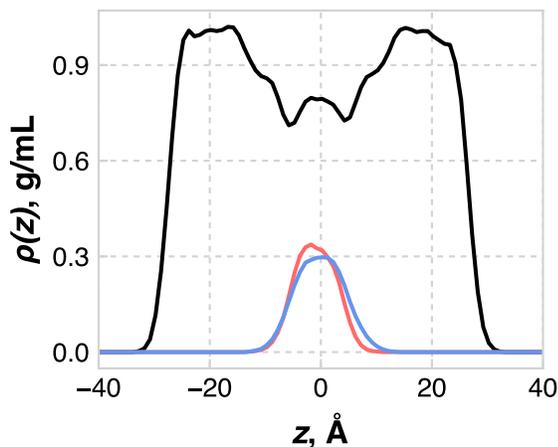


Figure B.3: Total lipid density profile (black line), and density profiles of the terminal part of the CER NS C24 fatty acid (red line) and FFA C24:0 tails (blue line).

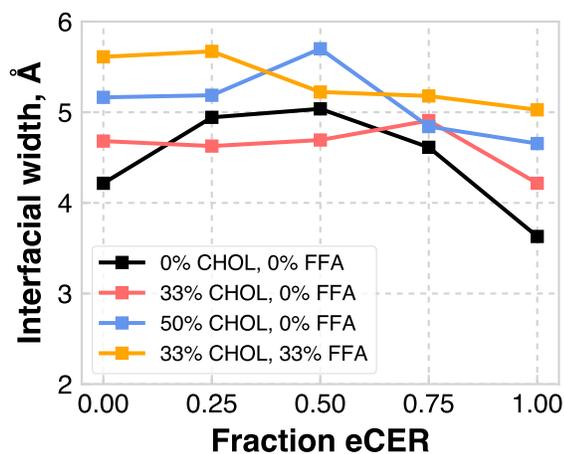


Figure B.4: Thickness of the lipid–water interface for each system considered. Note that the abscissa gives the fraction of CER NS that is eCER, not the total fraction of eCER in the system.

Table B.3: Hydrogen bonds between lipid components for the 1:1 CER NS:CHOL system with a 1:1 eCER:uCER ratio.

	eCER	uCER	CHOL
eCER	5.2	5.8	6.0
uCER	5.8	5.7	5.1
CHOL	6.0	5.1	0.1

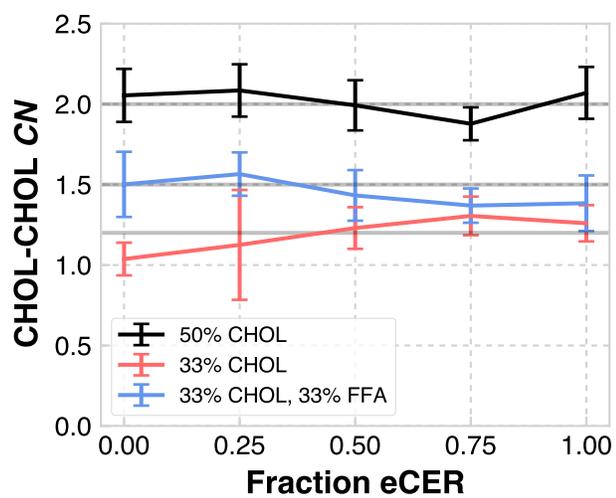


Figure B.5: CHOL-CHOL coordination number (CN) as a function of system composition. The light gray lines represent the CN value for random mixing in the bilayer plane. Note that the abscissa gives the fraction of CER NS that is eCER, not the total fraction of eCER in the system.

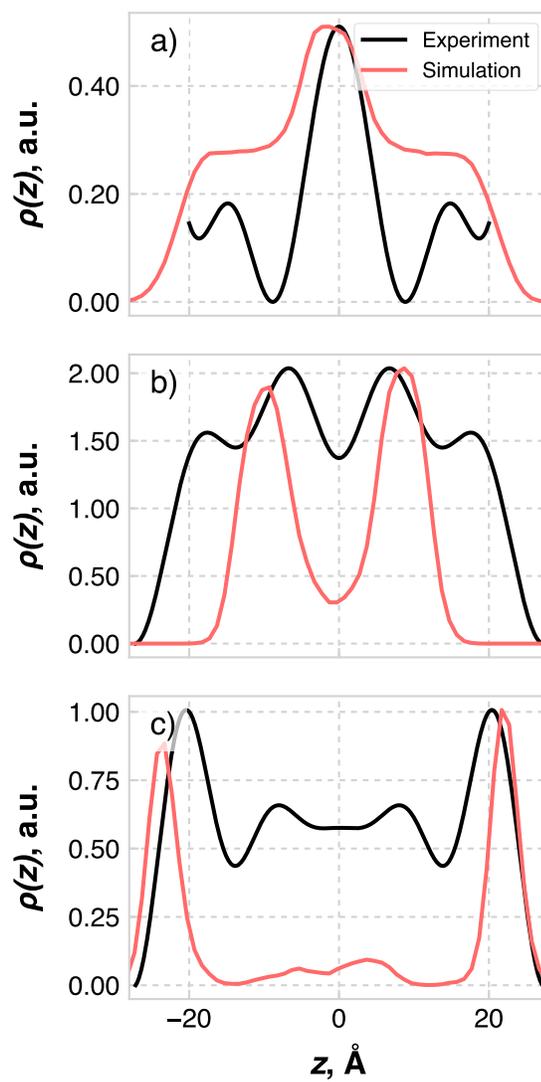


Figure B.6: Localization of different components calculated via neutron diffraction on selectively deuterated lipid samples (experiment) and from this work (simulation). a) uCER FA tail; b) CHOL tail; c) CHOL headgroup. Note that the curves were shifted such that the minima lie at  $\rho = 0$  and scaled to have the same height.

## APPENDIX C

### Appendix to Chapter 7

#### C.1 4-Site CER NS Headgroup RDFs and Pair Potentials

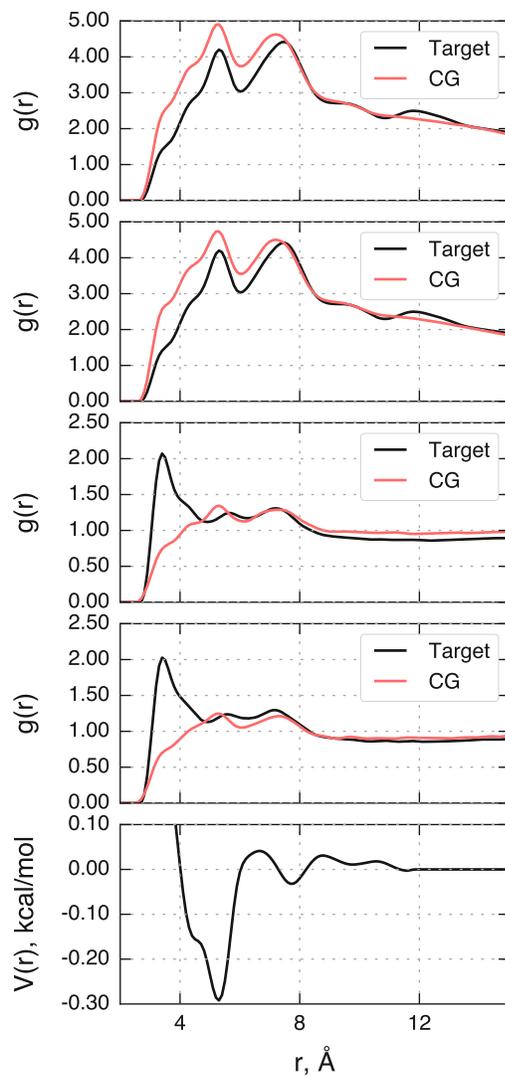


Figure C.1: AMIDE-AMIDE RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

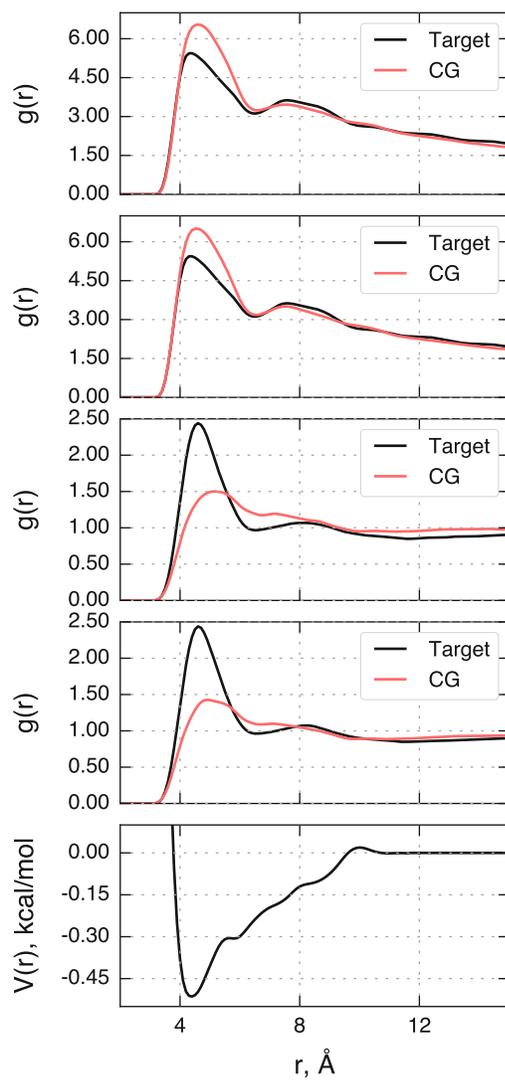


Figure C.2: AMIDE-MHEAD2 RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

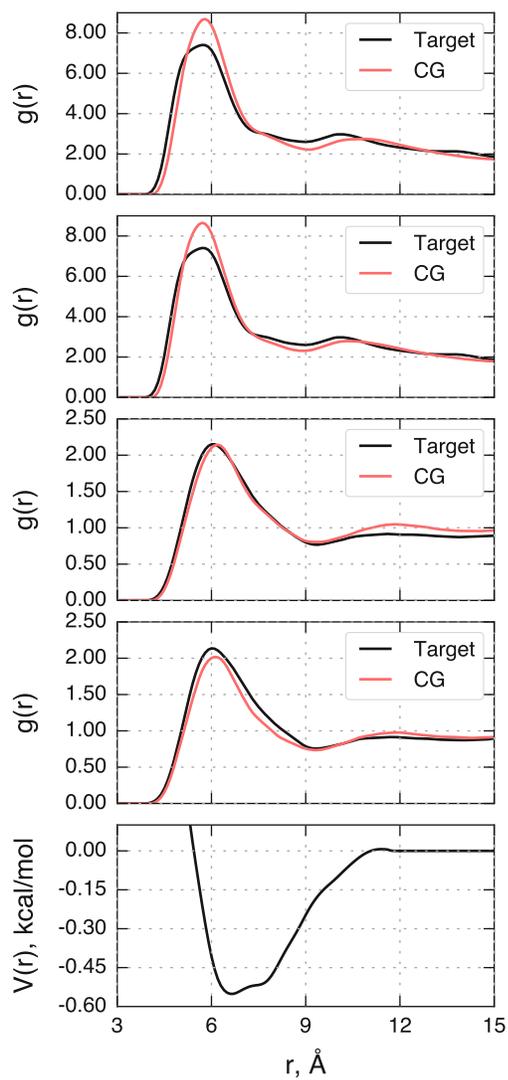


Figure C.3: MHEAD2–MHEAD2 RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

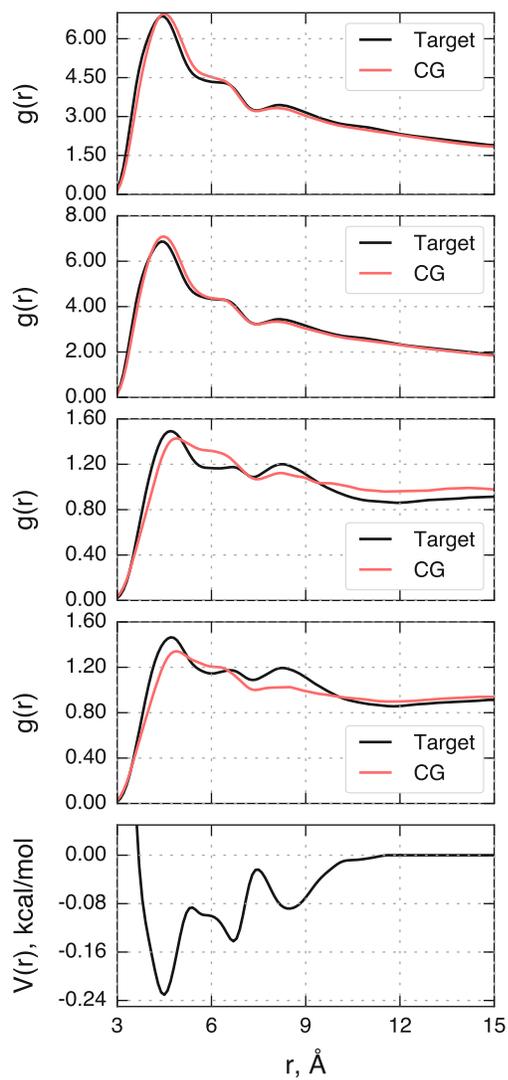


Figure C.4: MHEAD2–OH1 RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

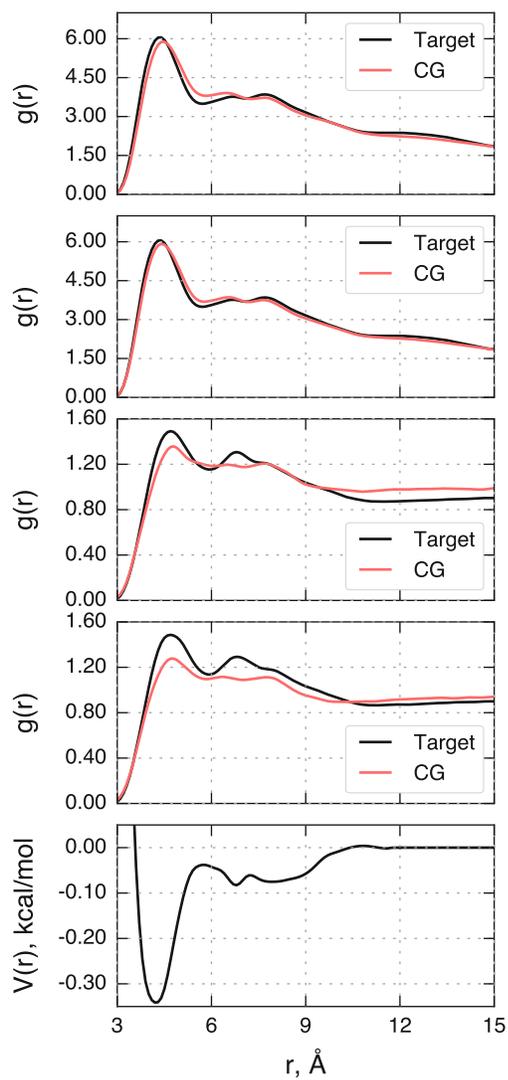


Figure C.5: MHEAD2–OH<sub>2</sub> RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

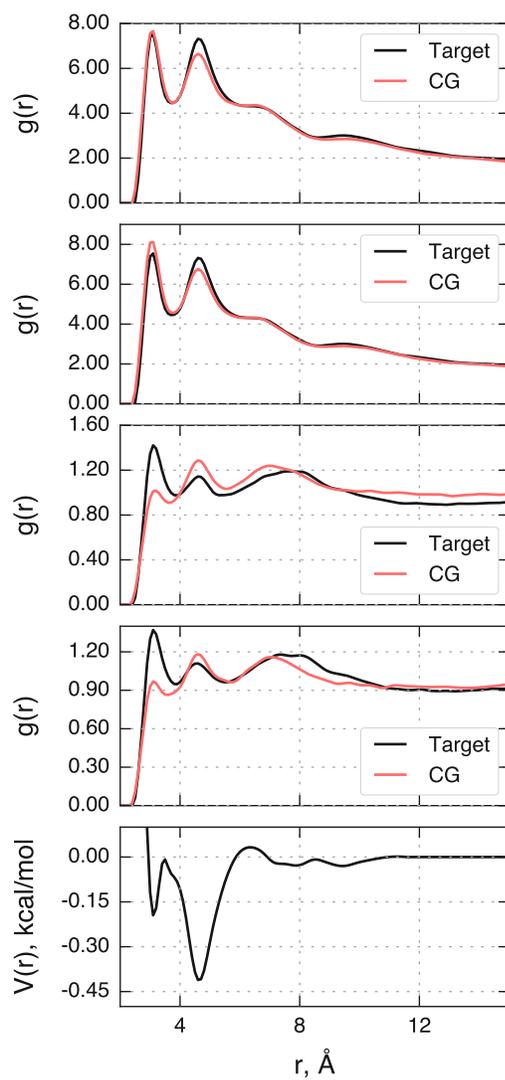


Figure C.6: OH1–OH1 RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

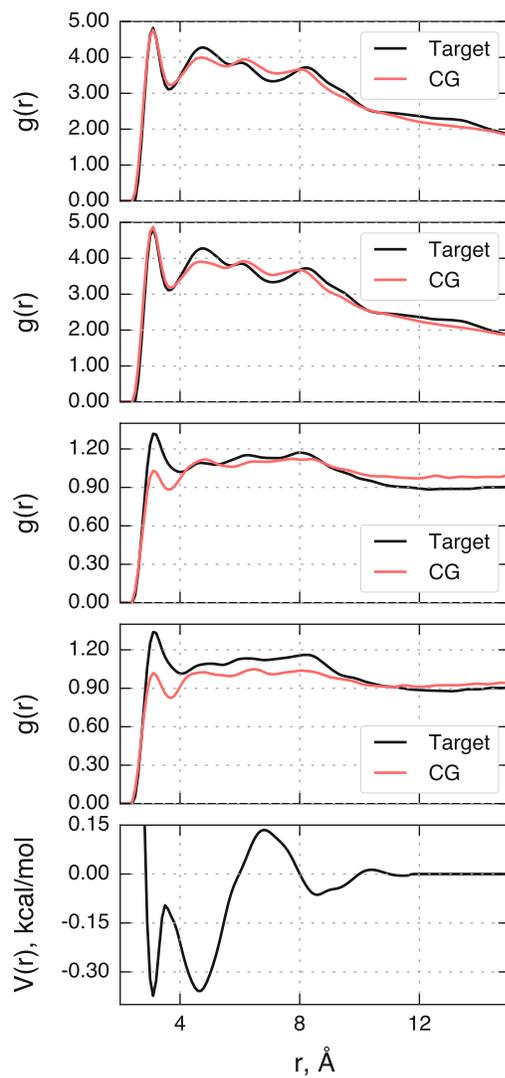


Figure C.7: OH<sub>2</sub>–OH<sub>2</sub> RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived pair potential.

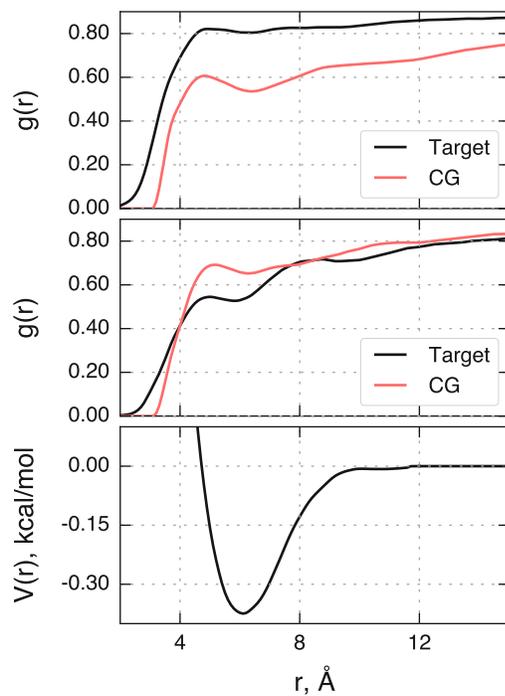


Figure C.8: AMIDE–WATER RDFs and pair potential. From top to bottom: CER headgroups in water RDFs; hydrated NPT bilayer RDFs; final derived pair potential.

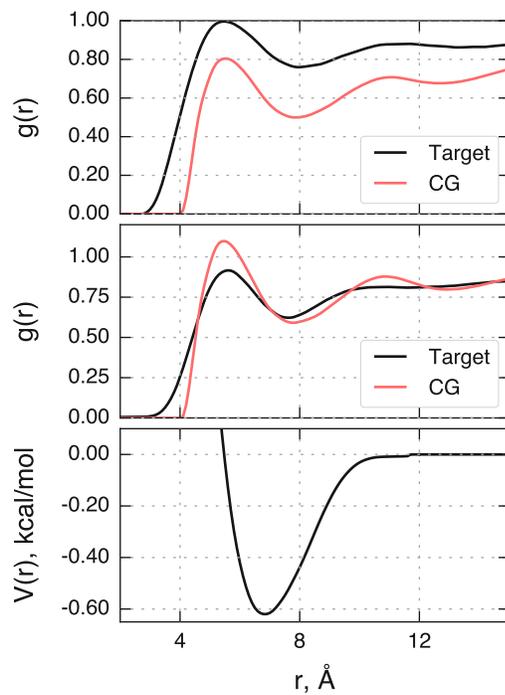


Figure C.9: MHEAD2–WATER RDFs and pair potential. From top to bottom: CER headgroups in water RDFs; hydrated NPT bilayer RDFs; final derived pair potential.

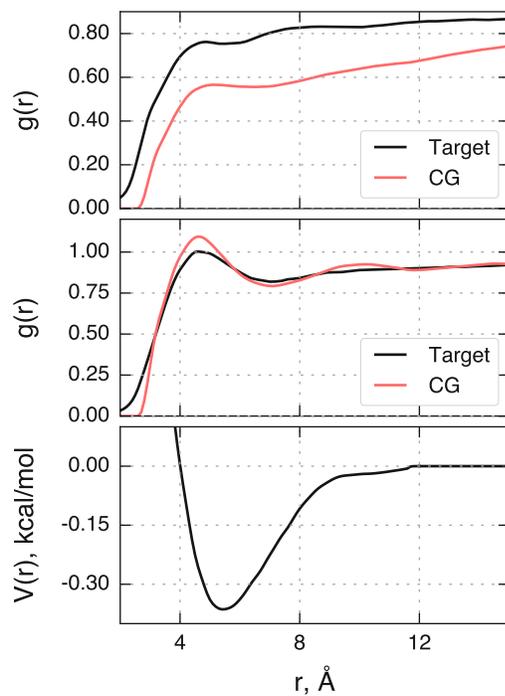


Figure C.10: OH1-WATER RDFs and pair potential. From top to bottom: CER headgroups in water RDFs; hydrated NPT bilayer RDFs; final derived pair potential.

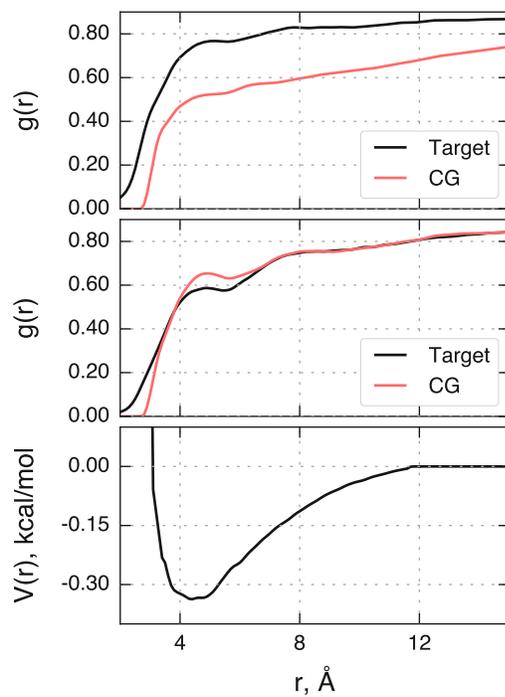


Figure C.11: OH2-WATER RDFs and pair potential. From top to bottom: CER headgroups in water RDFs; hydrated NPT bilayer RDFs; final derived pair potential.

## C.2 3-Site CER NS Headgroup RDFs and Pair Potentials

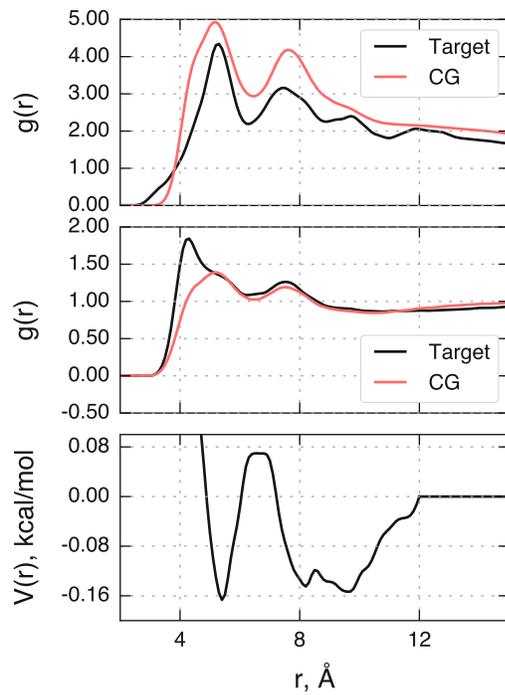


Figure C.12: NHEAD–NHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.

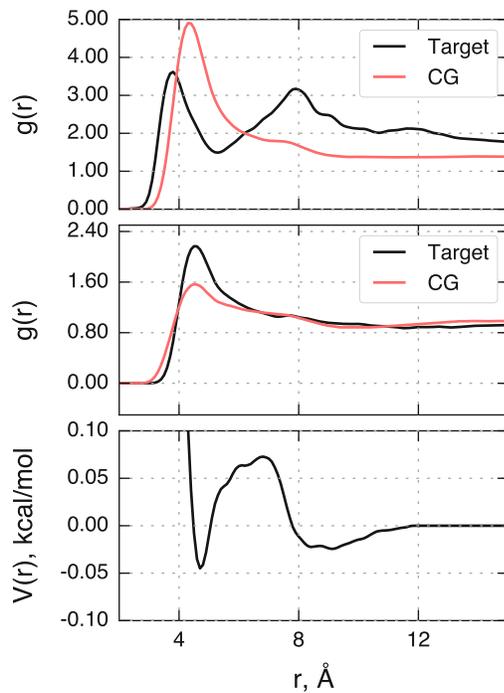


Figure C.13: NHEAD–MHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.

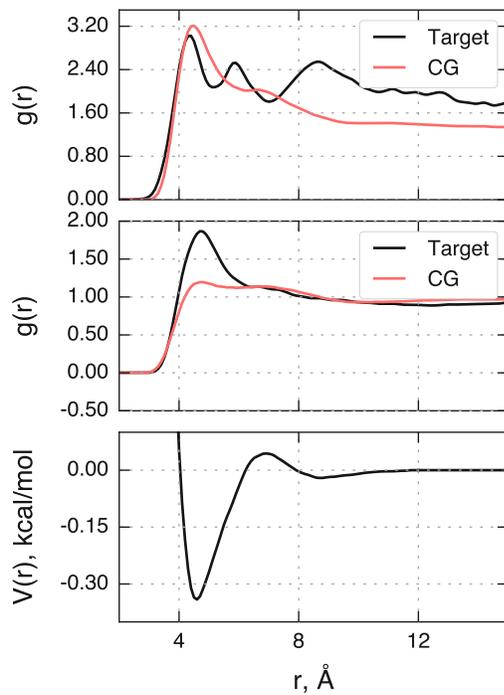


Figure C.14: NHEAD–SHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.

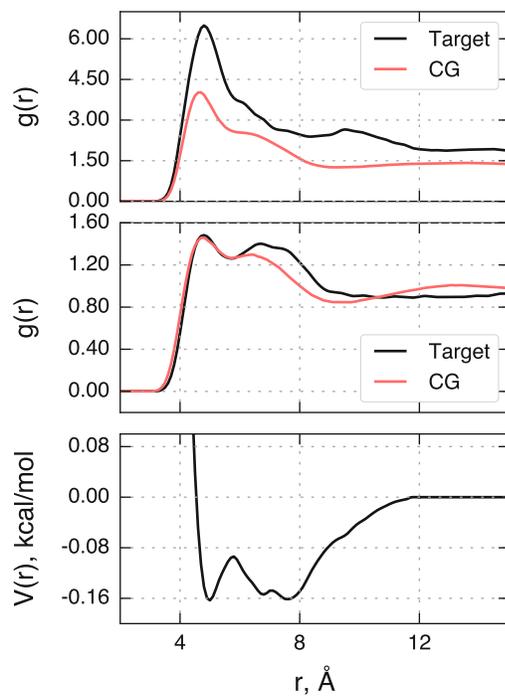


Figure C.15: MHEAD–MHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.

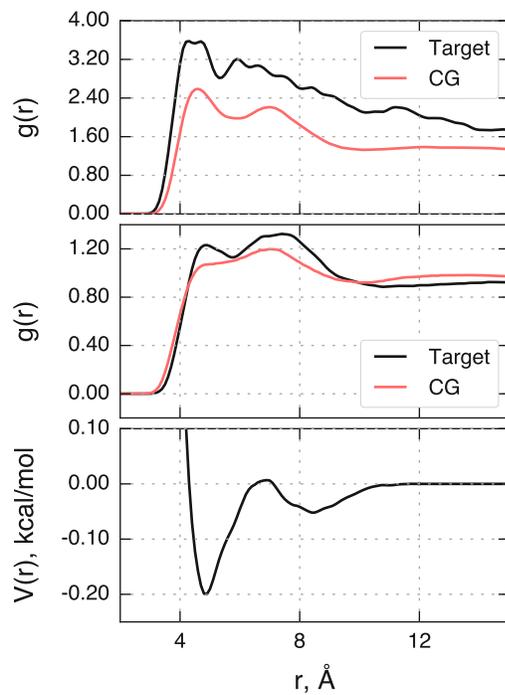


Figure C.16: MHEAD–SHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.

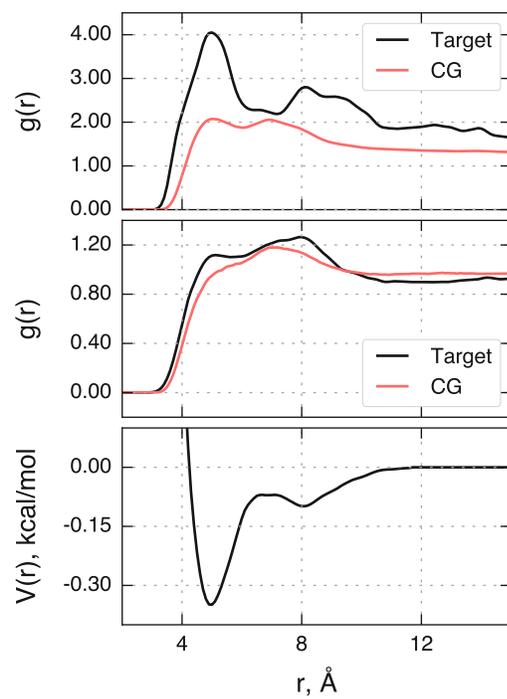


Figure C.17: SHEAD–SHEAD RDFs and pair potential. From top to bottom: dehydrated NPT bilayer RDFs; isotropic NPT fluid RDFs; final derived pair potential.