A Molecular Dynamics Study of the Self-Assembly of Human Skin Lipids

by

Kevin R. Hadley

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

December 2, 2009

Nashville, TN

Approved:

Dr. Clare McCabe

Dr. Peter T. Cummings

Dr. G. Kane Jennings

Dr. Scott A. Guelcher

Dr. Terry P. Lybrand

Dedicated

To

My Wife

ACKNOWLEDGEMENTS

I want to express my appreciation for people who helped me along the way. My advisor, Dr. Clare McCabe, has provided insight and has shown me what it takes to be a great advisor, researcher, and educator for when I enter the field of academia. Dr. McCabe, Dr. Annette Bunge, and the other members of my thesis committee (Dr. Cummings, Dr. Jennings, Dr. Guelcher, and Dr. Lybrand) contributed to the success of this research. In addition, I acknowledge the financial support of this work from the National Institute of Arthritis and Muscoskeletal and Skin Diseases through grant number R21 AR053270-02.

I would like to thank all of my close friends and family (especially my wife and parents, Bob and Bev). You are the people who made me who I am today. Without you, I wouldn't have the drive or the capacity to have completed. The members of the Cummings and McCabe labs also aided me along the way by providing advice, input, and distracting (although energizing) conversations. Specifically, I want to thank Dr. Patrick Redmill who took a lot of time out of his schedule to help me get started.

Although not related to this thesis, I would like to thank the people involved with helping me with my teaching certificate including Dr. Ken Debelak and Dr. Derek Bruff. In addition, I want to show my appreciation for the staff in the department (specifically Mary, Margarita, Rae, and Mark). They were always really nice to me and were always willing to help me with whatever I needed even if it didn't involve research.

Most importantly, I have to show my eternal gratitude towards my wife, Darcy Hadley. Through her support, her guidance, her love, and her infinite patience, I was able to finish with some sanity. I doubt I could have made it through without her.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

Dimethyl Sulfoxide ..... DMSO

Coarse-Grained ..... CG

Stratum Corneum ..... SC

Viable Epidermis ..... VE

Transmission Electron Micrograph ..... TEM

Short Periodicity Phase ..... SPP

Long Periodicity Phase ..... LPP

Ceramide ..... CER

Free Fatty Acid ..... FFA

Cholesterol ..... CHOL

Molecular Dynamics ..... MD

Radial Distribution Function ..... RDF

United Atom ..... UA

Reverse Monte Carlo ..... RMC

Reith, Putz, and Muller-Plathe ..... RPM

Periodic Boundary Conditions ..... PBC

van der Waals ..... VDW

Mean Squared Displacement ..... MSD

Chapter I

INTRODUCTION

The skin is the largest organ in the human body and serves as the interface with the surrounding environment[1]. The skin has many important functions including sensory perception, thermal regulation, excretion, and as a barrier against disease, chemicals, and internal water loss; the barrier function of skin prevents water from diffusing out of the body and allows survival in a non-aquatic environment. Understanding the mechanics of the skin's barrier function in order to develop treatment strategies for skin diseases and progress skin drug delivery research is the key motivation for this research.

Although the skin is an effective barrier, skin diseases disrupt the ability of the skin to prevent water loss and protect the body from chemical assault and infection[2-7]. An impaired barrier function has been linked to an altered molecular arrangement of the lipids present in the bilayers of the skin caused by an abnormal lipid composition[8].

With respect to other diseases in the body, there are many routes for delivering drugs to their target organs. Ingestion of drugs is a common way of administering treatment to the patient, but certain factors create difficulties in transporting the chemical to its desired location. The acidic environment of the stomach may not be compatible with certain drugs' chemical structure. Natural digestion may break down the drug rendering it inert before it reaches the treatment site. Also, digestion of the chemical may cause discomfort to the patient in question. Another problem is the difficulty in regulating the rate of drug absorption into the bloodstream, which may create problems in establishing dosage and frequency of intake. Finally, unless the diseased site is in the

digestive tract, it is impossible to localize administration to minimize the amount of drug needed for treatment.

Fortunately, the skin can overcome many of these problems with respect to proper drug delivery. One mechanism of skin drug delivery is through direct injections, but this is accompanied with problems of its own. Patients may forgo treatment to avoid the pain stemming from a needle puncturing their skin. Like ingestion, injections cannot regulate a constant administration, because the drug is delivered in one concentrated dose.

The other major mechanism of skin drug delivery is through passive diffusion. Placing a medium containing the drug to be delivered on the skin allows for localized delivery from the medium, and passive Fickian diffusion induces the drug to penetrate through the skin into the body. Transport across the skin is fairly consistent, so regulated administration of the drug is possible. Unless the medium containing the drug is a skin irritant, patient discomfort is minimized compared to inoculations and ingested drugs.

The skin barrier poses a serious hindrance for delivering benign chemicals into the body. The nature of some drugs either inhibits or absolutely prevents transport across the skin. As such, strategies are being developed to diminish the barrier function for delivery of drugs[9]. Some strategies involve disrupting the skin lipid structure with the addition of penetration enhancers like dimethyl sulfoxide (DMSO)[10].

Experimental research for skin drug delivery can be costly and time-consuming. A common approach is to develop transport models to predict absorption of either hazardous chemicals as a result of an accidental chemical spill or of a drug administered through a medium like a nicotine patch. Although cost-effective, these models are only

as accurate as the physical parameters (e.g. diffusion coefficients) describing the model, which can also be difficult and costly to obtain.

Barrier function is directly related to the lipid composition and packing of the molecules in the skin.[6,8] While experiments can measure lipid composition and barrier function, understanding the molecular-level packing through experiments is not feasible. Molecular simulations are capable of probing molecular level phenomena and may provide much needed insight into the connection between barrier function and skin lipid composition.

Skin diseases like psoriasis, atopic dermatitis, and ichthyosis exhibit depleted barrier function as a result of deficiency in one of many key structural skin lipids[2-7]. With simulations, a molecular model of the bilayer with a normal lipid composition and with a lipid composition associated with diseased skin could separately be developed and validated. With a model of healthy and diseased skin, exploring molecular packing and barrier function differences is possible and could be more cost-effective. As such, treatment strategies could be accurately and efficiently developed and tested. For example, a candidate for a new drug to aid psoriasis victims could be incorporated into our diseased model to observe its capability of restoring barrier function by measuring water diffusion rates. Initially testing drugs through computer simulations allows directed drug discovery and minimizes waste with respect to synthesizing and testing ineffective drugs.

The bilayer model for healthy skin can be applied to chemical penetration research, also. Measuring accurate diffusion parameters is possible, cost-effective, and efficient with the use of molecular simulations. To illustrate, drugs or hazardous

chemicals would be placed in the simulation, and diffusion across the structure would be measured and extrapolated to the large-scale diffusion model. Similar to testing treatment options for skin diseases, the effectiveness of penetration enhancers could be tested by observing their molecular level effects on the structure of the skin lipids. If an effective penetration enhancer is introduced to the bilayer model, the bilayer structure may be disrupted and exhibit high diffusion rates for water.

Although molecular simulations could provide insight into the molecular-level phenomena of skin barrier function, the ambiguity in the arrangement of the skin lipids hampers accurate research using molecular simulations. Experiments cannot determine exactly how the lipids are configured throughout the bilayer. As such, a major goal of this research is to use computer simulations to characterize the lipid arrangement and to provide molecular level insight into barrier function. The ultimate goal of the project is to develop a molecular skin lipid model capable of self-assembling into a bilayer exhibiting characteristics and features seen experimentally. After self-assembly, the molecular -level structure could be further studied in the fashion discussed throughout this chapter.

Although self-assembly can be observed through molecular simulations in an atomistic simulation (where every atom in the system is modeled and accounted for)[11], self-assembly in mixed lipid systems occurs on timescales not attainable on an atomistic level. Atomistic simulations are typically limited to running for 500 ns, but mixed lipid self-assembly may require at least 100 μs of simulation time. As such, simpler models are needed for modeling skin lipids capable of self-assembling into structures seen

experimentally. To simplify the model, groups of atoms can be represented by a collective site or bead on the coarse-grained (CG) level.

The level of complexity is immense for a simulation matching the experimental skin lipid composition and structure. As such, the goal of the work discussed later is to develop CG models of water and select skin lipids capable of self-assembling into an experimentally observed bilayer[12]. CG models have been optimized to match the structural features of their atomistic counterparts in a pure state and in simple mixtures. The robustness of these models has been tested in mixtures, crystalline states, and with respect to transferability. With respect to water, a novel method was employed to develop a CG water model with multiple waters mapped to single beads. As a final test of our CG methodology, the components are placed into a mixture and self-assembly is observed, providing evidence the molecular-level structure of the skin bilayers could be determined with CG models.

In chapters II and III, we summarize experimental findings and background molecular simulations, respectively. Our methodology for developing the CG models is detailed in chapter IV. The following chapters discuss the results from simulations of free fatty acids, cholesterol, and water, respectively. The heuristics for bilayer self-assembly and our comparison between simulation and experimental findings are found in chapter VIII. Finally, in chapter IX, the conclusions and required future work for this project are summarized.

Chapter II

EXPERIMENTAL BACKGROUND

The barrier function of skin is directly associated with the structure of the skin. Human skin contains two distinct regions as shown in Figure 2.1. The lower layer called the dermis contains a high concentration of blood vessels, sweat and oil glands, and hair follicles. The concentration of blood vessels in the dermis prevents the dermis from providing any hindrance to water loss or to substances spreading throughout the body[13]. The top layer of skin is called the epidermis and is composed of dead skin cells and decaying skin cells as seen on the right in Figure 2.1. The epidermis provides the largest contribution towards the skin's barrier function,[8] because of the components in the two layers of the epidermis: the viable epidermis (VE) and the stratum corneum (SC).

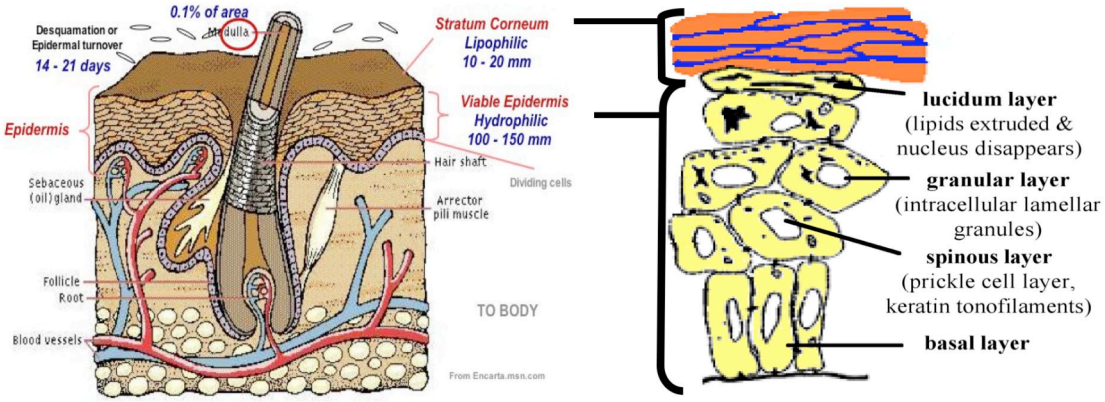


Figure 2.1 – Skin cross-section from Encarta.msn.com (left)[14] and cellular transformation of VE cells to corneocytes and lipid matrix (right)[15].

In the VE, the skin cells are slowly undergoing apoptosis, a form of programmed cell death. As the cells traverse the VE, the contents of the cell break down until the outer membrane bursts and the cell's contents seep out forming a complex lipid matrix in the SC, where absolutely no phospholipids exist unlike most cell membranes[15,16]. The large concentration of phospholipids makes the VE a hydrophilic medium.

Although the VE acts as an efficient barrier to hydrophobic molecules[17], the SC is the main contributor to barrier function due to its physical and chemical components and features. In the SC, a rich lipid matrix surrounds large hexagonal corneocytes composed of keratin, the protein found in fingernails, as shown in Figure 2.2. The corneocytes are the remnants of dead skin cells from the VE and are hypothesized to be effectively impenetrable and prevent any sort of passage of water or foreign agents[18]. An interwoven network of fibril proteins composes corneocytes and allows for absorption of water and swelling of the dead skin cells. As an aside, this phenomenon is the reason skin becomes wrinkled after a long soak in the tub.

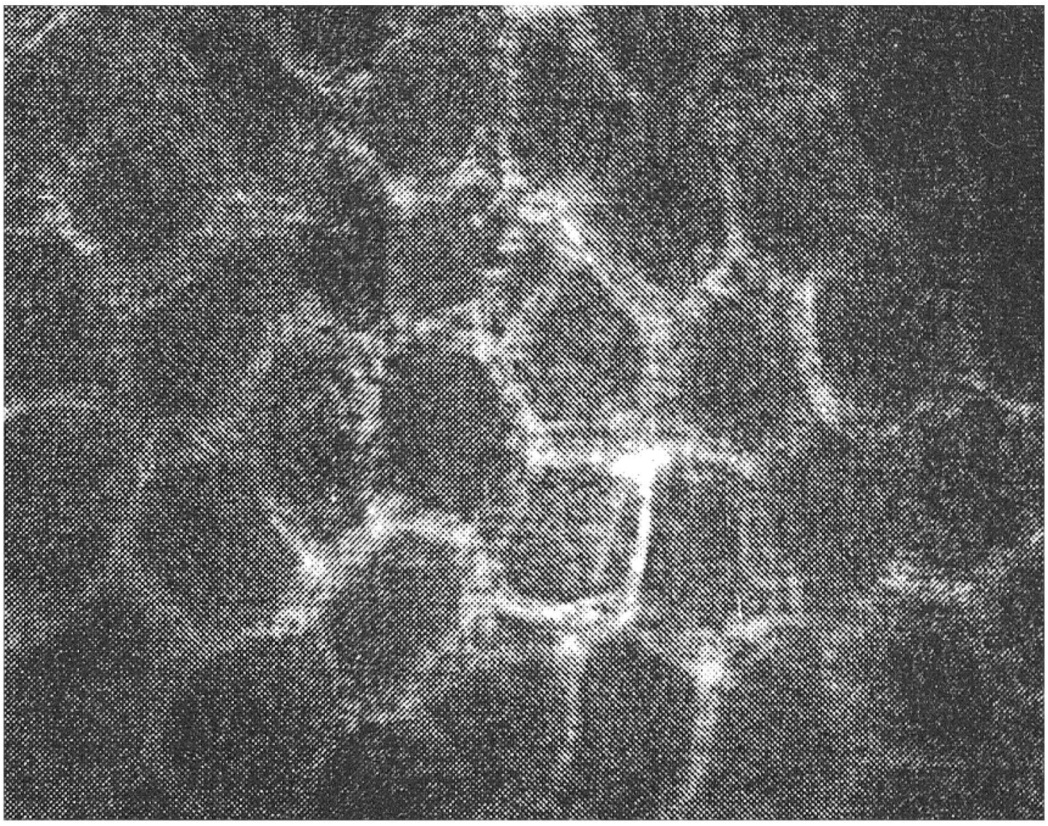


Figure 2.2 – Confocal micrograph with fluorescent stain of the stratum corneum with corneocytes (dark regions) embedded in the lipid matrix (light regions)[19].

In order for molecules to traverse the skin they must diffuse through the lipid region between the corneocytes as seen in Figure 2.2 because of the impenetrability of the corneocytes[18]. This region is a complex matrix of lipids arranged in a repeating lamellae crystalline structure. In contrast to the lipid membrane in skin, most biological membranes exist in a gel or fluid-like phase[20]. In addition, the matrix of SC lipids contains no phospholipids, which causes the matrix to be hydrophobic, as opposed to other membranes with high concentrations of phospholipids[15]. The high degree of hydrophobicity provides resistance to water diffusion and the densely packed crystalline structure restricts movement of other molecules, in general. For molecules with a high

affinity for the lipid matrix, the tortuous diffusion pathway acts as the final component of the efficient barrier function of the SC. As shown in Figure 2.3, the corneocytes embedded in the lipid matrix are arranged in a brick-and-mortar configuration preventing unidirectional diffusion.

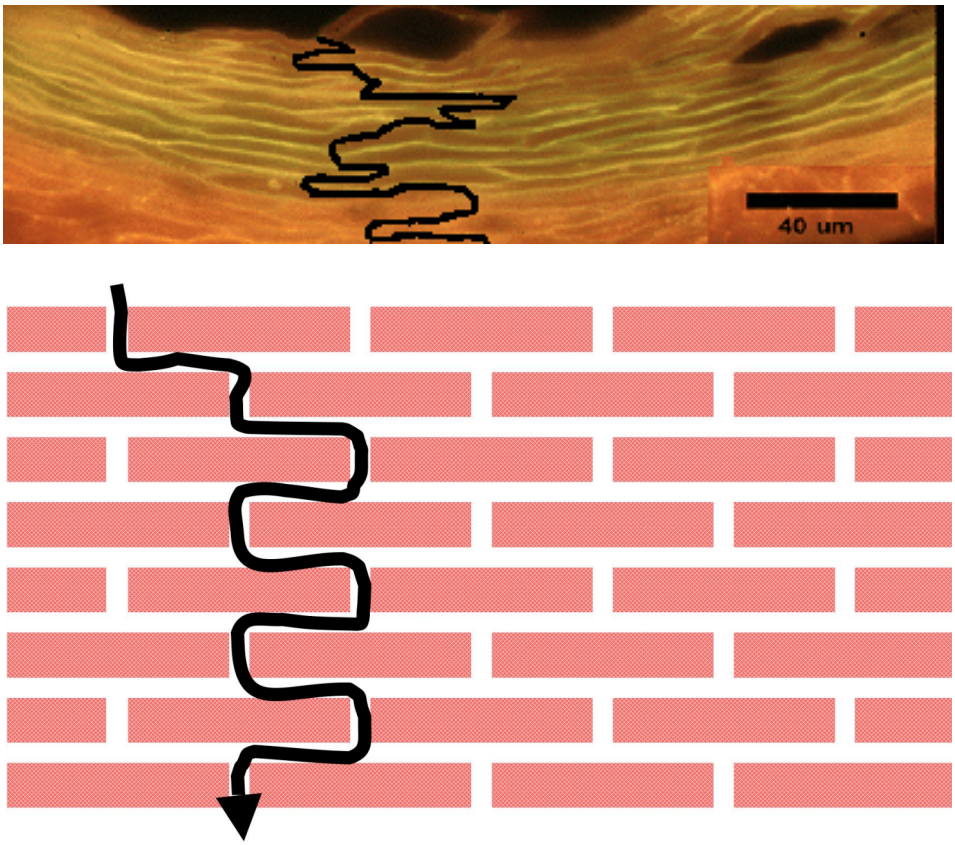


Figure 2.3 – Image of the stratum corneum with the diffusion pathway indicated by a solid line[21]. The tortuous pathway is analogous to what is seen in a brick-and-mortar arrangement.

The intercellular lipid lamellae of the SC are depicted in a transmission electron micrograph (TEM) in Figure 2.4. The lipids were treated with a strong oxidizer, RuO_4, which reacts with polar groups but not hydrophobic groups. This results in dark bands for polar regions and lucent bands where hydrocarbons are present. Unlike most lipid

bilayers in biological systems, there are two different bilayer structures in the SC as shown by the broad-narrow-broad lucent bands. Small and wide-angle X-ray diffraction data of human, pig, and mouse SC all show two periodic orthorhombic crystalline lamellar phases existing between the corneocytes, where the narrow phase has a periodicity of 6 nm while the broad phase has a periodicity of 13 nm[22,23]. The existence of a short periodicity phase (SPP) in conjunction with a long periodicity phase (LPP) is thought to be key to barrier function in the SC[24,25].

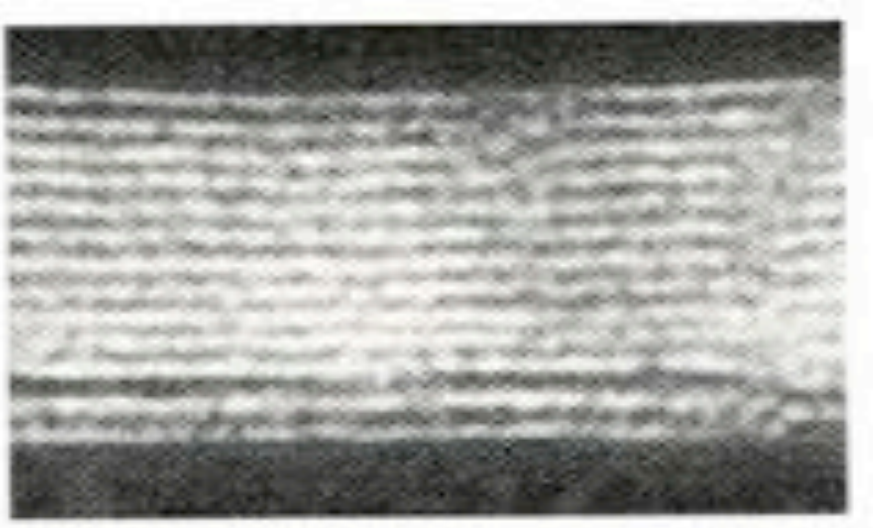


Figure 2.4 – TEM image of SC bilayer indicating broad-narrow-broad lucent bands[26].

2.1 Composition of the Lipid Matrix of the Stratum Corneum

Figure 2.5 shows the composition of lipids at different levels of the epidermis. The figure illustrates how the lipid composition shifts from mainly phospholipids in the basal layer to ceramides (CER), cholesterol (CHOL), and free fatty acids (FFA) in the stratum corneum. Most phospholipids are charged, making a portion of the molecule very hydrophilic. If these phospholipids were not converted to ceramides and fatty acids

before reaching the SC, water would have a higher diffusion rate or, in other words, the skin would exhibit impaired barrier function.

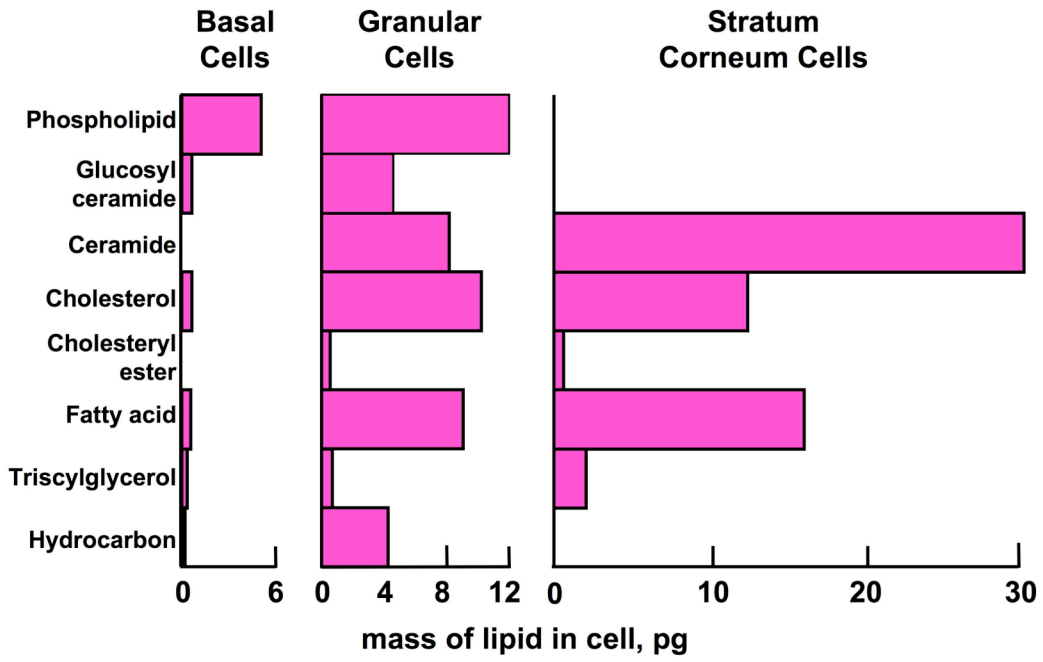


Figure 2.5 – Lipid composition of the layers of the epidermis[15].

The SC is mainly composed of free fatty acids, cholesterol, and ceramides, but there are other important components such as water, cholesterol sulfate, cholesteryl esters, and glucosylceramides.[27] By weight percent, CER represent 45-50% of the SC, CHOL 25%, and the FFA's 10-15%, with less than 5% of other lipids (e.g. cholesterol sulfate).[27] The tails of both CER and FFA contain a varied number of carbons, so an accurate molar composition is difficult to measure. Because of the variation in CER and FFA tail length, CHOL has the largest presence in the SC of any single lipid in the matrix. Based on the average mass of each class of lipid, investigators have reported a

1:1:1 molar ratio between the three main lipids, while others have estimated it to be 2:2:1 CER:CHOL:FFA.[27]

2.1.1 Free fatty acids (FFA)

Free fatty acids are molecules with one acid group (COOH) attached to a long carbon chain as shown in Figure 2.6. FFA's are the only major component of the SC capable of being ionized, but, at the pH of the skin (5.4), the head group remains protonated[28]. While the majority of the FFA's in the SC are completely saturated, monounsaturated molecules exist in negligible quantities[29]. FFA are denoted by the following notation, CΩ:Λ, where Ω is the number of carbons in the chain and Λ is the degree of unsaturation, e.g., hexadecanoic acid is denoted by C16:0.

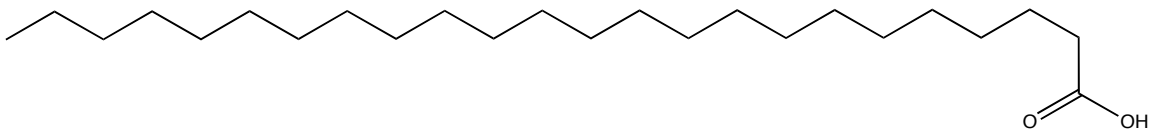


Figure 2.6 – An example of a free fatty acid found within the stratum corneum, C24:0.

The reported composition of FFA in human skin with respect to chain length is not consistent in the literature[29,30]. Wertz and Downing reported FFA with 16 to 26 carbons[30], while Norlén *et al.*[29] reported chain lengths ranging from 20 to 30 carbon atoms, with the majority having an even number of carbons. We consider the composition reported from Norlén *et al.* since their work is the most recent. Table 2.1 contains median mole % of the various acids as reported by Norlén, from which we note C24:0 and C26:0 dominate with approximately 39 and 23 mole percent, respectively.

TABLE 2.1 - Acid composition in SC[29].

Acid	C20:0	C22:0	C24:0	C25:0	C26:0	C27:0	C28:0	C29:0	C30:0
Mean mol%	5	11	39	10	23	3	8	1	2
Median mol%	4	11	39	10	23	3	8	0	2

The presence or absence of FFA in the SC affects the structure in several interesting ways. First, it is known a lack of FFA in native SC is a sign of ichthyosis, where the skin is abnormally dry and scaly (like a fish) because of impaired barrier function and results in a loss of body moisture at an elevated level compared to healthy skin[5]. Experimentally a shift from an orthorhombic crystal structure to a hexagonal structure is observed when FFA is absent and the reverse is true when an appropriate FFA amount is present[31]. This crystal change is also associated with a decrease in lipid density when going from an orthorhombic to a hexagonal configuration[31,32]. Why a hexagonal crystal structure reduces barrier function and how FFA induces an orthorhombic crystal structure is unknown, but the models discussed later in this work may eventually be able to elucidate this molecular level behavior.

2.1.2 Cholesterol (CHOL)

Cholesterol plays an important role in the properties of biological membranes and membrane dynamics. Specifically, it is known cholesterol is required for cellular viability and proliferation[33,34] and that similar sterols, such as ergosterol, cannot replace cholesterol in mammalian cells[35]. Therefore, the specific structure of cholesterol is believed to be related to the function of cholesterol in membranes and bilayers.

The molecular structure of cholesterol has both a rigid multi-ring arrangement and a flexible tail which is unique amongst other membrane lipids. Within the ring structure,

the presence of chiral methyls results in contrasting smooth and "rough" faces (shown in Figure 2.7), and the alcohol group contributes to the hydrophilic component of the molecule's amphiphilic properties and drives the orientation of cholesterol in bilayers[36].

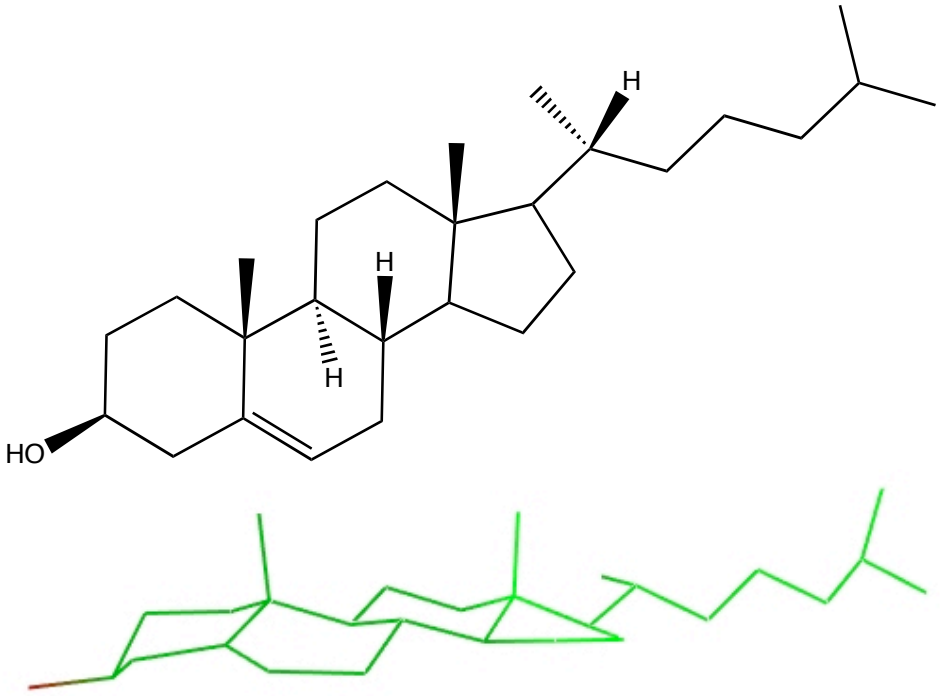


Figure 2.7 – Molecular structure of cholesterol (above) and a depiction of cholesterol (devoid of hydrogens) with the "rough face directed upwards (below).

The unique structure of cholesterol contributes to its properties in mixtures and in pure systems[34]. For example, Yeagle found cholesterol "rigidifies" bilayers by increasing order and decreasing molecular motion in lipid bilayers as observed by a change in phase from a fluid state to a gel state[37]. In pure bilayers, sections of the lipid tail become kinked (deviates from a trans-gauche conformation) and "push" that section of the neighboring tail producing a *domino effect* where all of the tails exhibit a similar kink[37]. In a bilayer containing cholesterol, if a tail tries to become kinked, the ring structure of

cholesterol is not flexible enough to change conformation; therefore, the rigid rings prevent the original kink and induce order in the bilayer structure. However, Yeagle also notes cholesterol inhibits bilayers from becoming too ordered,[37] because cholesterol is an "impurity" in the bilayer due to the mismatch in shape between cholesterol and typical membrane lipids (such as sphingomyelin or phosphocholine). By inhibiting the degree of structure in a bilayer, cholesterol induces a degree of fluidity in bilayers as seen in the SC lipid matrix, which contributes to the flexibility of human skin[38].

2.1.3 Ceramides (CER)

Ceramides are long double-tailed lipids composed of a sphingoid attached to a fatty acid tail via an amide linkage. To date, nine separate ceramides have been identified in the stratum corneum[30,39,40], as shown in Figure 2.8.

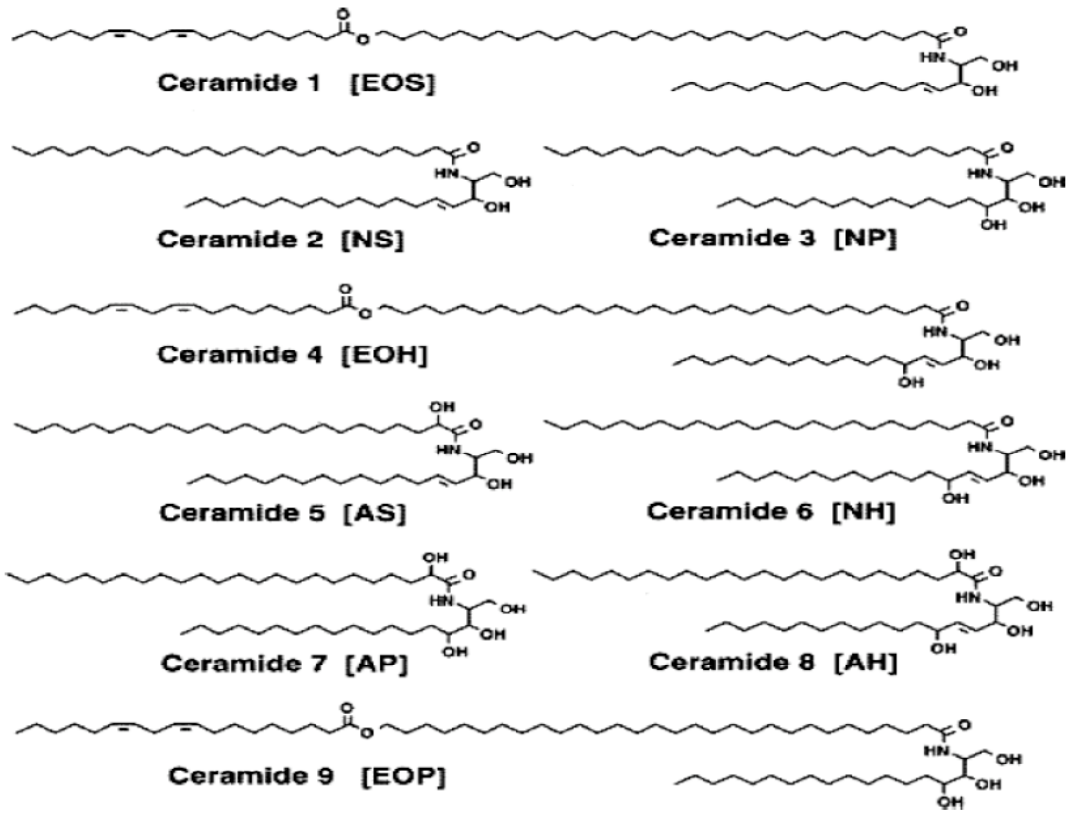


Figure 2.8 – Molecular structures of the nine CER from Kuempel *et al.*[41]. Notations are derived from Motta *et al.*[3]: ester-linked (E), ω-hydroxy (O), non-hydroxy (N), and α-hydroxy (A) fatty acids; sphingosine (S), phytosphingosine (P), and 6-hydroxysphingosine (H).

The sphingoid tail (the bottom half of the molecules found in Figure 2.8) in the SC has an average of 18 carbons in length. In CER1, CER2, and CER5, the sphingoid is composed of a sphingosine with a double bond at the fourth and fifth carbons. Similarly, CER4, CER6, and CER8 contain a sphingosine but with the addition of an alcohol group attached to the sixth carbon. In the remaining ceramides (CER3, CER7, and CER9), the sphingoid is composed of a phytosphingosine as well as an alcohol attached to the fourth carbon (as opposed to the double bond found in this location on a sphingosine). The acid tail component of ceramides is similar to the free fatty acids in the SC because the chain

varies in length throughout human skin with an average length of 24 carbons and is typically saturated. The major difference is the alcohol of the acid is replaced with an amide link and CER5, CER7, and CER8 contain an additional alcohol on the alpha carbon. Also CER1, CER4, and CER9 contain a linoleic acid ester linked to a 30-32 carbon ω-hydroxy fatty acid and are distinguished as acylceramides.

Table 2.2 lists the mean and average weight percents of the CER's in the SC as reported by Wertz and Norlén[42]. CER8 and CER9 are not included as they were only recently identified[30,39,40]. Ponec *et al.* discovered the presence of CER9 and recorded the concentration of the different ceramides found in human skin; however, their methods could not differentiate between CER5 and CER8, so the combined concentration was reported as shown in Table 2.3.[39]

TABLE 2.2 - Ceramide composition in SC as reported by Wertz and Norlén[42].

Ceramide	CER1	CER2	CER3	CER4	CER5	CER6	CER7
Mean wt%	8.3	20.8	19.4	8.2	24.2	5.7	22
Average wt%	6-12	12-25	11-34	4-9	17-27	4-11	10-27

TABLE 2.3 - Ceramide composition in SC as reported by Ponec *et al*.[39]

Ceramide	CER1	CER2	CER3	CER4	CER5/8	CER6	CER7	CER9
Median mol%	8.3	20.5	18.0	5.0	19.7	8.6	12.9	6.4

The presence of CER1 in the SC lipid matrix is of utmost importance as seen experimentally.[24,31,43-46] Diseased skin from patients with atopic dermatitis whose lipid matrix lacks ceramides (especially CER1)[2] and from patients with psoriasis whose skin has the same overall concentration of ceramides but a depleted level of CER1[3,4], showed severe depletion of barrier function.

2.1.4 Other lipid components

Other lipids present in the SC include cholesteryl-esters, cholesterol sulfate, glycerol, and glucosylceramides, but these do not affect bilayer structure and have been shown to be nonessential in regulating barrier function[47]. While they play important roles in the SC (e.g. regulate proliferation of the skin), they are beyond the scope of this work and will not be considered further.

2.1.5 Water

Due to its hydrophobicity, only small amounts of water exist in the SC and are tightly bound to the hydrophilic groups of the lipids[48]. In experiments, the lipid matrix structure is maintained with increasing amounts of water[22]. With addition of extreme amounts of water, the corneocytes swell and the lipid matrix undergoes a phase change and self-assembles into large globules similar to spherical micelles[9].

2.2 Lipid Role as Observed from In Vitro Self-Assembly Experiments

Elucidating the role each lipid plays in the self-assembly and structure of the SC lipid matrix is a difficult task with purely experimental methods. One shared goal is to reproduce the long periodicity phase (LPP) and the short periodicity phase (SPP) found in native stratum corneum by observing self-assembly from mixtures of the lipids found within the stratum corneum.

Early experimental mixtures were prepared with lipids extracted from human or porcine skin. McIntosh *et al.* mixed extracted lipids and were able to reproduce the 13 nm LPP and the 6 nm SPP and showed self-assembly does not require the presence of

proteins[43]. Bouwstra *et al*. used x-ray diffraction to study similar lipid mixtures and observed a repeating lamellar crystalline structure with a LPP and a SPP with spacing of 13 and 6 nm, respectively[49].

Although extracted ceramides provide the most accurate reproduction of the LPP and the SPP, synthetic ceramides allow consistent control and reproducibility between different experiments. Consequently, synthetic ceramides have provided more insight into the role each lipid plays and the requirements for LPP and SPP formation irrespective of the numerical accuracy of the LPP and SPP distances.

2.2.1 Role and properties of cholesterol in the stratum corneum

The role of cholesterol in the SC lipid matrix is very similar to the role it plays in other lipid bilayers. Cholesterol is believed to be key to the lipid matrix structure; in the absence of cholesterol in CER mixtures, the LPP cannot be observed[43,47]. In addition, a lack of cholesterol produces a highly crystalline state, so the lipid matrix would hypothetically be rigid and not have the flexibility of healthy skin[47]. Cholesterol's rigid ring structure induces order into the tails of the free fatty acids and the ceramides, but its mismatch in molecular shape prevents too much ordering and induces fluidity[37]. To elucidate this phenomenon, Zbytovska and coworkers looked at differing the composition of cholesterol in mixtures of cholesterol, CER6, C16:0, and cholesterol sulfate[50]. As the concentration was increased from 10 to 30 mol% cholesterol, the membrane density decreased by approximately 20% and the area per lipid increased by approximately 20% since the increased amount of cholesterol prevented efficient packing and allowed for increased bilayer flexibility.

In addition, Zbytovska *et al*. found increasing the temperature increased the solubility of cholesterol in SC lipid mixtures[50]. This finding was substantiated by the work of Pare *et al*. who observed a low-ordered bilayer of C16:0, water, and CHOL at an elevated temperature, but complete phase-separation at room temperature[12]. According to the authors, at temperatures above the FFA's melting point, bilayer formation was more energetically favorable than complete phase separation between the two lipids.

2.2.2 Role and properties of free fatty acids in the stratum corneum

Free fatty acids play the important role of increasing the saturation limit of cholesterol, but cholesterol is believed to be at its saturation limit in native skin[5]. In mixtures of SC lipids, cholesterol forms domains with higher concentrations of cholesterol and, in mixtures without an appropriate amount of FFA, will phase separate into a crystalline cholesterol-rich phase[5,43].

In addition, the presence of FFA in SC lipid mixtures promotes structures similar to what is observed in healthy human skin. An orthorhombic crystal structure forms only after FFA's are included into the bilayer structure[5,31]. However, it is difficult to determine if the presence of FFA's directly affects the crystal structure, or if the increased CHOL concentrations influence the crystal structure. Bouwstra and coworkers have also observed FFA's extend the LPP distance of SC lipid mixtures composed of extracted human ceramides from 12.2 nm to 13.0 nm[31,47,51].

Many have reported a match between the length of the ceramide tail and the length of the FFA is important for stabilization and the prevention of phase separation[46,47,49,52]. Lafleur and coworkers were able to show a mismatch between the

length of the FFA and cholesterol prevented formation of a low-ordered bilayer phase[12,53,54]. Kessner *et al.* were unable to observe LPP in a synthetic mixture with CER1, CER6, CHOL, and C16:0 and attributed it to choosing C16:0 over C24:0 as the FFA component[46]. Ruettinger *et al.* reported the effect of changing FFA chain length on a model bilayer containing synthetic CER6 as the main bilayer constituent[52]. They concluded CER serves as the main structural component of the bilayer, but the FFA has effects on the structure such as bilayer thickness. With shorter FFA's (C16:0-C18:0), there is only one phase present in a CER6 bilayer containing a 16-carbon fatty acid tail. With increasing FFA length (C24:0-C26:0), a fraction of the acids phase separate and form their own crystalline phase indicating long-tail FFA's are not very soluble in mixtures composed of short-tail CER's. In addition, the bilayer thickness decreased from 45.6 Å to 43.7 Å with C26:0 compared to C16:0. The authors hypothesize the tails of the longer acids extend far beyond the terminal carbon of the CER, so the tails bring the bilayer leaflets closer together to prevent vacuum-like regions and to maximize the van der Waals contacts.

2.2.3 Role and properties of ceramides in the stratum corneum

Bouwstra and coworkers have done extensive work on utilizing synthetic or commercially available CER to duplicate the structure of the SC lipid matrix[24,25,44,51,55,56]. They have determined the inclusion of an acylceramide as well as some level of CER1 is required to observe a LPP. Varying the concentration of other CER components has a negligible effect compared to varying the concentration of CER1[44,56]. Bouwstra *et al.* prepared different mixtures using CER1 as the main component, but switched the tail

structure of the ester-linked acid between the different mixtures[51]. They found if the ester-linked acid was fully saturated, no LPP could be formed, and if the ester-linked acid had one degree of saturation, a weak LPP phase could be detected, but a separate liquid phase also formed. As such, LPP formation is dependent upon the presence of an ester-linked acid at the end of the fatty acid tail of the CER and the ester-linked acid must contain two cis-double bonds (as seen in linoleic acid) to properly mimic the phase behavior and structure of the SC lipid matrix.

In [44,56], the authors shifted the composition of the acylceramide component from CER1 to CER9, incrementally. Even with small amounts of CER1, the LPP was present, but they produced a much weaker signal in the diffraction patterns when compared to mixtures with only CER1. However, when CER9 was the only acylceramide present, no LPP was detected and CER9 phase separated from the mixture into a pure CER9 crystal. As such, an acylceramide is not only required for LPP formation, but the head group architecture is important as well.

Lastly, de Jager *et al.* elucidated the importance of a varied fatty acid length distribution in the CER components[25,44,55]. When the number of carbons in the tail is uniform, the components form a highly crystalline phase. However, if there is a non-uniform distribution of carbons, the mismatch in chain length allows the structure to be more fluid and flexible, as is seen in native human skin. For this reason, the ceramides within the lipid mixtures must have a distribution of tail lengths if the structures formed are to appropriately reproduce the lipid matrix structure of the SC.

2.3 Experimental SC Structure Models

By integrating the observations from lipid structure, the composition, and the role each lipid plays in structure and self-assembly, a lot of conceptual *models* concerning the molecular level arrangement of the lipids in the SC have been proposed. These *models* all serve the purpose of describing the molecular level structure and the phase state of the lipid matrix to help explain the experimental observations and understand concepts like chemical penetration and impaired barrier function in skin diseases.

One of the earliest *models* for the structure of the SC lipids is the stacked monolayer *model* of Swartzendruber *et al*.[57] In this model representation, which can be found in Figure 2.9, there are no FFA present and the unbound CER molecules are all in the splayed conformation (fully extended) versus the hairpin conformation commonly found in phospholipid bilayers. This arrangement was proposed to make sense of the broad-narrow-broad bands seen in TEM images.

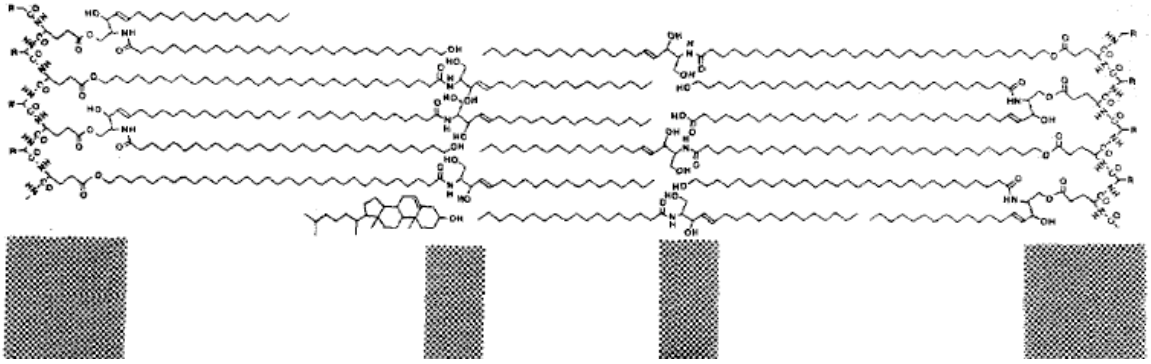


Figure 2.9 – Stacked monolayer model with the band lengths fitting the lengths from the TEM image[57].

The next major arrangement *model* was the domain mosaic *model* of Forslind (shown in Figure 2.10)[58]. In this model, there is no attempt to fit the broad-narrow-broad bands but rather the overall structure of the entire lipid region. In this *model*, the author

suggests a mix of two different phases: a crystalline phase and a more fluid phase separated by grain boundaries. The grain boundaries are believed to be where water diffuses through the lipid region.

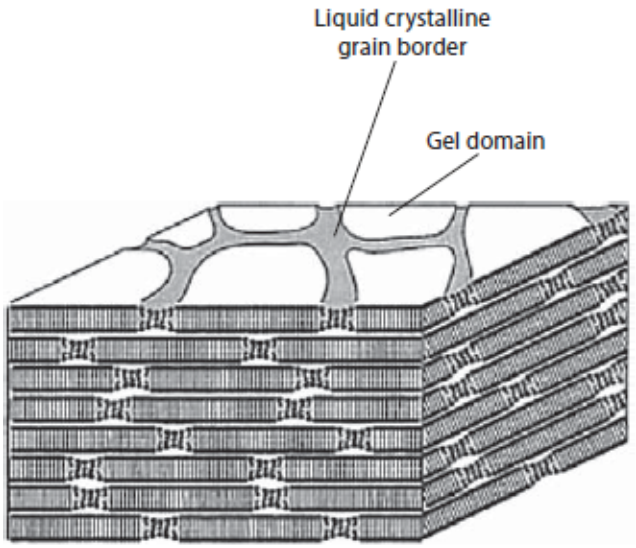


Figure 2.10 – Domain mosaic model[58] with the two phases and the grain boundaries taken from[59].

Work from the groups of Bouwstra[32] and Wertz[41,60] independently proposed the sandwich *model*, which reproduces the broad-narrow-broad bands where acylceramides serve as the key component to maintaining the LPP[32,41,60]. The sandwich *model*, shown in Figure 2.11, is a trilayer arrangement where the outside hydrophobic regions are in a more crystalline state and the inside region is more fluid-like.

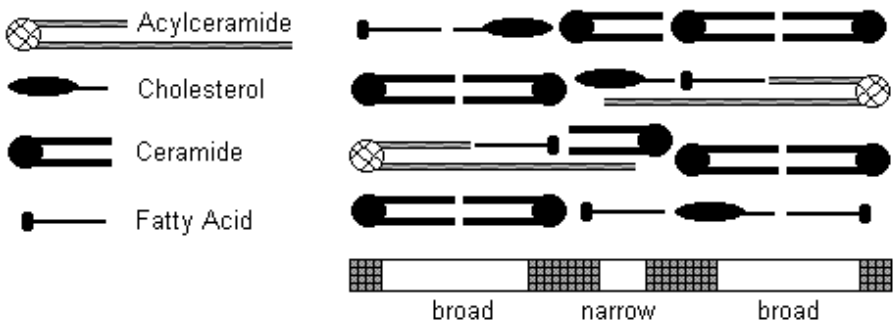


Figure 2.11 – Sandwich model with the broad-narrow-broad bands reproduced[32,41,60].

Norlén developed the single gel phase *model* as an alternative to the above *models*[59]. In this *model*, the lipids form a homogenous gel phase, splayed conformations and interdigitation of the CER are accounted for, and the lipids are layered as seen in Figure 2.12. To support his *model*, Norlén states the rate of water diffusion through the skin is not compatible with a two-phase medium as seen in the sandwich model and the domain mosaic *model*. In addition, it is argued a non-uniform phase would be very sensitive to environmental changes such as temperature and humidity, and as such, this design would be unnaturally inefficient. With respect to the sandwich *model*, Norlén believes interdigitation, tilting of the tails, and a mix of hairpin and splayed configurations for the ceramides need to be accounted for because these molecular arrangements are physically possible[61].

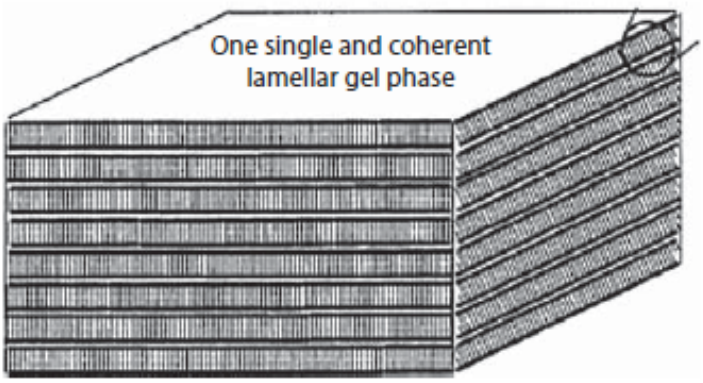


Figure 2.12 – Single gel phase model of the structure in the intercellular space of the SC[59].

More recently, Kiselev has developed a slightly different model called the armature reinforcement model, where there is an explicit mix of ceramides in splayed and hairpin conformations[62]. The splayed molecules serve as a binder between two bilayer regions as shown in Figure 2.13. With the addition of more water, a traditional bilayer conformation is observed, and it was hypothesized the splayed conformation is not energetically favorable in the presence of a significant amount of water. This is further supported when one considers the crystal structures observed by Dáhlen and Pascher, where the molecules adopt a V-shape or, in other words, a partially-splayed conformation[61,63]. Although the ideas generated by this model could be valid, the source of the information comes from lipid mixtures not representative of the native SC lipid matrix. Specifically, the mixtures used for Kiselev's experiments did not employ any acylceramides and contained only one ceramide type, CER6. In addition, the fatty acid of the ceramide contained 16 carbons, while the majority of SC ceramides contain 24 carbons in their fatty acid tail.

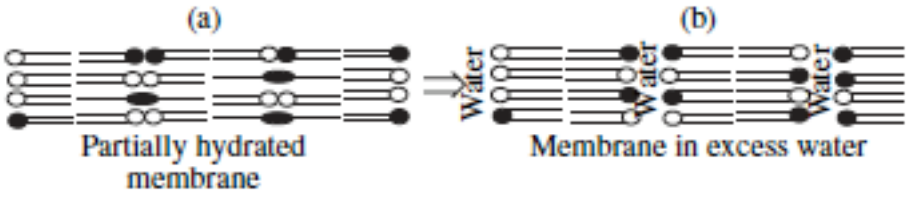


Figure 2.13 – Armature reinforcement model of the stratum corneum lipid matrix. Structural transition from a dehydrated state (a) to a fully hydrated state[62].

Of all of the models, only the sandwich model reproduces the composition and both the LPP and SPP length scales. The concepts found within the armature reinforcement model could be incorporated into sandwich model making the model more robust. With only experimental results, validating the models is difficult, and accurate description of how the lipids are organized on the molecular level cannot be explicitly detailed with current experimental methods.

Although experiments have provided a large amount of insight into the role of each lipid component, the phase behavior of SC lipid mixtures, and the large-scale structure of the native SC lipid matrix, many questions remain unanswered. One of the long-term goals of this work is to help elucidate the experimental findings and further validate the hypothetical structural models of SC lipids by using simulations to provide the molecular level insight absent from experimental work.

Chapter III

MOLECULAR MODELING BACKGROUND

Molecular simulations are becoming commonplace in the study of biological systems. Specifically, computer simulations have been used to study a variety of properties for simple and complex membrane systems[64]. These simulations allow analysis of the membrane properties at the molecular level and aid in relating microscopic properties to the macroscopic properties of a system. Simulations can observe structural effects, validate experimental findings, and observe molecular level phenomena to help understand macroscopic observations.

As discussed in earlier chapters, we believe molecular simulations could be used to provide much needed insight into the molecular level behavior of the skin lipids in the stratum corneum (SC). We believe we can observe the effects of altered composition and show how the structure within the skin membrane is altered with the addition of drugs and penetration enhancers. Without knowing the molecular level arrangement of the lipids, simulations of the SC lipids may not be valid. As a result, simulations need to be performed on the self-assembly of the lipids, and the resulting structure can then be validated by comparing the self-assembled structures to experimental observations.

One way of simulating the movement of molecules over time is through molecular dynamics (MD). MD simulations solve Newton's equations of motion by integrating the forces from a force field on each interaction site. A force field contains the mechanical intramolecular and intermolecular interactions of the molecules in an

analytical form. Many different force fields exist; therefore, the choice of force field depends on the system being simulated.

Typically, MD simulations are run on the atomistic level which details and accounts for the interactions between every atom in the system. Atomistic simulations can be very accurate because of the amount of detail incorporated into the system, but the calculations involved in studying complex phenomena like self-assembly of a mixed lipid system can be costly with respect to computer time. A "cheaper" alternative is to run simulations on the coarse-grained (CG) level, where groups of atoms are collectively represented as a single molecule. For specific examples, refer to the CG section later in this chapter. Since the degrees of freedom are less than that of atomistic models, CG models are computationally more efficient, allowing for the study of more complex systems, e.g. the self-assembly of a mixed-lipid system.

3.1 Utility of Molecular Simulations for Study of Bilayers

A large amount of insight can be gained from running atomistic simulations, and their results help direct future experimental work. Of relevance to our research, simulations have been performed on bilayers with double-tailed lipids to examine structure and observe the effect of incorporating other components into the bilayer. In addition, the effect of altering the structure of the lipid components has been observed in the overall bilayer structure. Also, atomistic simulations have been able to observe self-assembly of a lipid system into a bilayer[11].

Before elaborate simulations can be performed, force fields must be optimized and tested in simple bilayer simulations. At this stage, the test force field is utilized in a

bilayer simulation, and its utility is measured by how well it can reproduce structural features seen experimentally. Common observables to extract from bilayer simulations include z-density profiles indicating the location of water, hydrophilic groups, and hydrophobic groups, the average area or volume per lipid, lateral radial distribution functions of the hydrophilic groups, and the lipid tail order parameter. The order parameter is a measure of how frequently trans-gauche conformations are observed in a lipid tail within the bilayer phase.

After a bilayer structure is validated and established through simulations, the effects of altering the structure of bilayers can be compared between different simulations. Notman *et al.* have studied the effects of adding DMSO to CER2 bilayers as an analogous form of the lipid matrix in native human skin[10]. Small amounts of DMSO have an effect on how the lipids structure themselves, but larger amounts completely disrupt the structure, and the lipids undergo a phase transition from the gel phase to a fluid phase. By looking at snapshots from the simulations, they observe the DMSO essentially "steals" hydrogen bonds from the ceramides preventing the lipids from staying together. From these simulations, the authors were able to deduce the mechanism of how DMSO acts as a penetration enhancer, and they established a high degree of hydrogen bonding is needed to enable ceramides to form and retain a bilayer structure.

Peters *et al.* used a combination of simulations and experiments to observe how the addition of deprotonated, protonated, saturated, and unsaturated fatty acids to a phosphocholine bilayer affected packing[65]. In general, they found the degree of saturation had a minimal effect compared to a deprotonated head group. With the negatively charged acids, the head groups had a high tendency to interact at the water

interface, but the protonated acids tended to reside farther into the bilayer. As a result, protonated acids decreased the area per lipid and elongated the bilayer height while salts increased the area per lipid and decreased the bilayer height. This behavior may provide further evidence of how and why the inclusion of fatty acids in SC lipid mixtures induces an extension in the long periodicity phase (LPP) length and tighter packing arrangement seen in the crystal change from the hexagonal to the orthorhombic phase. Free fatty acids are required in the skin due the ability to extend the LPP length and induce a crystal change.

Simulations involving sphingomyelin and phosphocholines mixed with cholesterol have also been performed and these simulations reach similar results and conclusions when compared to each other[66-70]. As is seen experimentally, cholesterol is found to increase the order of the bilayer as shown by the order parameter of the carbon tails and by the decrease in area per lipid upon adding cholesterol to a pure bilayer. In addition, the simulations indicate the rough ring face and hydrogen bonding between cholesterol and the bilayer lipids dictate the orientation and packing of cholesterol in the overall bilayer structure.

Niemela *et al.* studied the effects of chain length and tail saturation on the packing of sphingomyelin bilayers[71]. As to be expected, increasing the tail length of the lipid increases the bilayer thickness, but they also observed an increase in area per lipid with unsaturated tails. In a sense, the kink in the tail caused by the unsaturated bond decreases the order found amongst the tails and decreases the packing density of the lipids as a whole. Pandit *et al*. simulated bilayers with varying degrees of saturation and

the addition of cholesterol and found the smooth face of cholesterol has an affinity for saturated tails and the rough face has an affinity for facing towards the unsaturated tails[68].

Lipid diffusion of mixed lipid bilayers is slow and self-assembly of a mixed lipid system cannot currently occur in the timescales available in atomistic simulations[72]. However, bilayer self-assembly can be observed atomistically for a pure phospholipid system as reported by Marrink *et al.*[11] Lower detail CG simulations can achieve longer time and length scales allowing for simulation of self-assembly of a large mixed lipid system.

3.2 Coarse-Graining (CG) Techniques

Many CG models have been proposed, where they all differ in approach and concept, but share common coarse-graining steps. The first step maps the molecule to a CG level, and the next step establishes the interactions between the molecules. While the details of these two steps differ with respect to the method chosen, each method shares similar guidelines.

When mapping the atomistic molecule to the CG level, the degree of coarse-graining must be determined. Traditionally, no more than six heavy atoms (atoms heavier than hydrogen) are represented by a single CG bead. When mapping to the CG level it is also important to retain functionality, i.e. polarity, hydrophobicity, hydrogen bonding capabilities, and charged groups. Although limiting the amount of heavy atoms to less than six heavy atoms to one bead is desirable, retaining functional groups within a bead is more important. For example, if we consider the phospholipids shown in Figure 3.1, we could represent the head group in at least two different ways. On the left, half of

the phosphate group is in one bead, with the rest mixed in with a portion of the choline group. On the right, the phosphate group is one bead and the choline is another separate bead. Even though the mapping on the right exceeds the recommended number of heavy atoms within a site, functionality due to charge is retained and therefore is more appropriate.

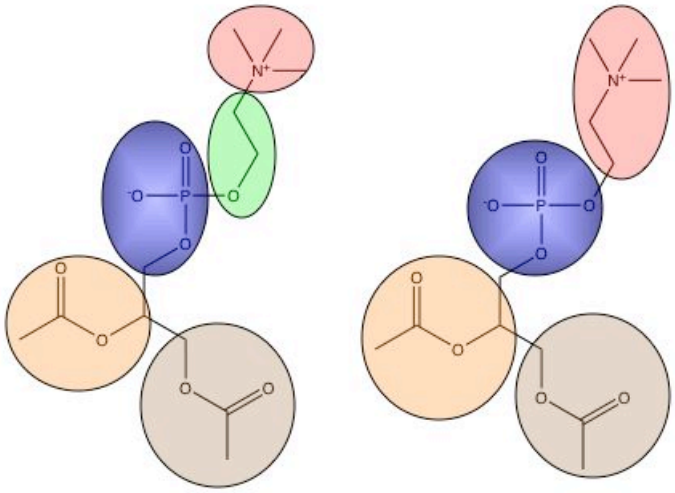


Figure 3.1 – Comparative mapping of the head group of a phospholipids. On the left is a misrepresentation of the head group and on the right is an appropriate mapping for the head group of the phospholipids.

After determining how to map the molecule, the dynamics, physics, and qualitative behaviors of the CG molecule are established by determining the parameters governing the model. The parameters are altered to match the behavior of the CG molecule to its target, typically atomistic simulations. Some target behaviors are mutually exclusive and cannot be matched concurrently. As shown by Izvekov and Voth[73], one cannot match the diffusivity and the radial distribution functions (RDF) in conjunction with each other. Likewise, Toth found energy and structure could not be

matched simultaneously[74]. An important decision when developing a CG model is to determine what capabilities of the atomistic molecule should be retained in the CG molecule. If the diffusivities match, the CG molecule will move like the atomistic molecule. If the RDF's match, the CG molecule will structure itself similarly to the atomistic molecule. Determining what behaviors are retained on the CG level is dependent upon the target properties, but also the choice of CG method dictates what properties can be matched. Some commonly used methods are listed below:

- United Atom (UA)
- Guess and Test
- Force-matching[75]
- Reverse Monte Carlo (RMC) [76]
- Self-consistent method of Reith, Pütz & Müller-Plathe (RPM)[77,78]

3.2.1 United atom (UA)

A united-atom (UA) model is one in which hydrogens are not explicitly represented. For example, ethane would be a 2-site UA molecule, but an 8-atom fully atomistic molecule. Although UA models are very accurate, a degree of detail is sacrificed due to the removal of the hydrogens. Also, even after the removal of the hydrogens, UA models still contain a higher number of degrees of freedom compared to other CG methods. Traditionally, force field parameters are developed from quantum mechanics calculations and these share force field components, which are calculated for fully atomistic models. UA models can be used for any system where the interaction of hydrogens does not

significantly affect the results and, thus, significant time is saved due to the reduction in the degrees of freedom and the ability to use higher integration timesteps.

3.2.2 Guess and test

When coarse-grained approaches to simulations were first developed, systematic parameterization techniques did not exist to develop an accurate CG force field. In early CG work, a form of the force field was chosen and the force field parameters were perturbed till the simulations produced the qualitative or quantitative observables desired[79-83]. Qualitative observations include bilayer formation and phase separation. Quantitative targets included area/lipid measurements in bilayers, melting points, and/or free energy. For example, Smit *et al.* used the guess and test method to develop the earliest work on CG models to observe micelle formation for a water/oil system[79]. By altering the parameters of a LJ potential, they observed the correct self-assembly in the oil and water phases.

Probably the most commonly used force field derived from the guess and test approach is the Martini force field developed by Marrink *et al.*[80,81] In this work, there is a general framework for determining the force field necessary for any CG molecule of interest. A harmonic oscillator defines all the bond and angle potentials while Lennard-Jones potentials define all non-bonded potentials. The interaction energies are fitted to match free energy calculations from reference systems and the hard sphere diameters are kept at 0.47 nm (the diameter of a 4-water cluster). The well depth of all interactions are taken from a table, which contains four general types of interactions: charged (Q), polar (P), non-polar (N), and apolar (A) with multiple subtypes. The bead types are chosen for

a molecule and the table contains the energy for the non-bonded force between the beads. The force field has also been extended to ring-type molecules like benzene and cholesterol.

Although the guess and test method enables a force field to predict some target behavior, determining the reference parameters is slow and inconsistent compared to more systematic methods available. Additionally, some details of the system cannot be duplicated; for example, the authors of the Martini force field were able to reproduce qualitative phase behavior and bilayer formation but believed fine structural details like crystalline repeat units cannot be duplicated[81].

3.2.3 Force-matching

Voth and coworkers initially developed force-matching for finding a force field of an atomistic simulation from *ab initio* calculations[75]. However, the concepts can also be utilized to develop parameters for a CG model from an atomistic model. To develop a CG force field, a set of reference forces are first collected from atomistic simulations, after which CG forces are matched to the reference forces using least-squares optimization. The objective function is shown in equation (3.1),

$$\chi^2 = \frac{1}{3LN}\sum_{l=1}^{L}\sum_{i=1}^{N}\left|F_{il}^{ref} - F_{il}^{p}(g_1,....,g_m)\right|^2, \qquad (3.1)$$

where N and L are the total number of atoms and atomic configurations, respectively, in the system and $g_{i,}$ where $1 \le i \le m$, are the parameters in the force field (e.g. the site diameter of a trimethyl group).

One drawback to this method is the difficulty in implementation. The fine details of force-matching are conceptually difficult to apply and understand, and developing the codes for implementation is labor-intensive.

3.2.4 Reverse Monte Carlo (RMC)

The RMC method was originally proposed as a way to develop atomistic force fields but, in a similar fashion as force-matching, the method can also be applied to CG force fields[76]. After an initial guess for the force field, the CG simulation is run and the target observables are recorded. The results are substituted into equation (3.2),

$$\Delta\langle A\rangle^k = \langle A\rangle^k - A^* = \sum_\gamma \frac{\delta\langle A\rangle^k}{\delta\alpha_\gamma^k}\Delta\alpha_\gamma^k + O\left((\Delta\alpha_\gamma^k)^2\right), \tag{3.2}$$

where A^* is the target observable, α_γ^k represents the input parameters, and the brackets indicate an average ensemble. From the definition of an ensemble average, the value of the gradient of the observable is estimated via equation (3.3),

$$\frac{\delta\langle A\rangle^k}{\delta\alpha_\gamma^k} = -\beta\left(\left\langle A^k \frac{\delta H}{\delta\alpha_\gamma^k}\right\rangle - \langle A\rangle\left\langle\frac{\delta H}{\delta\alpha_\gamma^k}\right\rangle\right). \tag{3.3}$$

where H is the total energy of the system, and β is equal to $1/kT$ (here, T is the temperature of the simulation and k is Boltzmann's constant). With equation (3.3), $\Delta\alpha_\gamma^k$ can be solved from equation (3.2), resulting in a new set of input parameters for the next simulation. In this iterative manner, the optimization scheme incrementally refines the input parameters of the force field. Termination happens when the change in the observables, the change in the parameters, or the gradients of the observables with respect to the parameters is near zero.

The major strength of this method is the built-in sensitivity analysis. Information from equation (3.3) provides an idea of which parameters are most influential on the observables.

3.2.5 Self-consistent method of Reith, Pütz and Müller-Plathe (RPM)

In the RPM method, a CG force field is developed through a multi-centered Gaussian fit for the bonded potential and an iterative Boltzmann inversion for the non-bonded potential. The RPM method has been used to develop a structurally-relevant CG model for polymers and liquid crystals[77,78,84-86]. When the RPM method was first developed, it utilized a simplex algorithm to fit the parameters of an analytical potential of the CG model[87]. However, for large parameter sets, the simplex algorithm was not well suited for the model optimization, necessitating an evolution to the current RPM method. The bonded potential includes bonds and angles between beads, but not four-body interactions like improper angles and dihedrals. The non-bonded potential is an isotropic numerical potential, which collectively represents the Van der Waals (VDW) and electrostatic interactions of a CG bead. Both pieces of the force field are developed from a target trajectory taken from atomistic simulations.

From the target trajectory, the distance between two bonded sites or the angle between three angled sites is added to a histogram and normalized. Although there is no explicit justification, we assume Milano *et al.*[78] weighted distributions for angles by a $\sin(\theta)$ term to provide for a preference for higher angles as shown in equation (3.4),

$$P(\theta) = f_n p(\theta) / \sin(\theta) . \tag{3.4}$$

The normalized distribution, $P(\theta)$, for both angles and bonds are fitted to a multi-centered Gaussian function, as given in equation (3.5), by altering the center (θ_{ci}), total area (A_i), and width (w_i):

$$P(\theta) = \sum_{i=1}^{n} \frac{A_i}{w_i \sqrt{\frac{\pi}{2}}} \exp\left[-2 \frac{(\theta - \theta_{ci})^2}{w_i^2} \right]. \tag{3.5}$$

Equation (3.5) is the same for the bonds, except θ is replaced with *r*. By applying a Boltzmann inversion to equation (3.5), one derives the potential seen in equation (3.6),

$$V(\theta) = -kT \ln P(\theta), \tag{3.6}$$

where *k* is Boltzmann's constant and *T* is the temperature.

Evaluating the bonded potential is an explicit calculation, but the non-bonded potential is derived through an iterative optimization. At each iteration step, the potential is updated via a Boltzmann inversion *viz*,

$$V_{i+1}(r) = V_i(r) + kT \ln \frac{g_i(r)}{g^*(r)}, \tag{3.7}$$

where $V_i(r)$ is the potential at iteration *i*, $g_i(r)$ is the radial distribution function (RDF) taken from the CG simulation, and $g^*(r)$ is the target RDF. Either a LJ potential or a previously optimized potential for a similar interaction is used as the initial guess. The termination occurs either at a set value of *i*, or when the merit function,

$$f = \int_{r_{\min}}^{r_{\max}} w(r) \left(g_i(r) - g^*(r) \right)^2 dr, \tag{3.8}$$

returns a value, which is below a set tolerance. In equation (3.8), a weighting function, $w(r)$, is put in place to diminish deviations at small values of *r* as defined by equation (3.9),

$$w(r) = \exp(-r). \tag{3.9}$$

The pressures calculated from the CG potentials are significantly different than what is measured from atomistic simulations. To address this discrepancy, a pressure correction term is employed,

$$\Delta V(r) = V_0 \left(1 - \frac{r}{r_{cutoff}} \right), \tag{3.10}$$

is added to the reoptimized potential and the corrected potential is reoptimized as defined in the previous paragraph; this iterative process is repeated until the output pressure of the optimized potential matches that of the atomistic simulation.

The goal of this work is to mimic the structure formed by SC lipids, so our CG models must be able to retain the structural capabilities and trends of its atomistic counterpart. RDF's are a measure of how molecules structure themselves, as such we chose to match RDF's on the CG level via the RPM method to ensure the skin lipids form experimentally observed structures. The RPM fitted the RDF's for all of the interactions involved with cholesterol, fatty acids, and water, and self-assembly of these molecules was observed upon mixing. In the next chapter, we discuss the details of the development of our models and the implementation of the RPM method to fit structural behaviors.

Chapter IV

METHODOLOGY

The goal of this research was to develop coarse-grained (CG) models for the components of the stratum corneum (SC) lipid matrix. The structure of the complex lipid matrix cannot be determined through experiments or atomistic simulations, but the structure may be verified by observing the CG lipids self-assemble into a structure similar to the SC lipid matrix. Assuming the CG model is accurate and the CG lipids structure themselves in the same way as they do in experiments, a great deal of insight may be gained from our simulations.

To develop a CG model, the potential must be optimized to fit a desired target. In this work, we have applied the self-consistent method of Reith, Pütz, and Müller-Plathe (RPM) to fit the radial distribution functions (RDF) from atomistic simulations in order to ensure the CG model structures itself the same as its atomistic counterpart. This chapter includes details for the atomistic and CG simulations performed, a discussion of the steps involved for the development of the overall force field, and a summary of our implementation of the RPM method. The simulation details found in this chapter summarize how the simulations were run for each of the systems reported for later chapters.

4.1 Coarse-Grain Model Development Strategy

There is a limited amount of experimental data available for SC lipids and there are no target RDF's in the literature for use in the RPM method optimization. Consequently,

atomistic simulations will serve as the basis for our CG models; to that end, the atomistic force field must initially be validated before further CG force field development. To verify the accuracy of the atomistic force field, we have measured the melting point and the parameters of the repeat unit of the lipid crystal structures to compare against experimental values.

Once the accuracy of the atomistic force field is confirmed, atomistic simulations can be run to provide target data for CG force field optimization. From each atomistic simulation, we extract a target trajectory containing 1,000 frames and map each frame to the CG level. A trajectory is essentially a series of snapshots of the configuration of atoms at regular intervals in the computer simulation. In each snapshot, the center of mass of each atom found within a CG bead, defined by the mapping, determines the target location of that bead on the CG level. This target CG trajectory serves as the source of all of the CG force field optimization targets.

The RDF's measured from the target trajectory are the optimization targets for the non-bonded potential within the RPM method as detailed in the previous chapter. Atomistic simulations were run with pure lipid systems and mixed lipid systems to provide target RDF's for the optimizations of pure bead interactions and of the cross-interactions between different bead types, respectively. The force field is intended to be transferable and to work well with mixtures. As such, interactions will only be optimized from a reference system and be transferred to other optimization systems as needed. For example, the TAIL beads are optimized from a free fatty acid and are transferred for use in CG simulations involving ceramides. However, transferred interactions will not be re-optimized while the other interactions are being optimized. For example, all of the

interactions between fatty acid beads were derived from pure fatty acid simulations, the interaction between water beads were derived from pure water simulations, but pure fatty acid and pure water interactions were held constant when optimizing the cross-interactions between water and acid beads.

For the pure systems, the CG force field was developed in both crystalline and amorphous configurations. A crystalline system is one where the molecules are arranged according to experimentally available crystal structures, and the amorphous configuration is one where the molecules are arranged in a random and disordered state. It is difficult to know *a priori* which state is more appropriate for the crystalline lipid matrix and is a subject of future work. There are no available experimental crystal structures for mixed lipid systems (regardless of which state is more appropriate for self-assembly). Therefore, mixed lipid simulations must be run in the amorphous state to provide target RDF's for optimizing cross-interactions between different lipids.

4.2 Simulation Details

In this work, we have performed molecular dynamics (MD) simulations, in which Newton's second law of motion,

$$\vec{F}_i = m_i \frac{d^2 \vec{r}_i}{dt^2}, \tag{4.1}$$

is solved to observe the dynamics of the system. In equation (4.1), m_i is the mass of atom *i*, $\vec{r}_i$ is the position of atom *i* at time *t*, and $\vec{F}_i$ is the force acting on atom *i*. The forces are integrated to determine the velocity and the position of each atom after a

segment of time (a timestep) has elapsed. The forces are calculated from the potential energy, V, as shown in equation (4.2),

$$\vec{F}_i = -\nabla_{\vec{r}_i} V \,. \tag{4.2}$$

The potential energy is defined by a force field, which dictates the forces acting on every atom or interaction site in the system studied. The force field can have a multitude of forms and typically involves details for both intramolecular and intermolecular forces. In this work, we use the CHARMM force field[88] as the basis for our lipid models because CHARMM accurately represents biological molecules and their interactions. In addition, Cournia *et al.* has developed a force field for atomistic simulations of cholesterol compatible with the existing CHARMM force field[89,90]. The general form of the chosen force field is shown in equation (4.3):

$$V_i = \sum_{bonds} K_b (r_{ij} - r_{eq})^2 + \sum_{angles} K_a (\theta_{ijk} - \theta_{eq})^2 + \sum_{dihedrals} A\left[1 + \cos(n\phi_{ijkl} - \delta)\right]$$

$$+ \sum_{impropers} K_{imp} (\phi_{ijkl} - \phi_{eq})^2 + V_{VDW} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \,. \tag{4.3}$$

Here, K_b, K_a, K_{imp}, and A are the bond, angle, improper, and dihedral force constants, respectively. The variables r, θ, and ϕ represent the value of the distance, the angle, and the dihedral (or improper) angle, respectively, and the subscript, eq, represents the equilibrium value of those parameters.

The non-bonded forces are split between a Van der Waals (VDW) potential, V_{VDW}, and an electrostatic potential. A Lennard-Jones (LJ) potential,

$$V_{VDW} = \sum_{VDW} \varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - 2 \left(\frac{\sigma}{r} \right)^{6} \right], \tag{4.4}$$

represents the VDW interactions, where σ equals the interaction distance at the potential's minimum energy, ε. Coulomb's law defines the electrostatic potential, and the reaction field model[91] was implemented to account for long-ranged electrostatic interactions because of its speed and accuracy. The form of the reaction field potential is show in equation (4.5),

$$V_c = \frac{1}{4\pi\varepsilon_0}\sum_{i<j} q_i q_j \left[\frac{1}{r_{ij}} + \frac{B_0 r_{ij}^2}{2R_c^3}\right], \tag{4.5}$$

where q_i is the charge of the atom i, R_c is the cutoff radius, and ε_0 is the permittivity of a vacuum. B_0 is given by equation (4.6),

$$B_0 = \frac{2(\varepsilon_1 - 1)}{(2\varepsilon_1 - 1)}, \tag{4.6}$$

where ε_1 is the dielectric constant outside the cutoff region. Due to the nature of the SC lipids on the CG level, each site is charge neutral and, therefore, CG simulations do not account for electrostatics, explicitly.

Simulations were performed either using the DL_POLY 2.14[92] or the LAMMPS[93] MD codes and the leapfrog Verlet integrator algorithm was used to solve the equations of motion[94]. The cutoff for the non-bonded interactions was set at 12.0 Å. If a temperature or pressure controller was used, it was implemented using the Nosé-Hoover thermostat and barostat with a time constant equal to 1.0 for both controllers. Finally, for all simulations, periodic boundary conditions (PBC) and the minimum image convention were employed.

4.2.1 Initial configuration and equilibration

For crystal simulations, the molecules were placed inside a unit cell according to the reported experimental crystal structure. The unit cell was replicated along the *a*, *b*, and *c* cell vectors to provide the initial placement of all of the molecules in the system. Over the course of 350 ps in an NVE ensemble (constant volume and constant energy), temperature scaling was slowly increased from 0.01 K to the simulation temperature (typically 298 K) with the scaling frequency slowly increasing from every step to every 1000 equilibration steps. The equilibration scheme for the crystal takes the initial crystal structure and slowly relaxes it without strong perturbations to the initial configuration.

The amorphous systems were constructed by placing the molecules on a grid and subsequently equilibrated at a constant volume. To achieve a random and disordered configuration, the initial amorphous configuration must undergo an equilibration scheme capable of dismissing its initial lattice configuration, so it is brought to a high temperature and then quenched to room temperature over a period of 900 ps.

4.3 Atomistic Simulations

Atomistic simulations serve as the basis for the CG force field optimization, but validating the atomistic force field must be performed first by comparing measurable quantities from the simulations to experimental values. All atomistic simulations were run from their equilibrated configurations for 1.0 ns with a 1.0 fs timestep.

The first step in validating the atomistic force field is to duplicate the unit cell of the crystal structure using the NσT ensemble. The NσT ensemble allows the size and shape of the simulation cell to change in order to maintain a constant stress on each

simulation face (indirectly keeping the pressure constant). I.e. the ensemble allows the crystal structure to change if the initial configuration is not stable. The average cell vectors from the complete simulation represent the most stable repeat structure, and as such, can be compared to the experimental crystal structure parameters.

Once the crystal structure from the atomistic simulations is verified against the experimental values, simulations are run at different temperatures above, below, and equal to the experimental melting point of the molecule to estimate the melting point using simulations. At temperatures below the melting point, an NσT ensemble is utilized, but an NPT ensemble is used for higher temperatures because the system is liquid and cannot be used in an NσT ensemble. From the trajectory of each simulation, the average density and self-diffusion coefficient is determined using the Einstein-Stokes equation[95],

$$MSD = \frac{\langle \Delta r \rangle^2}{6\Delta t}, \tag{4.7}$$

where the brackets indicate an ensemble average, Δr is the change in position of the center of mass of the molecule from one frame of the trajectory to the next, MSD is the mean squared displacement, and Δt is the time interval between frames. The slope of the data, when plotted as a function of temperature, has an abrupt change where the system shifts from a crystalline phase to a liquid phase; this change in slope determines the temperature of the melting point as measured by computer simulations.

Once the atomistic force field is verified, target trajectories are taken from the crystal simulations and the necessary amorphous simulations. The amorphous simulations are run for 1.0 ns with a 1.0 fs timestep in an NVT ensemble. The trajectory

from a simulation using the NPT ensemble is also used to "correct" the pressure of the CG models; this is discussed in the following section.

4.4 Implementation of the RPM Method

The RPM method, in its original form, is insufficient to develop an accurate force field to duplicate crystalline structures, in general, and for CG SC lipids, specifically. Although it is capable of reproducing structures in polymers[78,84,96] and liquid crystalline[86] systems, the structures of SC lipids cannot be reproduced without some alterations to the overall RPM method. The following summarizes how the RPM method was implemented and altered to develop a CG model for SC lipids capable of self-assembling into an experimentally observed bilayer.

The first addition to the method was the inclusion of an improper potential and a dihedral potential when needed. Improper potentials were derived in the same manner as the bond and angle potentials and were included in our force field to enforce chirality in the cholesterol molecule. A dihedral potential was added to ensure the rotation of the beads along a chain within the molecule better matched its atomistic target. The potential had to be derived differently compared to the bond and angle potentials because the potential is typically oscillatory and does not follow a Gaussian distribution. As such, the parameters for the dihedral potential (or in other words the bond rotation) are taken from the atomistic CHARMM force field. To help illustrate this concept, Figure 4.1 shows a CG and atomistic representation of an n-hexadecane molecule (equivalent of a lipid tail) with a rotational disc to represent the dihedral interaction in question. In parameterizing the torsion potential, the parameters dictating the dynamics of bond rotation for the

atomistic dihedral were used for the coarse-grained dihedral potential. In this fashion, the bond rotates its components in the same way on the atomistic and CG levels.

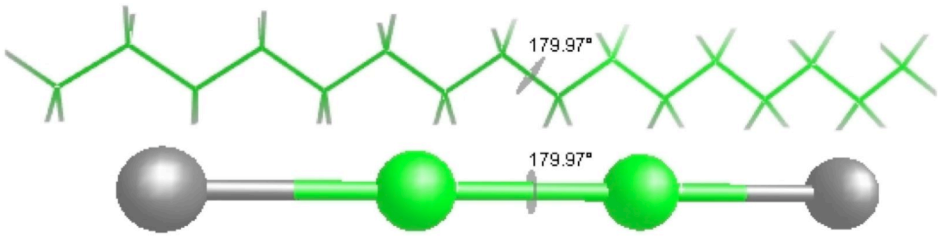


Figure 4.1 - Mapping of the dihedral dynamics to the CG level for n-hexadecane. The top is the atomistic molecule and the bottom is the CG molecule. The disc indicates the dihedral being considered.

A multi-centered Gaussian produces a complex bond and angle potential and would require alteration of the molecular dynamics codes. To simplify the bonded force field a single-peak Gaussian is fitted to the distribution. After the Boltzmann inversion, is applied to the distribution, the potential becomes a simple harmonic oscillator as seen in equations (4.8) through (4.10),

$$P(\theta) = \frac{A}{w\sqrt{\pi/2}} \exp^{-2\frac{(\theta-\theta_{eq})^2}{w^2}}, \tag{4.8}$$

$$V(\theta) = -kT \ln\left(\frac{A}{w\sqrt{\pi/2}} \exp^{-2\frac{(\theta-\theta_{eq})^2}{w^2}} \right), \tag{4.9}$$

$$V(\theta) = \frac{2kT}{w^2}(\theta - \theta_{eq})^2 + const\,. \tag{4.10}$$

For the non-bonded component of the original RPM method, three major changes were implemented. First, a damping factor, δ, was added to equation (3.7) *viz*,

$$V_{i+1}(r) = V_i(r) + \delta kT \ln \frac{g_i(r)}{g^*(r)}. \quad (4.11)$$

For complex mixtures and crystalline systems, we found the optimization of the potential was ill-defined due to the fact that the resultant RDF for each subsequent iteration had a diverging oscillatory response which resulted in large changes in the potential and caused the cell volume to either contract to a size too small for the defined cutoff or expand to a gaseous state. This behavior likely stems from the high order of repetition in a crystalline state, leading to an RDF composed of sharp, tall peaks and low valleys in contrast to a liquid or amorphous state, where a large degree of sampling with respect to orientation leads to broad, short peaks. By adding a damping factor, the changes are smaller and convergence can be achieved. Throughout the optimization, the magnitude of δ correlates with the stability of the optimization and is always positive and less than 1.0. In amorphous optimizations, the damping factor was not needed; therefore δ was set to 1.0 in equation (4.11). A damping factor has been included in the implementation of the RPM method by other researchers, but its importance has never been extensively discussed or explored[84,85,96,97].

We believe the first peak of the RDF represents the main component of the structure and, consequently, fitting this peak is more important than fitting near the beginning of the RDF. In addition, as was seen when comparing the target and the CG RDF, interactions with much higher peaks in the target RDF had large values of the merit function, equation (3.8), and interactions with much smaller peaks had smaller values of the merit function. As such, we found it necessary to normalize the merit function in a relative fashion and to maximize the penalties near the highest peak versus penalizing at low *r* values. To achieve this, the weighting function (equation 3.11) was replaced by

$$w(r) = g_*^{-2}(r_{\max})\exp\left(\frac{-|r - r_{\max}|}{r_{\max}}\right), \tag{4.12}$$

where g_* is the value of the target RDF and $r_{\max}$ is the distance where the highest peak is located. The value within the exponential term provides the largest weight at values near the highest peak and is normalized by a value of $r_{\max}$ to make that quantity dimensionless. The leading term was added to normalize equation (4.12) to allow for comparison of the values of the merit function between different RDF's. From a single optimization step, the new form of the weighting function allows for identification of which RDF's in the system have the largest relative deviations from their targets, thereby indicating which interactions are least accurate.

Lastly, the pressure correction of equation (3.10) did not provide accurate pressure and fitted RDF's in our optimizations. As such, we implemented an alternative strategy for correcting the pressure dependent upon which configuration was utilized. For crystal simulations, an NσT ensemble was used during the CG optimization, to allow for indirect optimization of the pressure along with optimization of the non-bonded potential. For amorphous simulations, equation (3.10) was used to bring the pressure close to the target, but re-optimization was performed in the NPT ensemble with the potential being fitted to RDF's derived from atomistic simulations run with the NPT ensemble. Once the density was within a prescribed error of the target and the NPT RDF's fit their target RDF, the force field was subsequently put back into the NVT ensemble and re-optimized one last time. We found this was only necessary for potentials, which used a Lennard-Jones potential as an initial guess and for very small molecules like water and n-pentanol. When using previously optimized potentials as the

initial guess for a new optimization, the pressure correction was unnecessary. The pressures from these potentials were accurate throughout the optimization and did not require further correction.

In this chapter, we have detailed our general methodology for developing the CG force field. In the following chapters, we will discuss specific exceptions and special cases which required alterations to the above methodology and strategy.

Chapter V

A COARSE-GRAINED MODEL FOR AMORPHOUS AND CRYSTALLINE FATTY ACIDS

Free fatty acids (FFA) are considered to be one of the key structural lipids in the top layer of the skin, the stratum corneum. In the stratum corneum, the hydrocarbon chains of the fatty acids range from 20 to 30 carbons in length[29], with the majority having an even number of saturated carbons and C24:0 (tetracosanoic acid, also known as lignoceric acid) and C26:0 (hexacosanoic acid) are found in the highest mol %. Here the notation C24:0 refers to an acid with 24 carbons and 0 degrees of unsaturation. The pH of the skin prevents deprotonation of the acid group, which cause the acid group to have a softer interaction when compared to the ionic groups of phospholipids.

While it is known that a deficiency of free fatty acids (as opposed to fatty acids bound to corneoytes) in the stratum corneum causes a structural phase change[31], little is known about the molecular level interactions causing the observed behavior. By performing molecular simulations of the lipids of the stratum corneum, we hope to gain insight into how and why a phase change occurs.

The chapter is organized as follows. The overall strategy used to develop a coarse-grained force field for fatty acids is first described. The simulation details are discussed in section 5.2 followed by an overview of how the chosen coarse-graining method was used. Given the semi-crystalline nature of the fatty acids in the stratum corneum, simulations were performed both in amorphous and crystalline states. Finally, the results obtained are presented in section 5.4 and conclusions discussed.

5.1 Coarse-Grained Model and Force Field Development

When developing a coarse-grained model it is essential to identify the target properties that the model should duplicate and the behavior of the target system that can be considered inconsequential to the model's goal. Since the key impact of the fatty acids on the stratum corneum is structural[31], our goal was to develop a coarse-grained force field capable of mimicking these effects. Due to the scarcity of experimental structural measurements for fatty acids, the necessary target data for the coarse-grained force field optimization were radial distribution function data taken from atomistic simulations of the fatty acids. Therefore, in order to ensure accurate results on the coarse-grained level, the atomistic model must accurately reflect experimental behavior.

The CHARMM force field[88] was chosen to describe the fatty acids at the atomistic level due to its success in modeling similar biological systems. Although limited experimental data is available in the literature for fatty acids against which the atomistic force field can be validated, melting point data has been reported for C16:0, C24:0, and C26:0 and the crystal structure for C16:0 has been determined by Moreno *et al*.[98] Additionally, Moreno hypothesizes that their data can be extrapolated to acids of longer chain lengths. Molecular dynamics simulations of C16:0, C24:0, and C26:0 have, therefore, been performed for validation purposes and simulations of C24:0 and C26:0 were subsequently chosen as the basis for the coarse-grained model development, since C24:0 and C26:0 are the predominant fatty acids in the stratum corneum.[29]

In Figure 5.1, we present the mapping of the C24:0 and C26:0 fatty acids to the coarse-grained level. In order to simplify the development of an accurate fatty acid

coarse-grained force field, the mapping was designed to use the minimum number of bead types possible, in order to minimize the number of interactions that need to be optimized, and enable the potentials to be transferable to molecules for which they were not optimized. As shown in Figure 5.1, the acid head group represents one bead (denoted HEAD), while every four subsequent carbons represents a tail bead (denoted TAIL). The two terminal beads (denoted TER2 or TER4) correspond to the last two or last four carbons in the hydrocarbon tail, respectively. This mapping was chosen for its simplicity and transferability. In this work, C26:0 was chosen as the basis for determining the intermolecular interactions for the coarse-grained potential. The optimized shared interactions (HEAD-HEAD, HEAD-TAIL, and TAIL-TAIL) were then used in simulations of the coarse-grained C24:0 model, while optimizing the missing non-bonded interactions (i.e., HEAD-TER2, TAIL-TER2, and TER2-TER2). In this way, an accurate transferable coarse-grained potential for pure fatty acids can be developed.

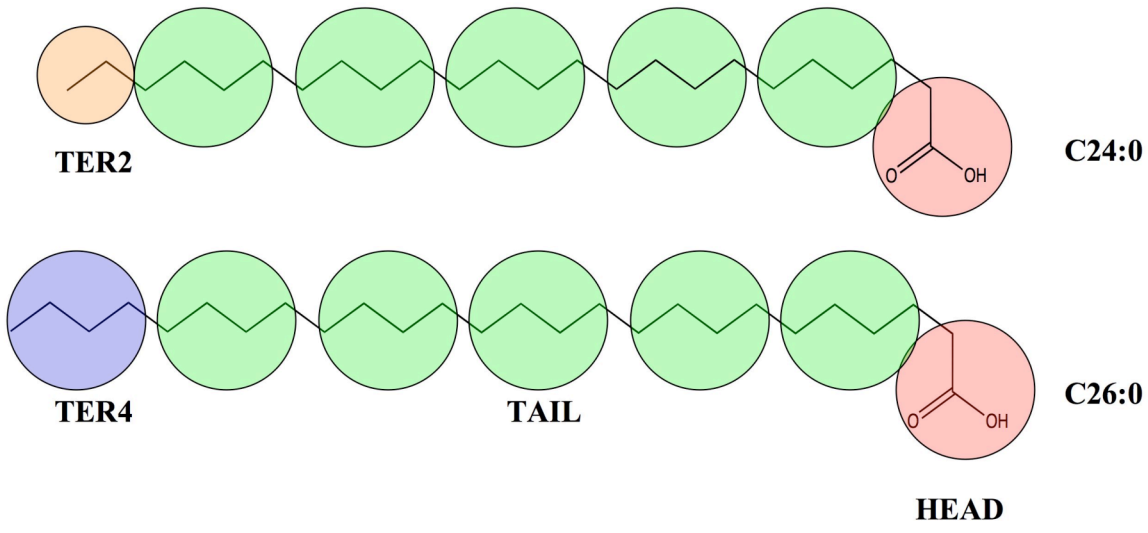


Figure 5.1 – Mapping of the atomistic C24:0 and C26:0 fatty acids to the coarse-grained level.

5.2 Simulation Details

All simulations were performed using the molecular dynamics package DL_POLY (version 2.14) in the isothermal-isochoric (*NVT*) and constant stress ($N\sigma T$) ensembles with a 1 fs time-step and the Nosé-Hoover thermostat and barostat for temperature and pressure control as needed.[92,99]

For the atomistic simulations of crystalline fatty acids, the molecules were initially arranged according to the crystal structure provided by Moreno *et al*.[98] For C16:0, 96 molecules were placed on a crystal lattice and the system equilibrated for 350 ps. For the larger acids studied, 160 molecules were used. The simulations were performed at 298 K and 1.0 bar in the $N\sigma T$ ensemble, which allows the shape and size of the simulation cell to change, if appropriate, from the initial configuration to a more stable crystalline structure. Each simulation was run for 1.0 ns and the system density and crystal parameters determined. To evaluate the melting point of the acids, simulations were performed using the same initial conditions at a series of different temperatures around the experimental melting point, and the density and self-diffusion coefficient determined to be used as signatures for the melting transition.

Although fatty acids found in the skin are predominately in a crystalline state, there is no experimental data for crystalline mixtures. Therefore, simulations in an amorphous state must be used to collect radial distribution function data for parameterization of the cross interactions in the coarse-grained model. For each amorphous simulation, 100 molecules were placed on an orthorhombic lattice and equilibrated for 900 ps, at which point the system contained no apparent defined structure. Production runs in the *NVT* and isothermal-isobaric (*NPT*) ensemble were then

performed for 1.0 ns at 298 K and 1.0 bar and the mapping in Figure 5.1 applied to the resulting atomistic trajectories based on the center of mass of the atoms in each bead to obtain the target RDF data on the coarse-grained level.

5.3 Methodology

While coarse-grained simulations have to date been successful in modeling liquid-like membranes (see for example [73,80,100,101]), polymers (see for example [77,78]), liquid crystalline[86] and glassy states,[102] we are unaware of any previous work on developing a coarse-grained model for a solid crystalline state, and, to our knowledge, the Reith, Pütz, and Müller-Plathe (RPM) method has in the past only been applied to liquids, liquid crystalline or amorphous solid systems. Within the RPM method both the intermolecular and intramolecular components of the coarse-grained force field are based on the target atomistic trajectory mapped to the coarse-grained level; the intermolecular interactions (electrostatic and van der Waals interactions which are consolidated into one numerical potential) are optimized iteratively against the target atomistic data while the intramolecular interactions (bond stretching, bond bending, and torsion motion between coarse-grained sites) are calculated explicitly from the target atomistic data. A summary of the parameterization process can be found in chapter IV.

5.4 Results and Discussion

5.4.1 Atomistic force field validation

In order to ensure the atomistic simulations generate an appropriate target trajectory for the coarse-graining procedure, the atomistic force field was first validated against

experimental data. For validation purposes, we focused on C16:0 as both experimental crystal structure and melting point data is available for hexadecanoic acid.[98] From Table 5.1, in which we compare the experimental crystal unit cell parameters and density with those obtained from the atomistic simulations, we can see the predicted crystal parameters are in good overall agreement with the experimental values. The density and the unit cell angles, are within 2% and 3% of the experimental data respectively, while the *a* vector is within 3% and the *b* and *c* vectors deviate by 10% and 8%, respectively. Fatty acids can exhibit multiple crystal structures at room temperature, which may account for the small discrepancies observed.[98]

Table 5.1 - Crystallographic data for C16:0 from experiment[98] and atomistic simulations.

	Experimental[98]	**Simulation**
Volume (Å^3)	1658	1630 ± 9
***a* (Å)**	35.62	39.07 ± 0.230
***b* (Å)**	4.949	4.568 ± 0.023
***c* (Å)**	9.406	9.160 ± 0.053
α	90.00^o	91.64^o ± 0.577
β	90.00^o	87.51^o ± 0.314
γ	90.45^o	87.44^o ± 0.479
ρ (g/cc)	1.027	1.044 ± 0.006

In Figure 5.2a, we present the self-diffusion coefficient for C16:0 measured as a function of temperature from atomistic simulations around the experimental melting point. As can be seen from the figure, the diffusion coefficients are small and exhibit little change with temperature until temperatures close to the experimental melting point are reached at which point the diffusion coefficient is seen to increase significantly and a distinctive change in the slope of the diffusion coefficient as a function of temperature is observed, indicating the onset of a phase change. We note a similar trend is observed for

the system density as a function of temperature (shown in Figure 5.2b), again indicating that the atomistic system is capturing the fatty acid phase behavior.

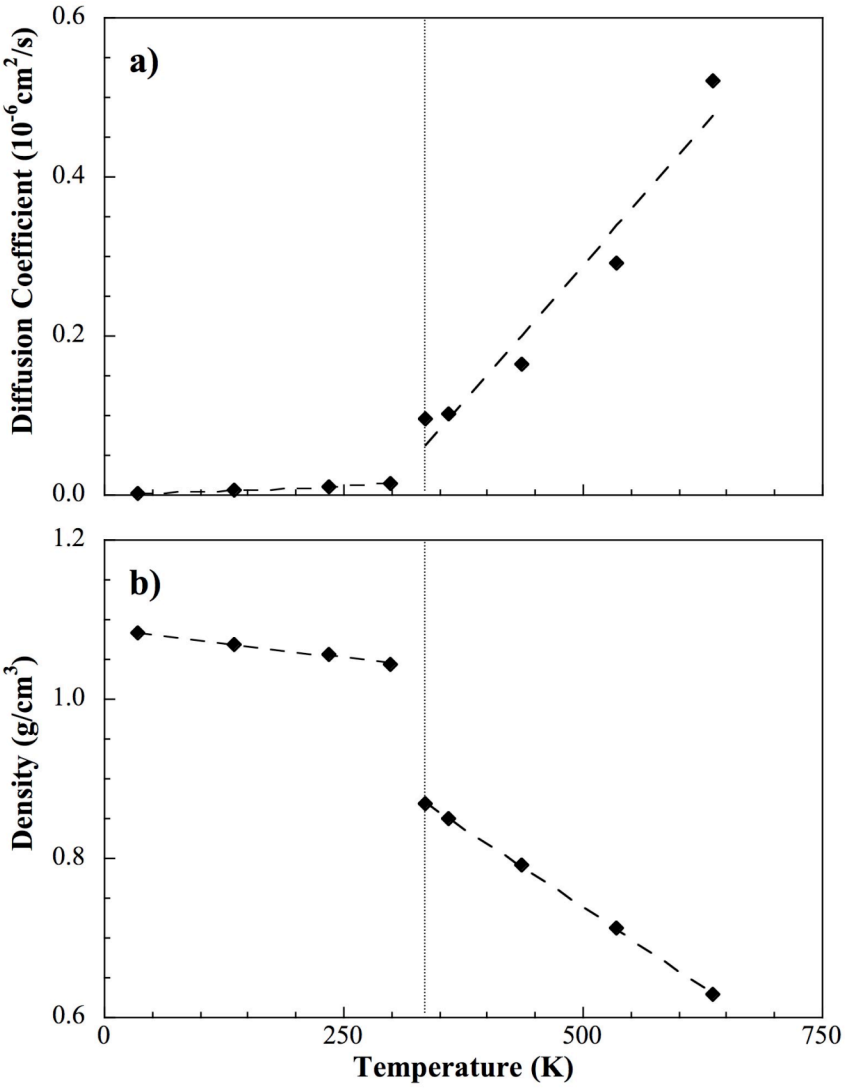


Figure 5.2 – Diffusion coefficients (a) and density (b) for C16:0 as a function of temperature from atomistic molecular dynamics simulations (diamonds). The dashed lines are provided as a guide and the experimental melting point indicated by the dotted line.

We now turn to C24:0 and C26:0 as these fatty acids are found in the stratum corneum. Although no experimental data is available for the crystal structure of these

fatty acids, we can compare the predicted crystal structures with those estimated from the data of Moreno *et al*.[98] The results, presented in Tables 5.2 and 5.3, for C24:0 and C26:0 respectively, exhibit similar trends compared to C16:0 with respect to the deviations in the simulation results compared to the estimated experimental parameters. For both acids, the predicted density and unit cell parameters are within 2% of the values extrapolated from the crystal structure of C16:0, with the exception of the γ angle for which the deviation is 2.9% for C26:0 and 3.5% for C24:0. The predicted melting point transitions are again in good agreement with experimental data as measured by the change in the diffusion coefficient (Figure 5.3) and density. Although these observations are indirect measures of the melting point, they indicate the CHARMM force field can capture the phase change behavior of the fatty acids and therefore the force field can be used to generate the target data on which to base the coarse-grained model.

Table 5.2 - Comparison of crystallographic data for C26:0 from atomistic and coarse-grained simulations and experimental data[98] extrapolated from the C16:0 crystal structure.

	Estimated	**Atomistic**	**Coarse-Grained**
Volume (Å^3)	2634	2618 ± 7	2592 ± 121
a (Å)	56.59	56.65 ± 0.18	60.07 ± 3.26
b (Å)	4.949	4.864 ± 0.017	4.869 ± 0.316
c (Å)	9.406	9.517 ± 0.023	9.062 ± 0.157
α	90.00°	91.37° ± 0.43	88.52° ± 0.29
β	90.00°	91.17° ± 0.57	89.10° ± 0.26
γ	90.45°	87.87° ± 0.38	81.76° ± 0.37
ρ (g/cc)	1.000	1.007 ± 0.003	1.017 ± 0.049

Table 5.3 - Comparison of crystallographic data for C24:0 from atomistic and coarse-grained simulations and experimental data[98] extrapolated from the C16:0 crystal structure.

	Estimated	Atomistic	Coarse-Grained
Volume (Å^3)	2439	2445 ± 8	2345 ± 84
a (Å)	52.40	52.79 ± 0.18	55.93 ± 1.88
b (Å)	4.949	4.874 ± 0.015	4.647 ± 0.282
c (Å)	9.406	9.517 ± 0.036	9.087 ± 0.168
α	90.00°	88.84° ± 0.54	90.41° ± 1.27
β	90.00°	89.69° ± 0.28	89.69° ± 1.45
γ	90.45°	87.26° ± 0.73	88.15° ± 6.86
ρ (g/cc)	1.004	1.002 ± 0.003	1.044 ± 0.033

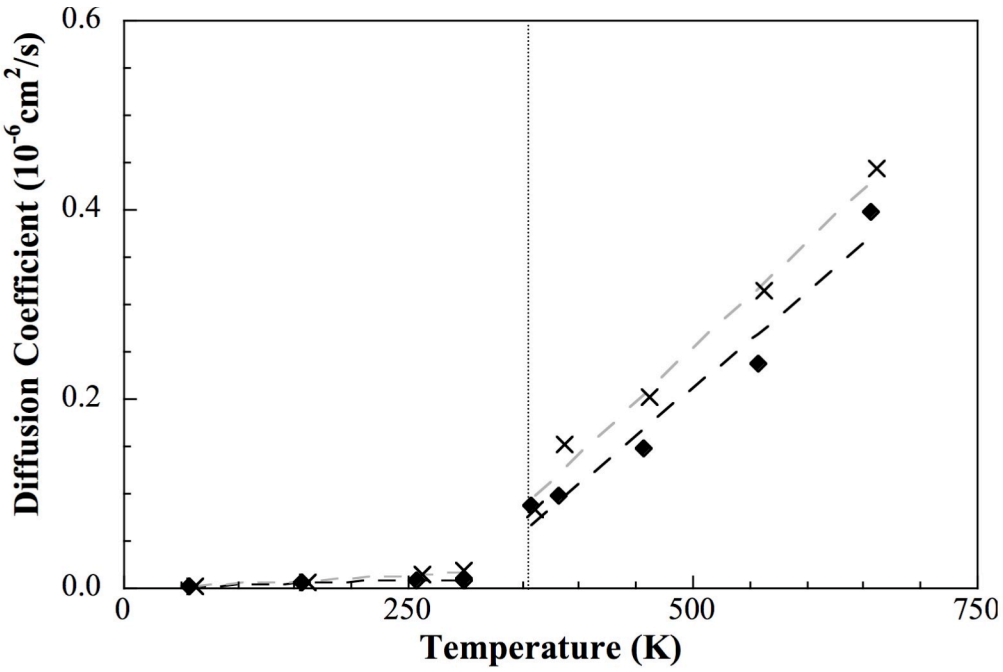


Figure 5.3 – Diffusion coefficients for C24:0 (diamonds) and C26:0 (crosses) as a function of temperature from atomistic molecular dynamics simulations. The dashed lines are provided as a guide and the experimental melting point indicated by the dotted line.

5.4.2 Coarse-grained bonded potential

For parameterization of the coarse-grained bonded potential, bond length and bond angle distributions mapped to the coarse-grained level were extracted from both the atomistic

amorphous and crystal simulations. To match the functional form of the CHARMM force field as described above, the bond length distributions, from both the amorphous and crystal simulations, were then fitted to single peak Gaussian distributions. For illustrative purposes, in Figure 5.4, we show the bond length distribution and the corresponding Gaussian fit for two TAIL beads from a crystal simulation of C26:0.

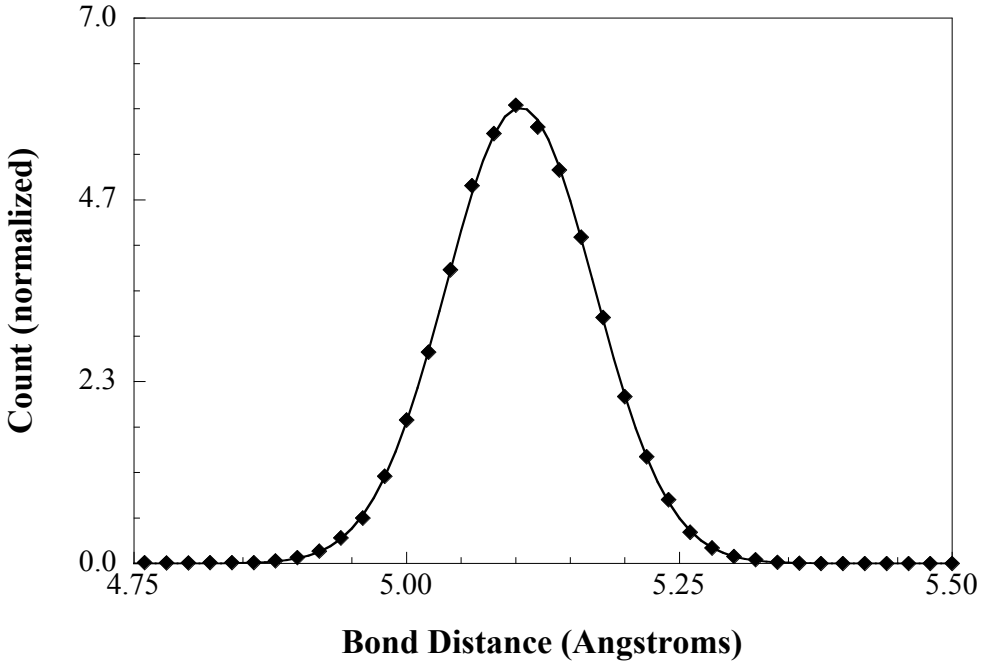


Figure 5.4 – Probability distribution for a TAIL-TAIL bond length from an atomistic trajectory (diamonds) and fitted by a Gaussian curve (solid line).

From the amorphous simulations, we found the bond distributions compared well to those obtained from simulations of a single molecule *in vacuo*; the latter approach was used by Peter and coworkers with the RPM method to parameterize a coarse-grained model for azobenzene liquid crystals.[86] As such, the bond parameters were arbitrarily taken from the amorphous simulations. For the bond angles, in the amorphous state or

from single molecule calculations, no definitive distribution is observed on the coarse-grained level. In both cases, the angles of the target coarse-grained molecule essentially have an equal probability of being measured at any value; hence it is not possible to apply the RPM method to determine the angle parameters. Since the parameters for the amorphous state angle potential could not be derived from the atomistic simulation, the parameters for hydrocarbon tails in the MARTINI force field[80] for coarse-grained dipalmitoylphosphatidylcholine (i.e., DPPC) were used. The DPPC hydrocarbon tails are structurally identical to those found in the fatty acids studied in this work and the DPPC MARTINI force field has been successfully used to study lipid bilayer systems.[80,81]

From the crystal simulations, if we compare the bond-length distributions to those obtained from a single molecule *in vacuo*, the location of the Gaussian peak was found to be at larger distances and the width of the peak much smaller, resulting in higher force constants. As such, the bonds are tighter and longer in the crystal than in the amorphous state (or measured from a single molecule *in vacuo*). For the angles, in contrast to the amorphous state, the angle distributions could be determined from the atomistic crystal simulation, and were found to be very similar to those used in the MARTINI force field, with the force constants being slightly higher in the crystal. In initial simulations of crystalline fatty acids the intramolecular force field parameters from the amorphous state simulations were used, as it was felt that the potentials derived from the crystal could make the coarse-grained molecules too stiff and inhibit the ability of the molecules to move around each other. However, model convergence issues were discovered, but they were alleviated if the longer bond lengths taken from the crystal simulations were used. Therefore, in order to match the target RDF and produce accurate crystal structures from

the coarse-grained model, the longer bonds seen in the crystal system need to be retained but the flexibility of the looser bonds from the amorphous simulations can be used. The bonded parameters used in this work are reported in Table 5.4.

Table 5.4 - Bonded potential parameters for amorphous and crystalline fatty acids.

Bond	Force Constant (kcal/mol/Å^2)	Amorphous Equilibrium Distance (Å)	Crystal Equilibrium Distance (Å)
HEAD-TAIL	12.022	3.8998	4.1229
TAIL-TAIL	6.205	3.9834	5.0902
TAIL-TRM4	7.111	3.9748	4.7917
TAIL-TRM2	6.01	3.1614	3.8790

Angle	Force Constant (kcal/mol)	Amorphous Equilibrium Angle	Crystal Equilibrium Angle
HEAD-TAIL-TAIL	0.018	180.0	180.0
TAIL-TAIL-TAIL	0.018	180.0	180.0
TAIL-TAIL-TRM4	0.018	180.0	180.0
TAIL-TAIL-TRM2	0.018	180.0	180.0

5.4.3 Non-bonded potential optimization

The numerical potentials for all bead types in amorphous C26:0 (i.e., HEAD-HEAD, TAIL-TAIL, TER4-TER4, HEAD-TAIL, HEAD-TER4, TAIL-TER4) were optimized to enable the coarse-grained RDF's to match the target atomistic RDF's mapped to the coarse-grained level indicating structural agreement between the atomistic and coarse-grained simulations. Representative radial distribution functions are shown in Figure 5.5 for the TAIL-TAIL, HEAD-HEAD, HEAD-TAIL, and TER4-TER4 interactions from a C26:0 amorphous simulation. To avoid repetition, the other RDF's are not shown as they all exhibit the same general trends and level of agreement between the coarse-grained and target RDF. The transferability of the coarse-grained potentials was then tested by using the HEAD-HEAD, HEAD-TAIL, and TAIL-TAIL interactions parameterized for C26:0

in coarse-grained simulations of C24:0 while optimizing the missing interactions (i.e., HEAD-TER2, TAIL-TER2, and TER2-TER2). The results of the optimized interactions are not shown as the agreement obtained is comparable to that seen in C26:0; however, the RDF's of the transferred interactions are presented in Figure 5.6 for the TAIL-TAIL, HEAD-HEAD, and HEAD-TAIL interactions to illustrate the retained agreement with their atomistic counterpart and transferability of the potential developed.

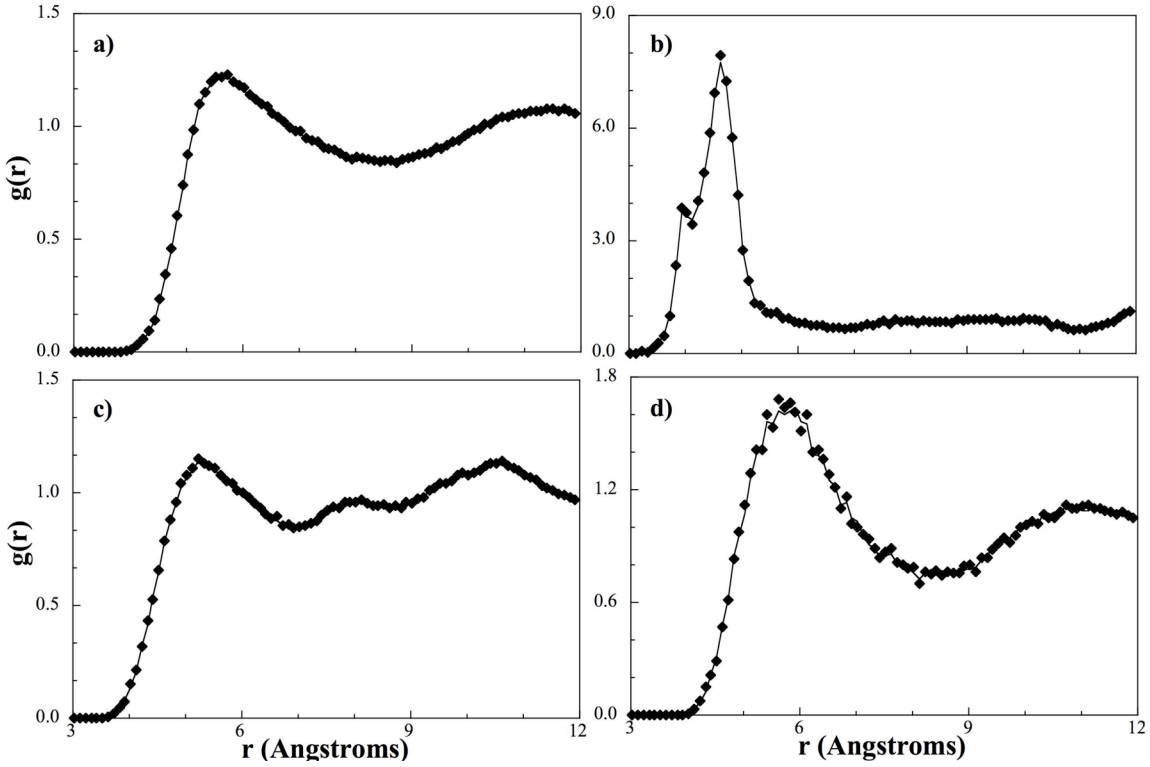


Figure 5.5 – Radial distribution function between a) tail (TAIL-TAIL), b) head (HEAD-HEAD), c) tail and head (TAIL-HEAD), and d) terminal (TER4-TER4) beads in C26:0 from a coarse-grained amorphous simulation (diamonds) and from the target amorphous atomistic simulation (solid line).

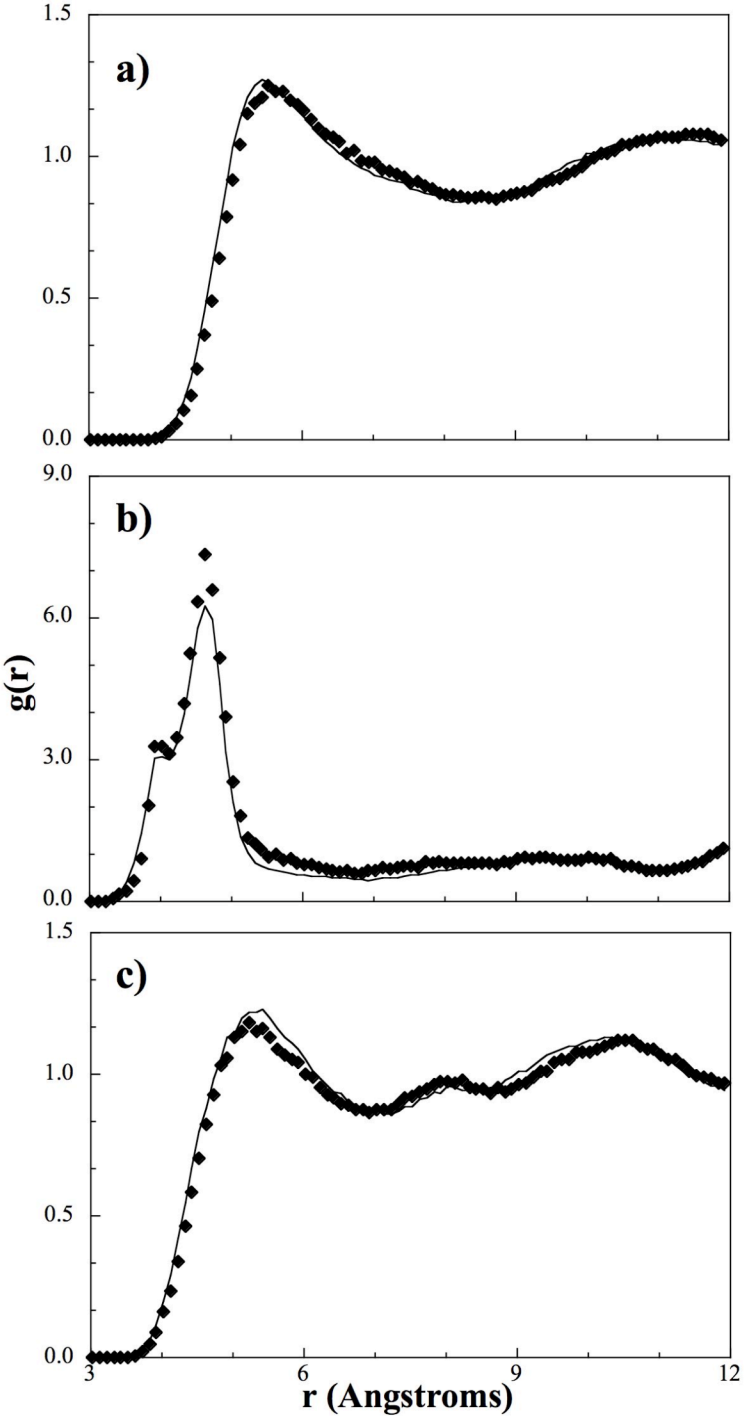


Figure 5.6 – Radial distribution function between a) tail (TAIL-TAIL), b) head (HEAD-HEAD), and c) tail and head (TAIL-HEAD) beads in C24:0 from a coarse-grained amorphous simulation (diamonds) and the target atomistic simulation (solid line). The coarse-grained potential for the interactions are transferred from C26:0.

As an additional test of the transferability of the amorphous fatty acid coarse-grained force field a mixed lipid system has been studied. The missing TER2-TER4 cross bead interaction needed to study mixed lipid systems was first optimized from a simulation of an equimolar mixture of C24:0 and C26:0. A more complex atomistic mixture of the stratum corneum fatty acids in the molar composition reported by Norlén *et al*.[29] (listed in Table 5.5) was then studied. The atomistic simulation was run from a random amorphous configuration and RDF data collected to provide target for validation of the coarse-grained force field. Since terminal beads for odd numbered acids are not considered in our coarse-grained force field, due to the relatively low mol % of odd-numbered acids in the stratum corneum, the odd-numbered acids were described as even-numbered ones in the coarse-grained system, i.e., half of the C25:0 molecules contained a TER2 terminal bead and the other half contained a TER4 terminal bead. A comparison of the target RDF's and the coarse-grained model prediction is shown in Figure 5.7 for the HEAD-HEAD, TAIL-TAIL, HEAD-TAIL, and TER4-TER4 interactions. Similar to what was seen in Figure 5.6 when using transferred potentials from C26:0 to study C24:0, very good agreement is obtained further supporting the accuracy and transferability of the coarse-grained models developed.

Table 5.5 - Composition of free fatty acids seen in the stratum corneum[29] and used in this work in the simulation of a mixed lipid system.

Acid	C20:0	C22:0	C24:0	C25:0	C26:0	C27:0	C28:0	C29:0	C30:0
Mean mol%	5	11	39	10	23	3	8	1	2
Median mol%	4	11	39	10	23	3	8	0	2

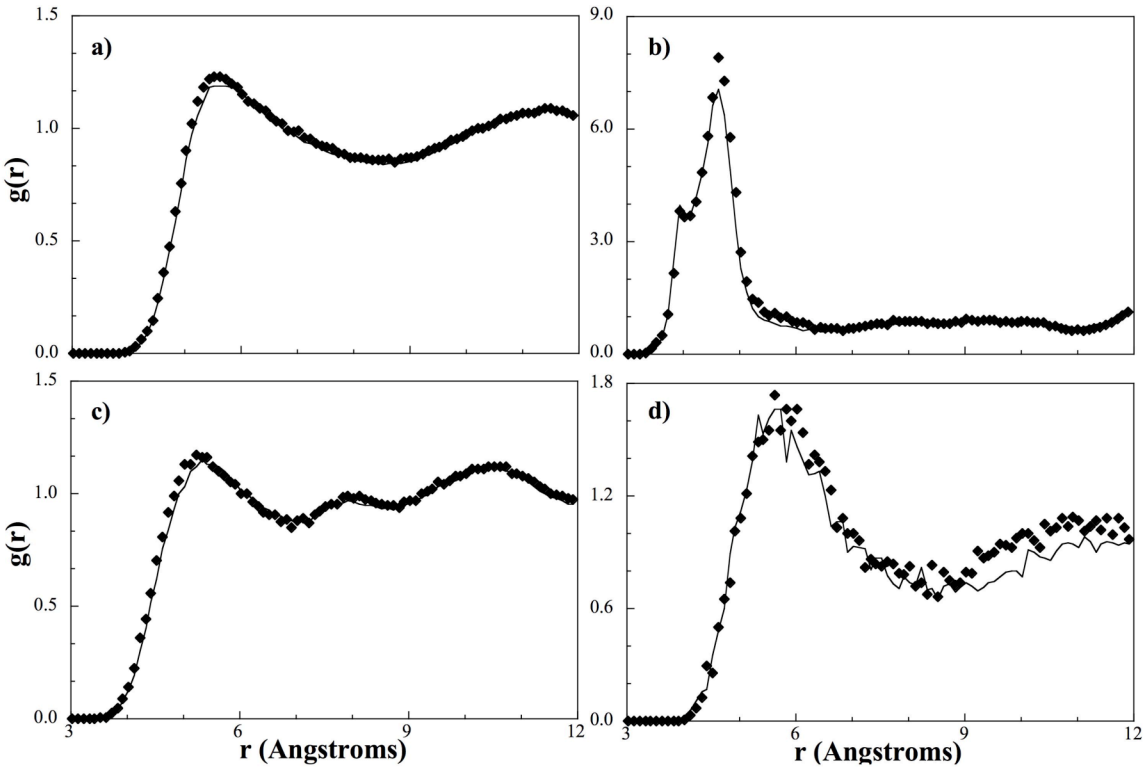


Figure 5.7 – Radial distribution function between a) tail (TAIL-TAIL), b) head (HEAD-HEAD), c) tail and head (TAIL-HEAD), and d) terminal (TER4-TER4) beads in C26:0 from a coarse-grained melt simulation (diamonds) and from the target melt atomistic simulation (solid line) for a mixed fatty acid system.

As discussed in chapter IV, the higher order of structure seen in the crystal compared to the amorphous simulations causes the optimization to be ill defined and requires the introduction of a dampening factor into equation (4.11) to enable convergence to be achieved. We also found that in general the optimization of the crystalline potentials requires more iterations than in the amorphous state. To illustrate the need for the dampening factor in equation (4.11) RDF's from iterations of the TAIL-TAIL potential for C26:0 is shown in Figure 5.8a. From the figure we can see the RDF has multiple sharp peaks, as expected for a solid state, and that without the damping factor the coarse-grained RDF is found to alternate between states A and B in the

optimization, with the magnitudes of the peaks becoming larger instead of converging towards the target. By making the changes in the potential smaller between iterations through the use of the damping factor, minimization of the difference between the coarse-grained RDF and the target RDF can occur and the resultant potential allows for the coarse-grained RDF to match its target (state C). Although the optimized structure (state C) produces higher peaks and lower valleys than the target, the position of the peaks and valleys is retained. Similar agreement is obtained for the other pure coarse-grained interactions (i.e., HEAD-HEAD, HEAD-TAIL, and TER4-TER4), as shown in Figure 5.8b, 5.8c, and 5.8d, again indicating that the coarse-grained crystal has the same structure as seen in the atomistic simulations. This can be further confirmed by comparing the crystal parameters from the coarse-grained simulation to those obtained from the atomistic simulation and the extrapolated experimental values in Table 5.1. The coarse-grained crystal parameters, reported in Table 5.2, are found to be in good agreement with their atomistic counterparts, with only two parameters (the γ angle and the length of the *a* vector) having a percent difference greater than 5%. The small discrepancy between the RDF of the coarse-grained system and the target may account for the observed deviations.

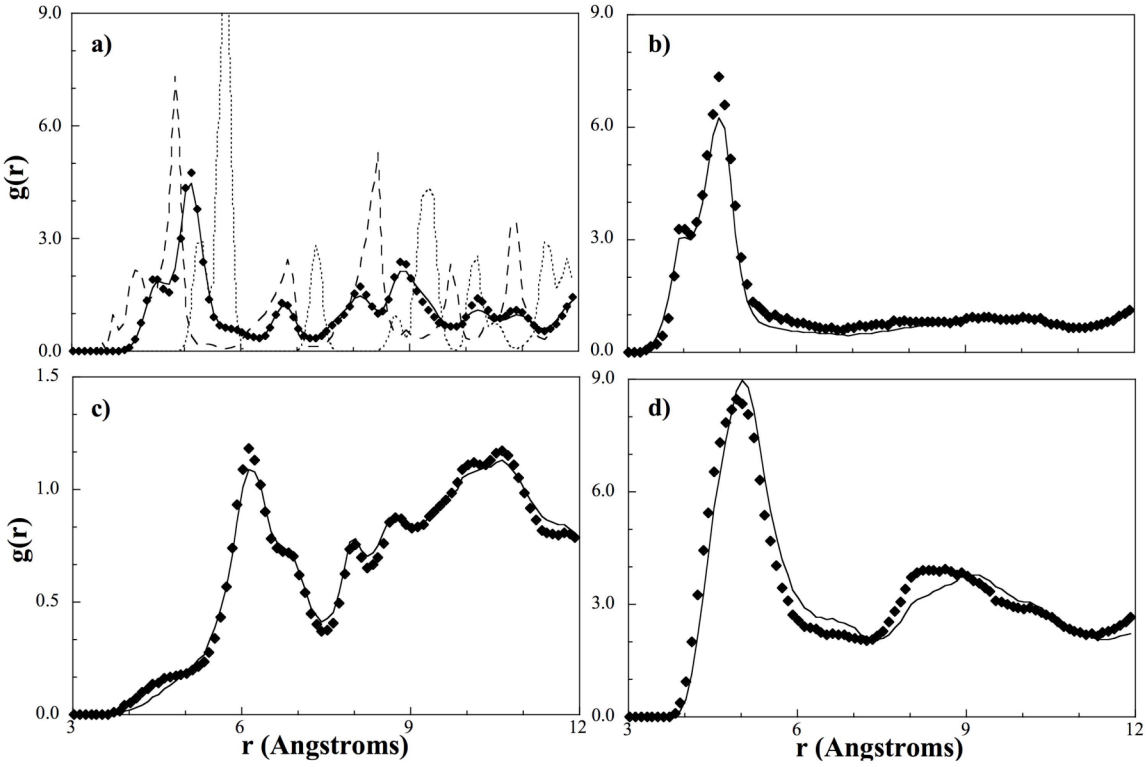


Figure 5.8 – Radial distribution function between a) tail (TAIL-TAIL), b) head (HEAD-HEAD), c) tail and head (TAIL-HEAD), and d) terminal (TER4-TER4) beads in C26:0 from a coarse-grained crystal simulation (diamonds) and the target atomistic simulation (solid line). In (a), State A (dotted line) and B (dashed line) are from sequential iterations in the CG model optimization without the use of a damping factor and state C (diamonds) is from the final iteration in the CG model optimization utilizing a damping factor.

As in the amorphous simulations, the HEAD-HEAD, TAIL-TAIL, and HEAD-TAIL crystal potentials of C26:0 were then transferred and used in a crystalline C24:0 simulation in which the missing interactions (HEAD-TER2, TAIL-TER2, and TER2-TER2) were then optimized. The optimized potentials were found to, again, be in good agreement with the target data and the transferred potentials performed well for C24:0 as shown in Figure 5.9 for the TAIL-TAIL, HEAD-HEAD, and HEAD-TAIL RDF's. While the peaks in the RDF shown in Figure 5.8 are higher in the coarse-grained simulation compared to the atomistic target, they are in the correct positions. This would indicate that the "size" of the coarse-grained beads is appropriate, but the interactions could be too strong, promoting a higher degree of structure than is seen in the atomistic simulation. This could also be due to the use of the force constants from the amorphous state (which were weaker than those derived from the crystal simulations), for which the non-bonded interactions need to be stronger to promote the retention of the crystal symmetry and structure. As seen in Table 5.3, the predicted crystal structure for C24:0 is however still in good agreement with the pseudo experimental data and the atomistic system; the only two values with a greater than 5% difference than the extrapolated values are the *b* and *c* vectors.

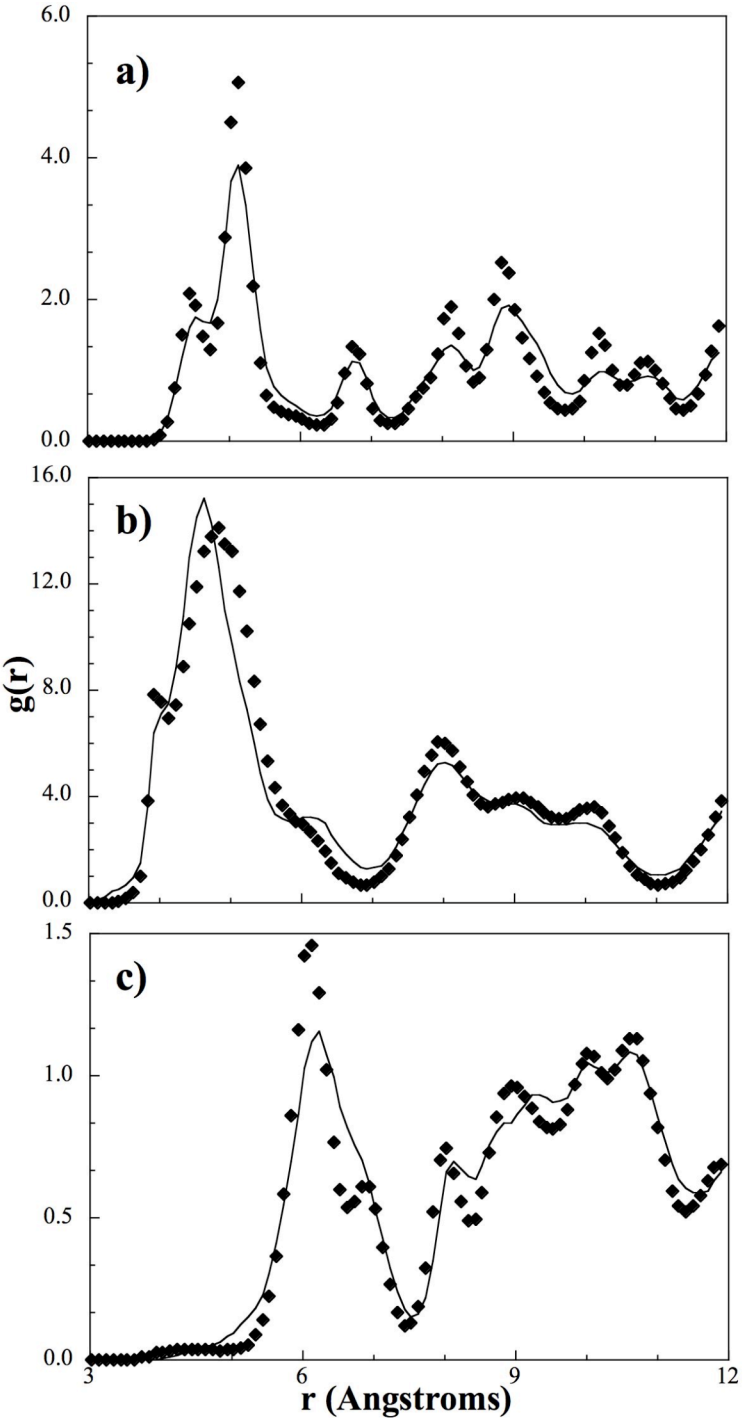


Figure 5.9 – Radial distribution function between a) tail (TAIL-TAIL), b) head (HEAD-HEAD), and c) tail and head (TAIL-HEAD) beads in C24:0 from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line). The coarse-grained potential for the interactions are transferred from C26:0.

Finally a snapshot from the simulation of C24:0 at both the atomistic and coarse-grained levels is shown in Figure 5.10. From the figure, and confirmed by the agreement in the RDF's, we can see the structures are in good agreement. Specifically, the hydrogen-bonding network is contained within the same cross-sectional plane as in the atomistic crystal and the angle between the plane and the tails of 37.28° ± 2.76 in the atomistic simulations is reproduced on the coarse-grained level (36.97° ±3.51). The retention of the hydrogen-bonding network can be verified by comparing the average distance between hydrogen-bonded HEAD beads in both simulations as depicted in Figure 5.11. The average distance from the target trajectory is 5.563 Å with a standard deviation of 0.785, while the average distance from the coarse-grained simulation is 5.113 Å with a standard deviation of 0.429. Finally, the retention of the spacing of the molecules in the coarse grained model compared to the atomistic simulation can be visually verified from Figure 5.10.

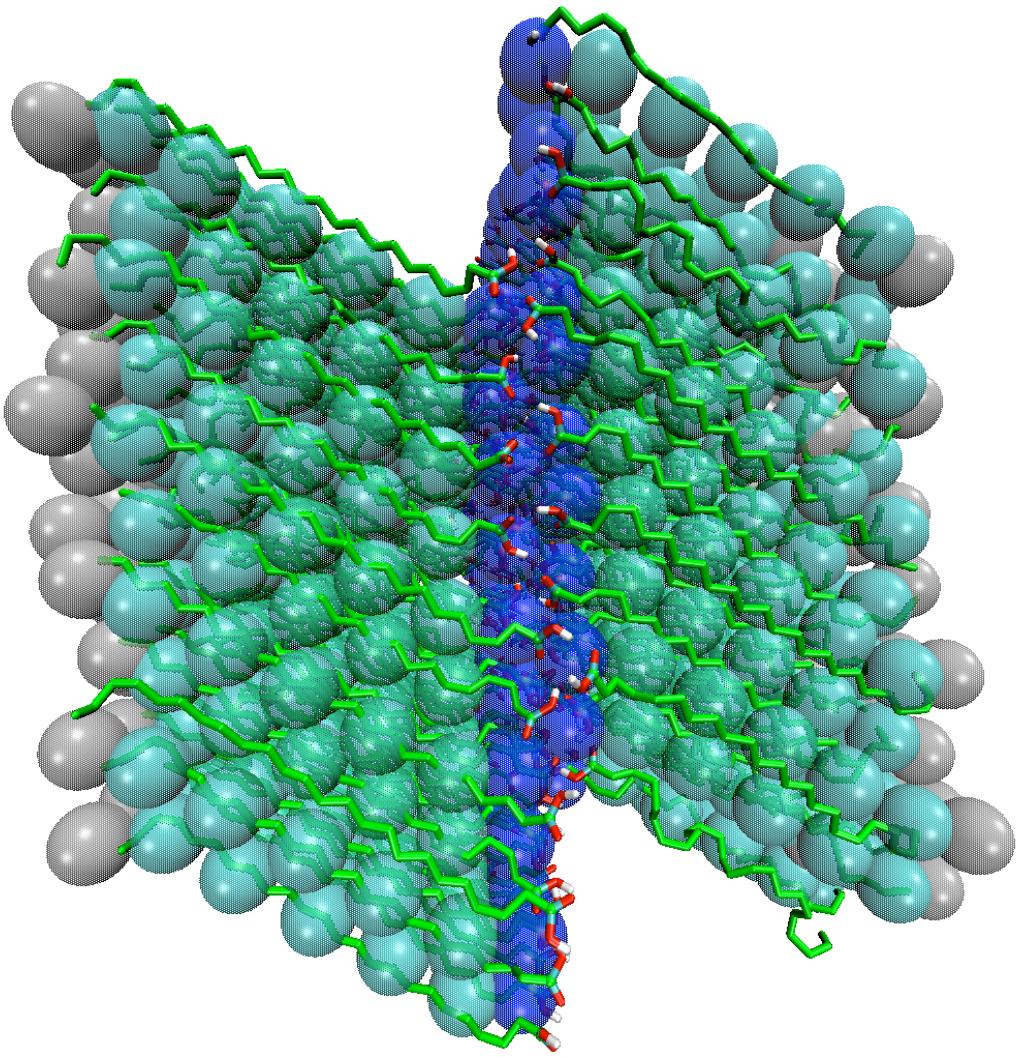


Figure 5.10 – Snapshot form simulations of the C24:0 crystal in the coarse-grained (transparent spheres) and atomistic (solid lines) simulations.

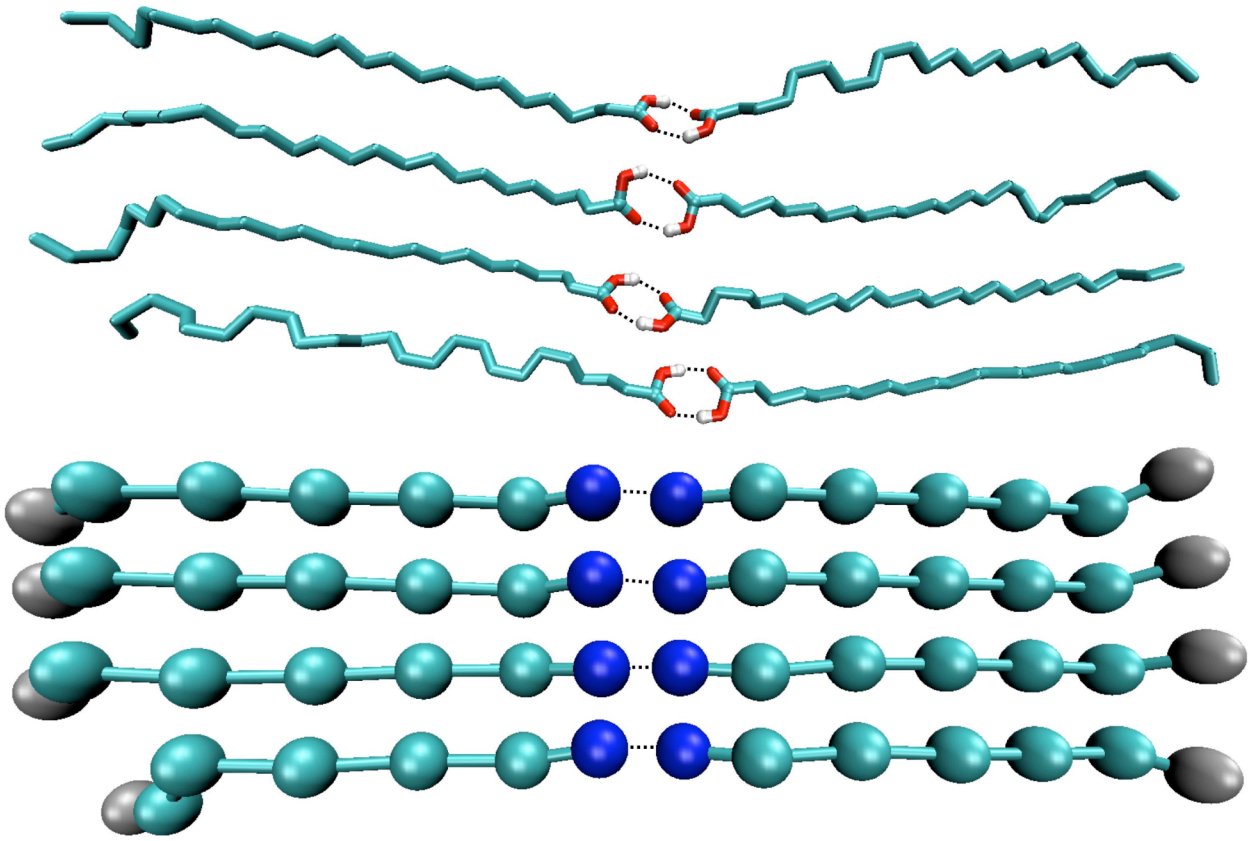


Figure 5.11 – Close-up of hydrogen bonding (indicated by dashed lines) between atomistic acids (top) and between CG acids (bottom).

5.5 Conclusions

An accurate and robust coarse-grained force field has been developed for fatty acids in both the crystalline and amorphous states. When developing a coarse-grained force field, one must decide which observable properties are to be retained on the coarse-grained level, whilst eliminating unnecessary atomistic detail. Since we are ultimately interested in studying the structure of the fatty acids in the stratum corneum, matching atomistic radial distribution functions using the coarse-grained methodology developed by Reith, Pütz, and Müller-Plathe was chosen as the target property. The ability of the CHARMM

all atom force field to accurately describe the fatty acids, and therefore provide reliable target data, was also verified by reproducing experimental melting point and crystal structure data.

By incorporating a damping factor into the optimization scheme, the RPM method was able to optimize coarse-grained potentials to match RDF data from atomistic crystal simulations. The resulting crystal parameters were found to be in good agreement with experimental data indicating the coarse-grained model exhibited the same structural behavior as its atomistic and experimental counterparts. This result is significant given that the development of a coarse-grained model capable of reproducing a crystalline state has, to our knowledge, not yet been reported in the literature. Good agreement between the target RDF's and coarse-grained models was also obtained for the amorphous state. Using the derived potentials to study different molecules and mixtures in a transferable fashion, we tested the robustness of the coarse-grained force fields developed for both the amorphous and crystalline states. In all cases, the correct structural behavior was predicted demonstrating the accuracy and transferability of the coarse-grained models developed.

Chapter VI

A STRUCTURALLY RELEVANT COARSE-GRAINED MODEL OF CHOLESTEROL

Cholesterol plays an important role in the properties of biological membranes and membrane dynamics. Specifically, it is known that cholesterol is required for cellular viability and proliferation[33,34] and that similar sterols, such as ergosterol, cannot replace cholesterol in mammalian cells[35].

As a membrane lipid, cholesterol has a unique molecular structure with both a rigid multi-ring arrangement and a flexible tail. Additionally, within the ring structure, the presence of chiral methyls on one plane of the molecule results in contrasting smooth and "rough" faces, while the alcohol group contributes to the hydrophilic component of the molecule's amphiphilic properties and drives the orientation of cholesterol in bilayers[36]. Cholesterol's unique structure contributes to its properties and function in membranes and bilayers[34]. For example, it has been shown by Yeagle that cholesterol acts as a molecular rigidifer to increase order and decrease molecular motion in lipid bilayers[37] whilst also acting as an impurity to prevent bilayers from becoming too structured[37] through the mismatch in shape between cholesterol and typical membrane lipids (such as sphingomyelin or phospholipids). Cholesterol is also believed to act as a rigidifier and fluidizer in the membranes found in the top layer of the skin, the stratum corneum.[38] The stratum corneum consists of dead skin cells embedded in a lipid-rich environment and plays a central role in controlling the passage of substances across the skin by acting as a barrier to foreign agents entering the blood stream and preventing

water loss from the body. The main lipids of the stratum corneum are ceramides, free fatty acids, and cholesterol,[15,27] which is an unusual lipid combination in that it does not contain any phospholipids and is believed to have evolved from the need for a relatively impermeable protective barrier that is sufficiently flexible to permit movement.[24,31,38,43,47]

Experimentally, cholesterol exhibits a triclinic crystal structure that undergoes a transition at 304.8 K[103,104], to a less symmetric structure, in contrast to what is typically seen in nature, making cholesterol somewhat of an oddity amongst biological molecules[104]. At room temperature (298 K), the unit cell contains 8 unique molecules, but at body temperature (310 K), the unit cell contains 16 unique molecules decreasing its level of symmetry relative to the room temperature cell. In each unit cell, cholesterol displays a degree of pseudosymmetry, in that the molecules exist in similarly structured pairs throughout the cell and the alcohol groups directed towards the center of the unit cell, allowing the molecules to hydrogen bond both within and across the molecular pairs.

In an effort to provide insight into the role of cholesterol in membranes several simulation studies involving sphingomyelin and phosphocholine bilayers containing cholesterol have been reported in the literature. They investigate the ordering and fluidizing effect of cholesterol, the importance of hydrogen bonding, and the trends involved with the different faces of the cholesterol ring structure. For example, Chiu *et al.* performed a series of simulations on phosphocholine bilayers with a wide range of cholesterol concentrations and have shown that the initial addition of 4 mol% cholesterol induces a high degree of structure in the lipid tails, but increasing the concentration had little additional effect.[105] Also, the area per lipid was found to increase in a linear fashion with respect to the amount of cholesterol, indicating the cholesterol was fluidizing the

bilayer. These trends were reproduced in subsequent simulations involving sphingomyelin and other phospholipids.[66,68-70,106]

Using computer simulations, Rog *et al.* extensively studied the hydrogen bonding networks found within phospholipids bilayers containing cholesterol.[69] Although others have observed extensive hydrogen bonding networks in bilayers[66,106], Rog *et al.* found the hydrogen bonding in sphingomyelin and cholesterol was much more extensive than what is found in phospholipid/cholesterol bilayers containing a phosphocholine group.[69]

Zhang *et al.* investigated the effects of unsaturation in the lipid tails with respect to cholesterol's structure effects.[70] Essentially, they found the smooth face of cholesterol has a preference for fully saturated tails, and the rough face of cholesterol packs more efficiently with mono-unsaturated tails. These results are further supported by the work of Pandit *et al*., who found cholesterol packs most efficiently in bilayers with a mix of saturated and unsaturated tails in the main lipid component.[68] In a sense, the saturated tail lies against the smooth face and the unsaturated tail lies against the rough face.

Although atomistic simulations can provide a wealth of information at the molecular-level they are limited to studying short length and time scales making self-assembly in mixed lipid systems, which can occur on the nano- to micro-second timescale, difficult to probe via atomistic simulation.[73,80,82,100,101,107] As a result, coarse-grained models, in which some of the atomistic detail is sacrificed, are commonly used in studies of biological membranes.

In the CG DPPC/cholesterol bilayer simulated by Murtola *et al.*, each molecule (including cholesterol) was represented by a single interaction site[100]. They found qualitative features like cholesterol-rich regions could be duplicated, but a lot of

quantitative features like bilayer height could not be calculated. At a similar level of coarse-graining Khelashvili and coworkers modeled cholesterol as a hard cylindrical rod with which they were able to reproduce the ordering effect of cholesterol in a DPPC bilayer, but they mention the lack of transferability of the cholesterol model to temperatures below the chain-melting phase transition and bilayer systems containing different phospholipids[108]. In more related work, Izvekov and Voth have developed a 4-site and a 7-site model for cholesterol using the force-matching algorithm[109]. In the 4-site model, the molecule is fairly linear with two beads representing the multi-ring structure and the flexible tail mapped to the remaining two beads. In the 7-site model, the flexible tail in is mapped in the same way, but five separate beads describe the ring structure in order to duplicate the planar nature of cholesterol. Both models were found to be capable of duplicating the radial distribution functions measured from atomistic simulations indicating the packing of the cholesterol within the bilayer was retained, though Izvekov and Voth preferred the four-site model due to its computational efficiency. Unfortunately, according to the authors, the coarse-grained models developed can only be used for pre-assembled bilayers and are therefore inappropriate for studying self-assembly. A similar degree of coarse-graining is used in the Martini force field[81] model for cholesterol in which the ring structure and 4 carbons of the flexible tail are described using 6 beads in a cross-linked multi-ring structure and the terminal end of the flexible tail is a separate bead. As such, the single-site tail does not have the same flexibility as that found in the Izvekov and Voth model. Simulations of DPPC/cholesterol bilayers have shown that this model does increase the order in phosopholipid bilayers[81], but we hypothesize the effect may be exaggerated because the rigid ring structure is larger on the

CG level than the atomistic level. In addition, the authors note that the cholesterol model is only appropriate for fluid-like systems.[81]

While, most lipid bilayers containing cholesterol are in a liquid-ordered phase, some (such as those found within the stratum corneum) are in a more solid-like crystalline state. Therefore, in this chapter, we have developed a coarse-grained model for cholesterol that is capable of reproducing the atomistic cholesterol crystalline structure by capturing the dualistic nature of the cholesterol molecule in the coarse-grained mapping scheme. The remainder of the chapter is organized as follows. The details of the atomistic simulations and the development of the coarse-grained model are discussed in section 6.1. The atomistic and coarse-grained simulations performed on crystalline systems in order to develop the cholesterol model are then presented in section 6.2 and the results and conclusions discussed in sections 6.3 and 6.4, respectively.

6.1 Coarse-Grained Model and Force Field Development

A coarse-grained model for cholesterol has been developed with the goal of mimicking the structural properties of cholesterol observed experimentally and in atomistic simulations. The target data for the coarse-grained force field optimization were taken from atomistic simulations of cholesterol using the CHARMM cholesterol force field developed by Cournia *et al.*[90] mapped to the coarse-grained level; using the center of mass of atoms within a site, the location of the target CG beads can be determined.

The non-bonded potential (electrostatics and van der Waals interactions) were determined using the method developed by Reith, Pütz, and Müller-Plathe (RPM),[77,78] which utilizes radial distribution functions (RDF) as the target for optimization of the

intermolecular interactions. The bonded potential was derived explicitly using Gaussian-based potentials as found in the original RPM method[78] resulting in harmonic potentials to described bond stretching and bond angle bending as seen in the CHARMM force field.

Since RDF data from atomistic simulations is used to parameterize the coarse-grained model, the atomistic model was first validated against experimental data in order to ensure accurate results on the coarse-grained level. The CHARMM atomistic force field for cholesterol[89,90] was developed by fitting to the vibrational spectra[89] and dynamics of the alcohol group, in order to capture the hydrogen-bonding capabilities of cholesterol,[90] and the force field shown to predict the experimental room temperature crystal structure of cholesterol. Therefore, as an additional test of the force field, in this work, we studied the crystal structure at body temperature (310 K), which is relevant to most biological membranes.

When mapping cholesterol to the coarse-grained level, the number of beads employed was minimized whilst retaining the key structural features of cholesterol as described below. The first mapping scheme considered is shown in Figure 6.1 from which we see that the rigid multi-ring structure is retained by using five beads (denoted ALC, 1RNG, 2RNG, 3RNG, and 4RNG), with the ALC bead also capturing the amphiphilic nature of the molecule. The chiral methyls on the planar ring are mapped as their own explicit beads (denoted CHM) in order to give the coarse-grained molecule a distinction between the rough and smooth faces, as seen atomistically. As discussed in detail in the results section, the rough face was found to be critical to duplicating the crystalline structure of cholesterol and is the major distinction between our coarse-

grained cholesterol model and those proposed in the literature to date.[81,100,108,109] Finally, the cholesterol tail is described by two beads (TAILC and TERC) of four carbons each, with the beads separate from the rigid CG ring structure to promote the tail's flexible nature.

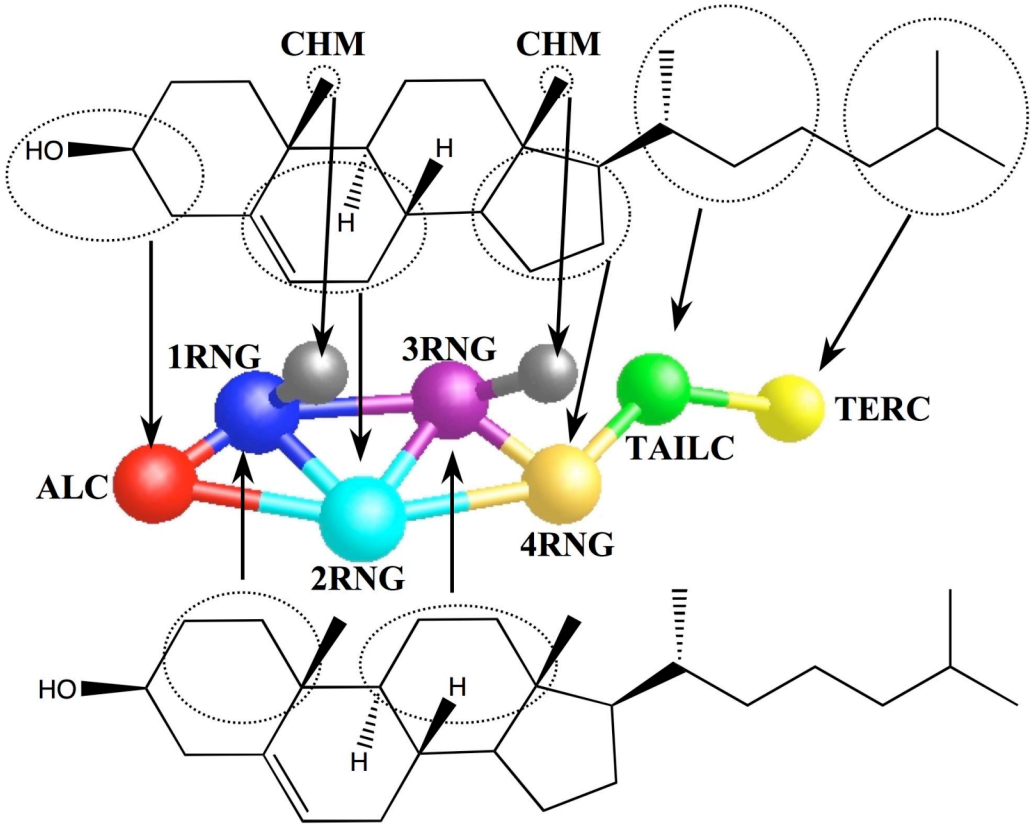


Figure 6.1 – The mapping scheme for cholesterol to the coarse-grained level.

A possible problem with this mapping is the large number (8) of different bead types, which results in a total of 36 interactions that need to be optimized in a pure cholesterol system. While the ALC, CHM, and tails (TAILC and TERC) beads have unique interactions and roles within the model, and must therefore be explicitly represented in the mapping, the four different ring beads could be replaced by four beads

that share the same non-bonded potential. Therefore, in our other model, only one type of ring bead is used, resulting in only 5 different bead types and 15 interactions to optimize. We refer to this model as the homogenized cholesterol model and the benefits and drawbacks of both models are discussed in the results section.

6.2 Simulation Details

The molecular dynamics code, DL_POLY (version 2.14), was used to model each system on the atomistic and coarse-grained levels.[92] The constant stress ($N\sigma T$) ensemble was used in the crystal simulations in order to allow the shape of the simulation cell to change in order to produce the most stable repeat structure of the crystalline unit cell. In the atomistic simulations, 96 cholesterol molecules were initially arranged according to the crystal structure of Shieh *et al.*[110] and the system equilibrated for 350 ps followed by 1 ns of simulation during which the average unit cell parameters were determined. The Nosé-Hoover barostat and thermostat was used to maintain the temperature and pressure as needed and a 1.0 fs timestep used. Simulations were also performed over a range of temperature around the experimental melting point to estimate the melting point of the crystal structure and compare with experimental data. The self-diffusion coefficient and average system density were measured during each simulation and used to determine the approximate temperature at which a phase-change is observed.

In subsequent simulations with the coarse-grained model, the same simulation conditions and initial configurations were used as in the counterpart atomistic simulations. The coarse-grained crystal potential was derived from simulations

performed in the $N\sigma T$ ensemble, and as a result, nullifies the necessity of the pressure correction found in the original RPM method.

The coarse-grained force field was derived from the target atomistic trajectory mapped to the CG level. The RPM method was used to optimize the CG potential in the manner discussed in detail in chapter IV.

6.3 Results and Discussion

6.3.1 Atomistic force field validation

As discussed above, although the CHARMM atomistic force field for cholesterol has been used to predict the experimental crystal structure at room temperature, we further tested the force field by studying the crystal structure at body temperature and the melting point.

From simulation of the cholesterol crystal at 310 K, the average crystal parameters, including lattice vectors and cell density, were determined and are compared to the experimental results[104] in Table 6.1. From the table we can note that the predicted density and unit cell angles are in very good agreement with experiment, with the unit cell lengths showing deviations of less than 3% compared to the experimental values. We also note that our results are similar to the work of Cournia *et al.* who obtained experimental deviations on the order of 3% for the unit cell parameters in their study of the room temperature crystal.

Table 6.1 - Crystallographic data for cholesterol measured from experiments,[104] atomistic simulations, heterogeneous CG model simulations, and homogenous CG model simulations at 310 K.

	Experimental[104]	Atomistic	Heterogeneous Coarse-Grained	Homogenous Coarse-Grained
Density (g/cm^3)	1.0119	1.0117 ± 0.0053	0.9986 ± 0.0233	0.9962 ± 0.0338
***a* (Å)**	27.565	28.204 ± 0.140	28.158 ± 0.361	28.864 ± 1.413
***b* (Å)**	38.624	37.859 ± 0.430	38.306 ± 0.478	36.772 ± 1.412
***c* (Å)**	10.748	10.770 ± 0.064	10.868 ± 0.083	10.806 ± 0.250
α	93.49°	91.10° ± 0.53	94.12° ± 1.55	96.73° ± 3.54
β	90.90°	90.32° ± 0.55	90.38° ± 0.94	88.86° ± 3.30
γ	117.15°	117.56° ± 1.33	118.03° ± 2.97	114.17° ± 2.72

To estimate the melting point of the high temperature crystal simulations were ran over a range of temperature from 25 K – 725 K and the average self-diffusion coefficient and density calculated to provide an estimation of the melting point. The results for the self-diffusion coefficient as a function of temperature are reported in Figure 6.2, from which the melting point is determined to be the point at which a distinct change in the slope of the self-diffusion coefficient as a function of temperature is observed. The estimate of the melting point determined from the intersection between these two lines occurs at 519 K, which is almost 100 K higher than the experimental value of 422 K; however, the determination of the melting point from a molecular simulation using periodic boundary conditions will tend to an overestimation of the melting point because of periodic effects.[111] The gradual change in magnitude of the self-diffusion coefficients at temperatures below the observed melting transition is presumably due to the flexible and more mobile nature of the cholesterol tail compared to the multi-ring structure. As a result the distinction between the solid phase and the liquid phase is less sharp than seen for chain molecules. For example, in simulations of atomistic fatty acids, in which in the crystalline state the hydrocarbon tails are elongated, adopting mainly trans-gauche conformations, and are relatively motionless (i.e. have low diffusivities),[112] a sharp

transition is observed at the melting point. In contrast, while the cholesterol tails behave similarly to the fatty acid tails, the rigidity and general bulkiness of the cholesterol ring structure inhibits motion of the molecule as a whole.

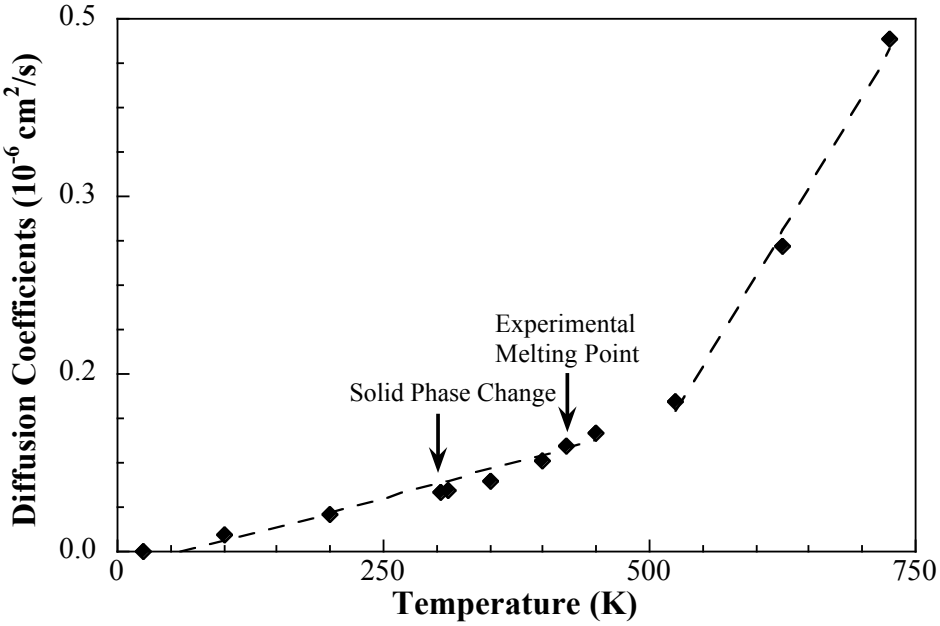


Figure 6.2 – Diffusion coefficients of CHOL as a function of temperature. The slopes of the solid and liquid phase (dashed line) with the greatest degree of fitting to the average values (diamonds). In addition, the experimental melting point matches the simulation's melting point.

A possible cause for the trend seen in the melting point calculation may be the alterations made to the force field by Cournia *et al*. to maximize the hydrogen-bonding capabilities[89,90]. With a lower degree of hydrogen bonding, the solid may lose its structure at higher temperatures, but the room temperature and body temperature crystal structures could not be retained without the hydrogen bonding alterations.

6.3.2 CG Bonded Potential

Following the testing of the atomistic force field the atomistic simulation trajectories were then mapped to the coarse-grained level using the mapping scheme shown in Figure 6.1. To find the parameters of the bond potentials, the distribution of bond lengths were measured and fitted to a single peak Gaussian as seen for example in Figure 6.3 for a bond between a CHM bead and a 1RNG bead. Utilizing the RPM method, the parameters for the harmonic bond potential can be extracted from the distribution as described in chapter IV. Similarly, the parameters for the angle potential can be extracted from a distribution of angles between sites as seen in Figure 6.4 which shows the distribution of the 1RNG-ALC-2RNG angle as a representative example.

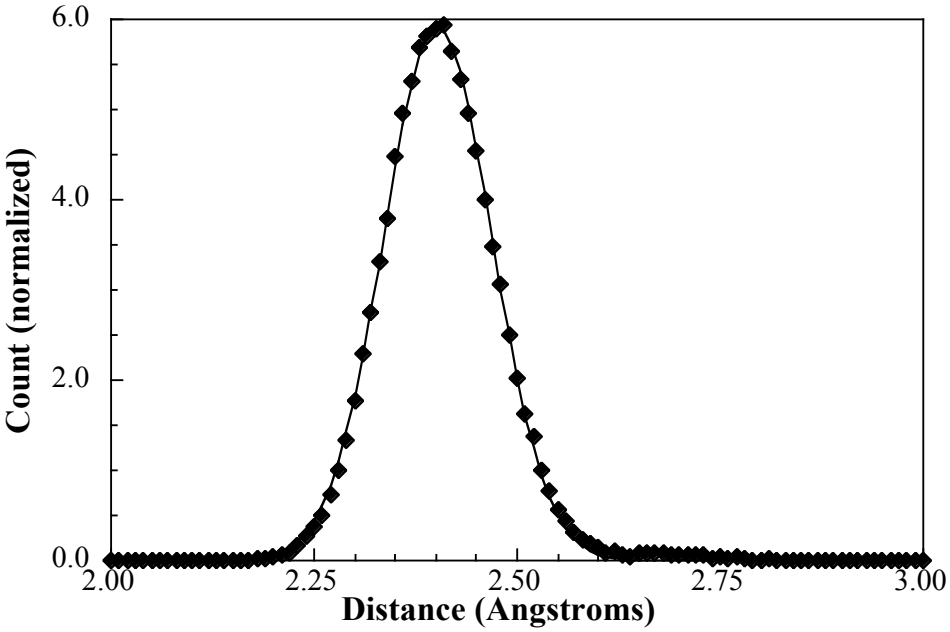


Figure 6.3 – The normalized distribution of the distances between bonded CHM and 1RNG beads (diamonds) and the fitted Gaussian (solid line).

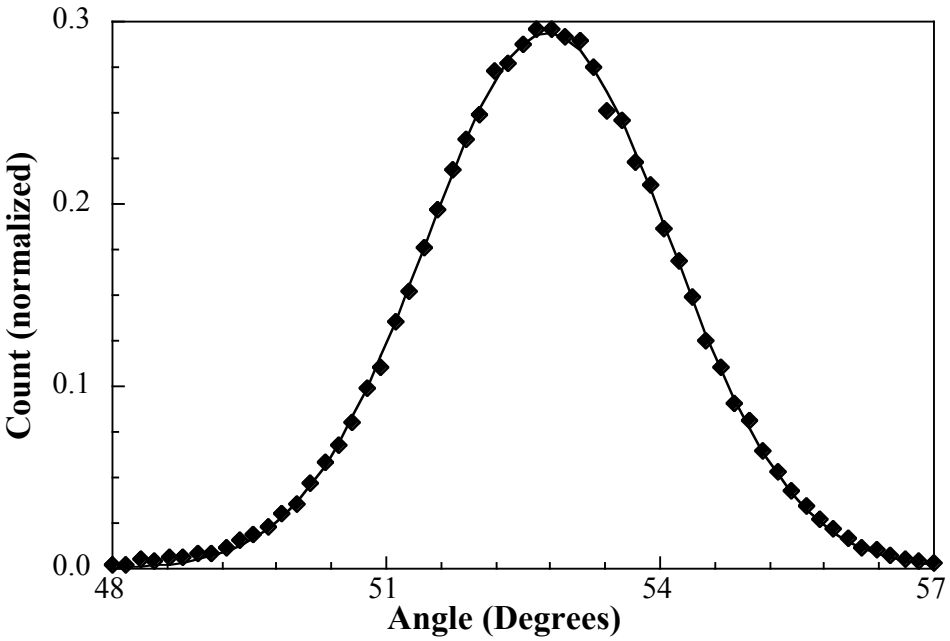


Figure 6.4 – The normalized distribution of the angle between the 1RNG-ALC-2RNG beads (diamonds) and the fitted Gaussian (solid line).

The rigid nature of the ring structure is reflected in the force constants for the bond and angle potentials as recorded in Table 6.2. Compared to the potentials found for the tail bond between the TERC and TAILC beads, the ring force constants are almost an order of magnitude higher.

Table 6.2 - Force constants for coarse-grained bonds of a fatty acid compared to that of cholesterol.

Bond	Force Constant (kcal/mol/Å^2)	Equilibrium Distance (Å)	Angle	Force Constant (kcal/mol)	Equilibrium Angle
ALC-1RNG	355.1	2.591	1RNG-ALC-2RNG	1070	52.8
ALC-2RNG	141.4	4.200	ALC-1RNG-2RNG	631.0	89.1
1RNG-2RNG	204	3.347	ALC-1RNG-3RNG	356.7	142.6
1RNG-3RNG	124.8	4.008	ALC-1RNG-CHM	127.1	107.1
1RNG-CHM	132.9	2.402	2RNG-1RNG-3RNG	1835	54.0
2RNG-3RNG	319.4	3.394	2RNG-1RNG-CHM	218.1	62.7
2RNG-4RNG	155.6	4.180	3RNG-1RNG-CHM	367.3	62.8
3RNG-4RNG	366.8	3.041	ALC-2RNG-1RNG	2379	38.1
3RNG-CHM	145.3	2.622	ALC-2RNG-3RNG	457.0	110.8
4RNG-TAILC	155.8	3.717	ALC-2RNG-4RNG	284.4	155.9
TAILC-TRMC	10.8	4.227	1RNG-2RNG-3RNG	864.4	73.0
			1RNG-2RNG-4RNG	628.8	118.6
			3RNG-2RNG-4RNG	2647	45.9
			1RNG-3RNG-2RNG	1324	52.9
			1RNG-3RNG-4RNG	485.5	133.4
			1RNG-3RNG-CHM	113.9	129.3
			2RNG-3RNG-4RNG	1040	80.8
			2RNG-3RNG-CHM	280.7	98.1
			4RNG-3RNG-CHM	778.5	53.9
			2RNG-4RNG-3RNG	1913	53.3
			2RNG-4RNG-TAILC	363.8	149.1
			3RNG-4RNG-TAILC	495.7	96.1
			4RNG-TAILC-TRMC	68.0	135.8

6.3.3 Non-bonded potential optimization

Both cholesterol models share the same target RDF's for interactions involving ALC, CHM, and flexible tail beads, and the RPM method can optimize the interactions to the same degree of accuracy as seen in Figure 6.5, where the target RDF and coarse-grained RDF are displayed for the ALC-ALC (6.5a) and CHM-CHM (6.5b) bead interactions. In both Figure 6.5a and 6.5b, the high degree of structure can be verified by the existence of multiple regularly spaced sharp peaks. This level of structure is seen in the other RDF's, and the RPM method also matches the CG RDF to its target to the same degree as these interactions.

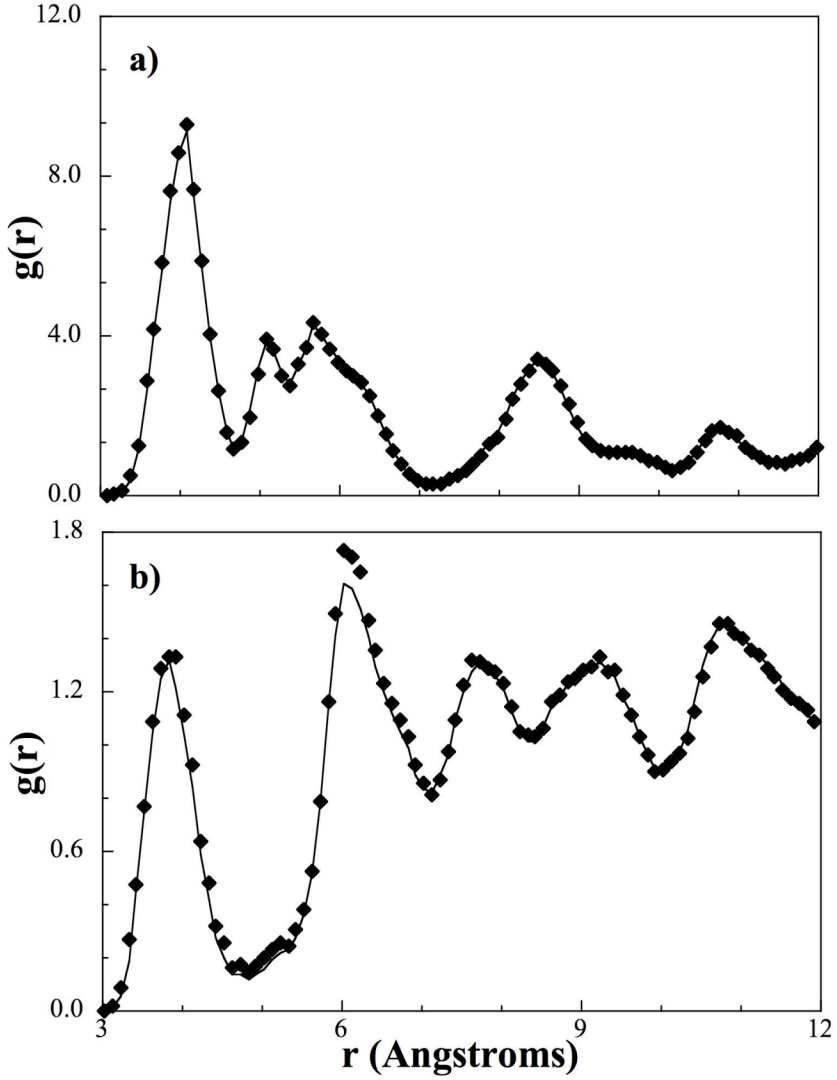


Figure 6.5 – Radial distribution function between a) alcohol (ALC-ALC) and b) chiral methyl (CHM-CHM) beads from a coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

Figure 6.6 shows the RING-RING RDF of the homogenous ring model along with the pure 1RNG, 2RNG, 3RNG, and 4RNG RDF's, where the effects of homogenizing the ring beads can be seen. With explicit interactions between the different ring beads, the RDF is more structured because there is only one bead of the same type at one specific interaction distance on the nearest neighbored cholesterol.

With the homogenous model, the nearest neighbored cholesterol contains 4 beads of the same type all at different interaction distances, so the structuring becomes more generalized (less localized) and results in a broader less structured RDF. The RING-RING RDF is not as structured as the other crystal RDF's, but there is more structure than an RDF from an amorphous configuration. However, the modified RPM method is still capable of fitting the RDF to a high degree of accuracy as seen in Figure 6.7.

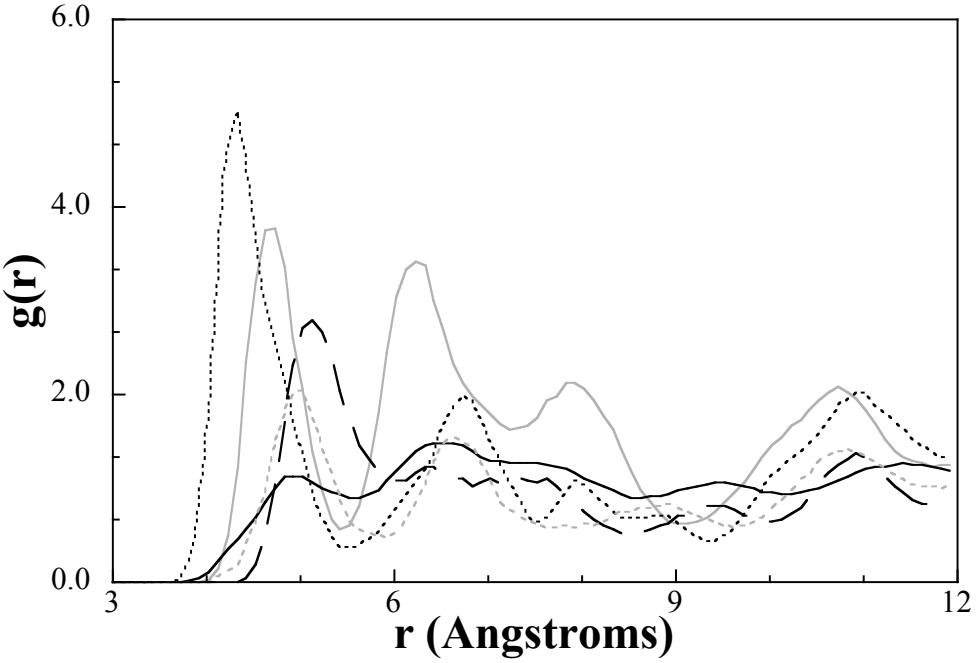


Figure 6.6 – Target radial distribution function for 1RNG-1RNG (solid grey), 2RNG-2RNG (fine dashed black), 3RNG-3RNG (long dashed black), 4RNG-4RNG (fine dashed grey), and RING-RING (solid black).

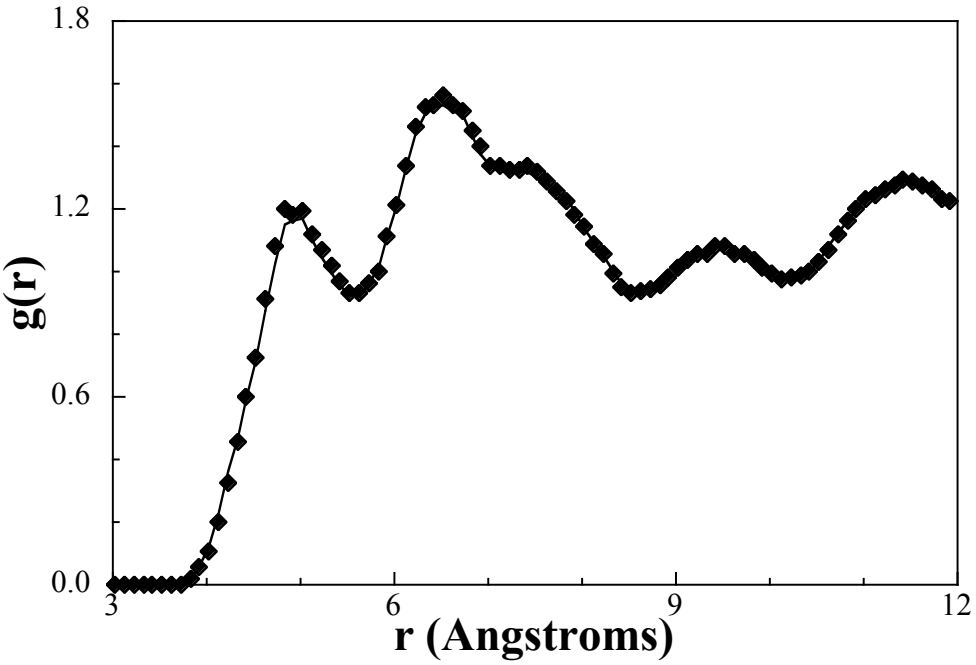


Figure 6.7 – Radial distribution function between ring beads (RING-RING) from a crystalline coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

Throughout the cholesterol crystal optimization, the magnitude of the damping factor (~0.01) was generally lower than what was used in optimizing other crystal potentials (~0.10). Our hypothesis is the large number of different interaction added further complications to the optimization, forcing the optimization steps to be smaller. In Peter *et al.*, they found they could not achieve convergence with the original RPM method for interactions between azobenzene sites using simulations of azobenzene[86]. As a result, they optimized pieces of the model in simpler simulations such as pure benzene. We believe we encountered a similar problem with cholesterol because of the presence of 36 different potential optimizations. In general, the damping factor had a larger value in the homogenized ring model, so convergence was achieved in a fewer number of iterations. With larger changes in a more stable optimization, the progression along the

optimization pathway towards the minimum happens at a faster rate. As such, the homogenous model allows for less optimization in two separate ways: there are less total interactions to optimize, and the optimization for each individual interaction is more stable allow for a faster optimization rate. We believe the complications stem from the high degree of coupling between the potentials. I.e. the ALC-ALC RDF is not solely dependent upon the ALC-ALC potential, but all of the potentials. More fitting targets complicate the coupling scheme and add to the ill definition of the optimization, but the coupling can be minimized with fewer potentials to optimize.

We also found a crystalline structure could not be retained if a different CG mapping scheme was used to obtain the target trajectory. Specifically, we observed convergence was not possible without explicit mapping of the chiral methyls. In an initial mapping, the methyls were mapped into the site they are currently attached to, namely the 1RNG bead and the 3RNG bead in the 9-site model. In a non-crystalline system, convergence could be observed, but as soon as the optimization was attempted for the coarse-grained crystal, the optimization never progressed. The damping factor prevents the system from having a diverging response, but the optimization was locked into an unacceptable local minimum. In addition, the average density of the system, without the CHM beads, was that of a liquid.

The inclusion of the chiral methyl bead had a significant effect on the fitting. By solely optimizing the CHM-CHM interaction and keeping the other interactions constant, the density of the system improved dramatically. From the same starting point, if all of the interactions were optimized simultaneously, the density only improved marginally. Although anecdotal, these qualitative observations suggest cholesterol cannot maintain a

crystal structure without the explicit mapping of the chiral methyls in order to preserve the rough face of the molecule. Within the simulation cell, the rough face of some molecules within different unit cells face towards each other, which may serve as a cross-linker or binder to maximize the cohesive energy of the system as a whole. Without the methyls, the coarse-grained cholesterols can slip past each other much more easily and leads the system to exhibit a liquid state.

The presence of a rough face combined with the damping factor for the update algorithm not only allows for fitting of the RDF's from crystal simulations, but also allows the CG molecules to have the same unit cell parameters as the atomistic and experimental crystals. In Table 6.1, the parameters measured from experiments,[110] the atomistic simulations, and our coarse-grained simulations are compared. With respect to our coarse-grained crystal, the parameters not only match the atomistic target well, but are also in good agreement with the experimental data. Although not shown in the table, the largest percentage difference between the CG crystal and the experimental values has a 3.02% difference between the γ angles. The other parameters with differences larger than 1.00% were the density and the length of the *c* vector. Compared to the atomistic results, the discrepancies are smaller than when compared to the experimental results. Only three of the parameters have a percentage difference between 1.00% and 3.00%: the density, the γ angle, and the *b* vector length.

The accuracy of the crystal structure can also be verified qualitatively by looking at pictures from the simulation as shown in Figure 6.8, where the atomistic structure is shown along with both CG structures. With the original 9-bead model, the overall shape

of the crystal is retained and the spacing between the molecules and their orientations are in good agreement with that seen atomistically.

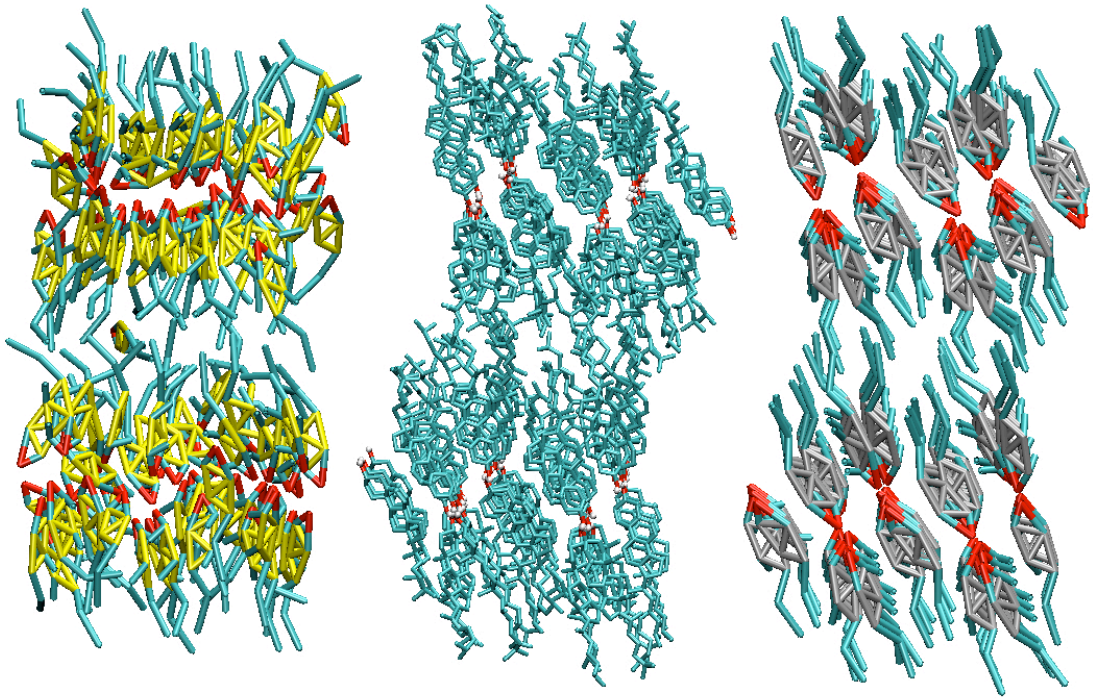


Figure 6.8 – Snapshots taken from the atomistic simulation (center), the heterogeneous CG simulation (right), and the homogenous CG simulation (left) for the cholesterol crystal in the xy-plane.

Additionally the original model exhibits the same hydrogen-bonding network, as seen in Figure 6.9. The retention of the hydrogen-bonding network can be verified quantitatively by measuring the distance between hydrogen-bonding sites in the coarse-grained simulation and in the target trajectory. In the crystal, the alcohol groups hydrogen bond with two different sites, one within the molecular pair and one between pairs in the y-direction. Within a pair, the average hydrogen-bonding distance is measured at 5.68 ± 0.24 and 5.73 ± 0.41 Å in the atomistic and coarse-grained simulation, respectively. Between molecular pairs, the atomistic trajectory produces an

average distance of 4.09 ± 0.22 Å, while the distance in the coarse-grained simulation is only 0.33% different at a value of 4.10 ± 0.44 Å.

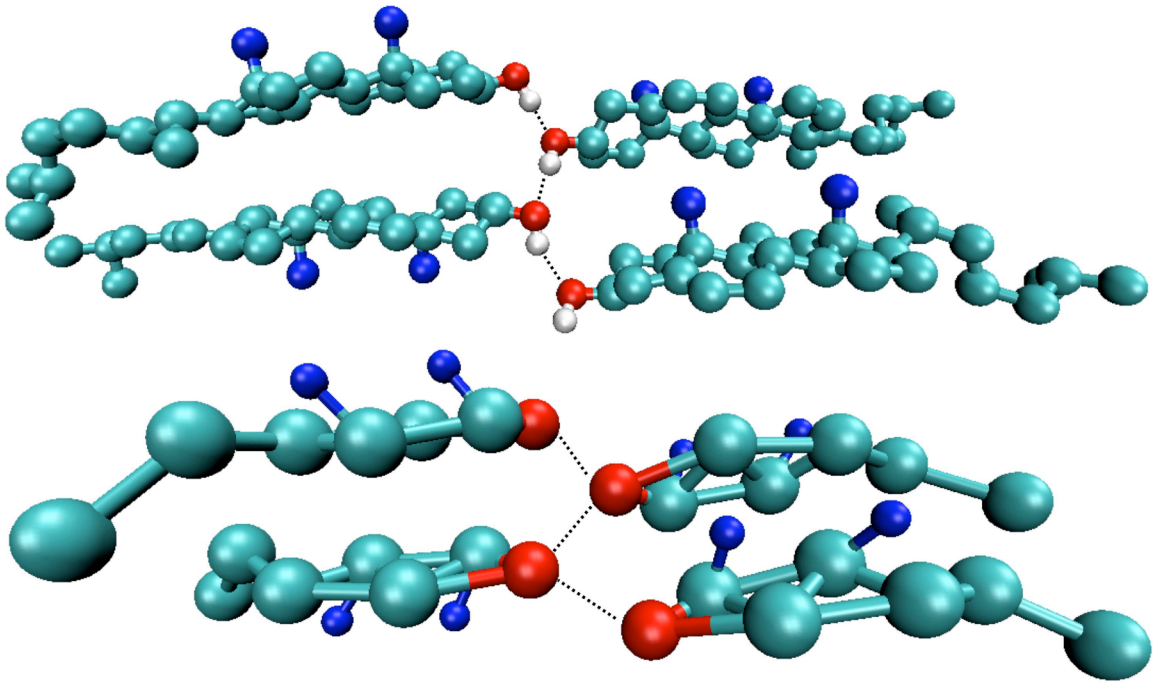


Figure 6.9 – Close-up of hydrogen bonding (indicated by dashed lines) for atomistic cholesterol (top) and for CG cholesterol (bottom).

Finally, in Figure 6.10, the yz-plane of the atomistic, heterogeneous model, and homogenous crystal structure are shown, where the coarse-grained molecules with heterogeneous ring beads maintain sheets like those observed in the atomistic snapshot. In addition, the heterogeneous model qualitatively displays similar spacing between its sheets as the atomistic figure.

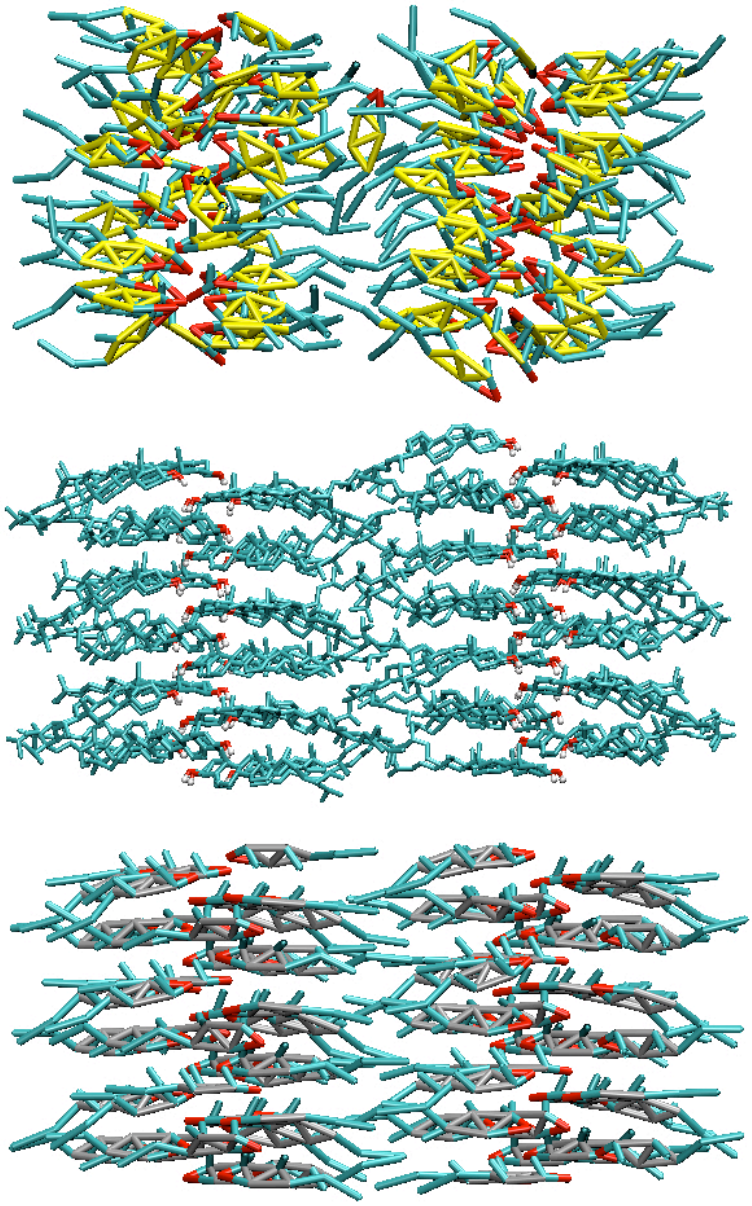


Figure 6.10 – Snapshots taken from atomistic simulation (center), the heterogeneous CG simulation (bottom), and the homogenous CG simulation (top) for the cholesterol crystal in the yz-plane.

The crystal structure obtained using the homogenized model is also included in Table 6.1. From the table, we can see that the homogenized model also reproduces the target density and lattice lengths accurately, indicating the correct volume, length, and cross-sectional area for cholesterol. However, with the homogenous ring structure, there

are more arrangements available to the molecules that can result in the same target RDF's but represent different molecular arrangements within the unit cell. Whereas, in the original model, by fitting the 10 separate RDF's associated with the 4 different ring beads, the structural intricacies can easily be retained and the molecules exhibit a packing arrangement very similar to that seen in the atomistic simulation as previously seen in Figures 6.8 and 6.10. From the homogenous model simulations, the conformations of the individual molecules are found to be significantly altered from that seen in the atomistic configuration. Mainly, the molecules rotate along the axis of the molecule, so the ALC groups are arranged the same, but specific packing of the molecules is altered. In the homogenized model, the ALC beads are still hydrogen bonding, but the molecules change hydrogen-bonding partners from the original crystal structure. The change in hydrogen bonding pairs is not systematic, so it is difficult to track which molecules are involved with a hydrogen bond, so a numerical analysis cannot be performed. I.e. The hydrogen bonding partner for molecule A of the unit cell does not switch from molecule B to molecule G in each unit cell of the overall simulation crystal. However, the ALC-ALC RDF can verify retention of a hydrogen-bonding network, even if it is different from the network in the experimental crystal.

The planar tilt found for the crystal structure of the homogenized model can also be seen in Figure 6.10. Although the rings have a significantly different tilt, the spacing of the molecules along both axes is similar to the atomistic target.

In the context of modeling the pure crystal structure of cholesterol, the model with heterogeneous ring bead types is clearly more appropriate, but, for the study of bilayers, the homogenous ring model may be adequate. Although the relative conformations of the

molecules were altered, the positions of the molecules did not change throughout the simulation. This is attributed to fitting the specific ALC-ALC interaction; the positions of the molecules cannot change significantly without the ALC-ALC RDF containing large deviations, so this interaction allows cholesterol to retain an aspect of its structure in the same manner as the specific ring bead interactions.

6.4 Conclusions

Cholesterol has unique structural features, and these features may result in unique properties in larger systems. To properly study cholesterol in any system on the coarse-grain level, a structurally relevant model is required. To ensure the CG model can exhibit the same structural capabilities as what is seen in nature, the structure of the CG molecule must be similar to the real molecule. From this manuscript, the CG molecule for cholesterol has a rigid multi-ring structure attached to a flexible tail, a rough and a smooth face on the ring structure, and amphiphilic properties.

Before the details of the model could be developed, the atomistic targets needed to be validated. In this work, we have continued testing the CHARMM cholesterol force field by comparing results for the crystal structure at body temperature. The results presented are as accurate as those reported by the original force field developers for the room temperature crystal. From this, we can conclude the force field is ample for researching the structural capabilities of cholesterol utilizing computer simulations.

We have also begun to investigate the effects of some of the unique features cholesterol possesses. Investigation of the melting point for cholesterol using self-diffusion coefficient data provides evidence the dualistic rigid/flexible nature of

cholesterol allows it to, in a way, hybridize a solid and a liquid state. In addition, from CG model optimization, we found the chiral methyls may play an important role in crystal retention for the molecule.

Optimization of a non-crystalline CG model suggests a damping factor is not only necessary for highly repeatable systems (i.e. a crystal), but also for simulations with a large amount of potentials to be fitted. With the damping factor and with our unique mapping, the CG CHOL in this paper is capable of mimicking the structural capabilities of atomistic CHOL as seen from the RDF's, the crystal parameters, and snapshots comparing atomistic and CG simulations.

A simpler homogenous coarse-grained model for cholesterol with the ring beads all of the same type was also developed, but it was not as accurate as the heterogeneous CG model. This model retains many features of cholesterol and will minimize the optimizations involved for future CG model development. Although the specific ring model is more accurate, the benefits of the homogenous ring model outweigh the drawbacks from specific crystal structure retention. Izvekov *et al*. Also found their more complex model was more accurate, but the simpler model saves research time in the long run.[109] The future use of the homogenized ring model for cholesterol was justified for this same reason.

Chapter VII

INVESTIGATION OF COARSE-GRAINED MODELS FOR WATER: BALANCING COMPUTATIONIAL EFFICIENCY AND THE RETENTION OF STRUCTURAL PROPERTIES

Water is unquestionably the most common solvent in experimental and computational studies,[113] particularly for biological systems. It serves as the basis for many properties (e.g. specific gravity and heat capacity) and is the most abundant chemical on earth. Water also has many unique properties, such as expansion upon freezing and a high surface tension, which complicates its dynamics and physics compared to other solvents.[114] Both its ubiquitous use and unusual properties make water an interesting and challenging system to study computationally and has led to the development of many different models to describe the properties and behavior of water (see for example the review of Guillot[115] for water models proposed up to 2002 and the more recent work of Vega at al.[116]).

The models for water essentially vary in terms of their choice of interaction potential and purpose. For example, the commonly used water models, such as TIP3P and SPC, are three site models with rigid bonds and angles, and point charges on the oxygen and hydrogens of the water to describe polarity. While these models reproduce many experimental properties such as enthalpy and liquid densities, other properties such as self-diffusion coefficients are not well reproduced.[117] More recent models have incorporated additional complexity to improve the ability of the model to reproduce additional properties such as the density maximum at 4˚C and isothermal

compressibilities by incorporating flexible bonds and angles,[118] polarizable effects,[119] and delocalized charges;[120,121] however, more detail does not always lead to increased accuracy. For example, Hess and van der Vegt[122] applied the SPC, TIP3P, and TIP4P-Ew water models to study the solvation of amino acids and found accuracy in reproducing experimental thermodynamic properties did not correlate with the complexity of the water model.[122] In particular, it was found the more complex TIP4P-Ew model underestimated the hydration heat capacity compared to the SPC and TIP3P models, which were in good agreement with the experimental values.

An alternative approach to the development of increasingly realistic water models is to explore the ability of simplified models to predict certain features of the properties of water. For example, work done by Nezbeda and coworkers[123] has focused on the so-called primitive model (PM), which describes water as a spherical segment with an attractive square-well potential to model hydrogen-bonding interactions and a hard sphere potential for all other interactions. Although the Nezbeda model is not suited to molecular dynamics simulations because of the discontinuous nature of the interaction potential, it is computationally efficient and able to quantitatively describe the PVT behavior of water.[123] In similar work, Dill and coworkers developed the simple Mercedes-Benz[124] water model in which water is described as a planar Lennard-Jones disk with three orientation-dependent hydrogen bonding arms. While the Mercedes-Benz model can capture, at least qualitatively, a number of the anomalous properties of water, the original version is only a 2-dimensionsal model. A 3-dimensional version has been developed by Bizjak *et al.*,[125] however, implementation within common open source codes is difficult due to the orientation-based hydrogen-bonding potential.

Although these models, compared to TIP3P[120] and SPC[126] water models for example, involve simple interactions, they are still very computationally expensive when used in studies of biological systems involving large quantities of water to be simulated over long time-scales. As a result, an additional class of water model has been developed for use in coarse-grained, rather than atomistic, simulations.[127] In a coarse-grained model each bead or site in the model represents the collective interactions of a group of atoms and are typically developed using center of mass-based methods, in which the trajectory from an atomistic simulation is mapped to the CG level and the center of mass of the atoms within a bead determines the center of mass of the coarse-grained bead. By fitting the interactions of the coarse-grained model to reproduce specific target properties or aspects of a system, an accurate model on the coarse-grained level can be derived. While several different center of mass-based techniques, including force-matching[75], reverse Monte Carlo[76], and the iterative method developed by Reith, Pütz, and Müller-Plathe (RPM),[77,78] can be used to develop coarse-grained models; they all result in numerical potentials for the coarse-grained model interactions.

For typical solute molecules, mapping schemes develop simply and naturally from the covalently bonded structure of the molecule. Subsequently, one can easily produce a target coarse-grained trajectory from an atomistic trajectory based on the center of mass of the heavy atoms within each coarse-grained bead. While, for computational efficiency and consistency one would like to map several water molecules to a single bead, center-of-mass based methods are limited to one water molecule per bead due to the loose association of water molecules through hydrogen bonding interactions. Therefore, while 3 or 4 water molecules may be closely associated in an atomistic

simulation and could be assigned to a single bead centered on the collective center of mass of the water molecules, those 4 waters are only loosely associated on the atomistic level and so will move somewhat independently of each other during the atomistic simulation, requiring reassignment of the waters to coarse-grained beads. Without an efficient method to dynamically re-assign atomistic waters to the coarse-grained beads, to date, center of mass based methods, map water on a 1:1 basis, thus limiting the computational efficiency. We note that while coarse-grained models in which multiple waters are represented by a single coarse-grained bead have been developed, for example by Marrink[81] and Klein,[128] such models are not compatible with center of mass-based coarse graining methods (i.e., an atomistic trajectory cannot be mapped to the coarse-grained level and then used to derive target properties for parameterization of the cross interactions with respect to solute molecules).

In this work, in order to develop an efficient coarse-grained model for water with multiple water molecules mapped to one bead, we utilize a clustering algorithm, called the K-means algorithm.[129] Several degrees of coarse-graining have been studied and tested to determine the optimum balance between computational efficiency and accuracy for biological systems of interest. In sections 7.1 through 7.3, we describe the new coarse-grained water model, provide details of the simulations performed to develop and test the model, and then discuss the general strategy and methodology adopted to develop and parameterize the model. In section 7.4, we present the results for the different multi-water models, consider the most appropriate level of coarse-graining, and then apply the chosen model to a simple system of biological interest.

7.1 Coarse-Grained Model and Force Field Development

All coarse-grained models are designed to retain key features from the atomistic simulations, typically at the cost of accuracy in other properties. We are primarily interested in developing a coarse-grained water model to be used in biological simulations of self-assembly; therefore, our model will be optimized to retain structural features rather than focus on phase and PVT behavior. To aid in the model development, atomistic simulations have been performed for pure water, selected pure solutes, and solute/water mixtures, and the trajectories mapped to the coarse-grained level, as described above, using the center of mass of the atoms contained within a specified site. To ensure the structural behavior of water and the solutes are retained on the coarse-grained level, radial distribution functions (RDF) from the atomistic trajectory mapped to the coarse-grained level serve as the target properties for the optimization. The coarse-graining method of Reith, Pütz, and Müller-Plathe (RPM)[77] has been used to determine the coarse-grained force field by iteratively optimizing the interactions until the coarse-grained RDF matches its target.

In order to develop a coarse-grained model for water that maps multiple water molecules to one bead to be as coarse-grained as a typical coarse-grained beads containing 3 - 5 heavy atoms per bead, we have explored the use of the K-means algorithm.[129,130] The K-means algorithm finds the optimal grouping of a large number of data points, which, in our application, corresponds to the coordinates of the atomistic water molecules. The K-means procedure locates which waters (the data points) are clustered together and determines the coordinates for the cluster (the location of the coarse-grained bead). As such the K-means algorithm allows for a dynamic mapping

scheme, i.e., it allows for the allocation of specific waters to their coarse-grained beads to change from frame-to-frame of the atomistic simulation trajectory.

The algorithm is schematically illustrated in Figure 7.1; the first step requires allocation of the number of clusters, *k*, to be used. For our model development, *k* is equal to the number of water beads used to model water on the coarse-grained level and so relates to the degree of coarse-graining of the water. As shown in the example in Figure 7.1, if we map four waters (the squares) to each bead (the circles), a system with twelve waters would contain three beads (i.e., *k* equals three). The next step is to determine an initial location for each of the *k* clusters, which is chosen from the coordinates of random water oxygens found within the simulation. After initializing the positions of the beads, each data point (the location of water oxygens) is allocated to the cluster with the smallest distance between the cluster and data point. In the example provided, two waters are allocated to the green cluster, six waters are grouped within the red cluster, and four waters belong to the blue cluster. Once the allocation is determined, the center of mass of the waters within a cluster is calculated as the new coordinates for that cluster, as indicated by the arrows in Figure 7.1. These steps are repeated until the termination criteria,

$$\sum_{n=1}^{k}\left(r_n^{i+1} - r_n^{i}\right)^2 < tol\,, \tag{7.1}$$

has been satisfied, where r_n^i represents the location of bead *n* at iteration *i* and *tol* is the tolerance set by the user (0.01 Å was used in this work). In the example provided, convergence is achieved in two iterations. For each subsequent frame of the trajectory, the location of the clusters from the previous frame is used as the initial guess for the next frame, to reproduce a more continuous target trajectory.

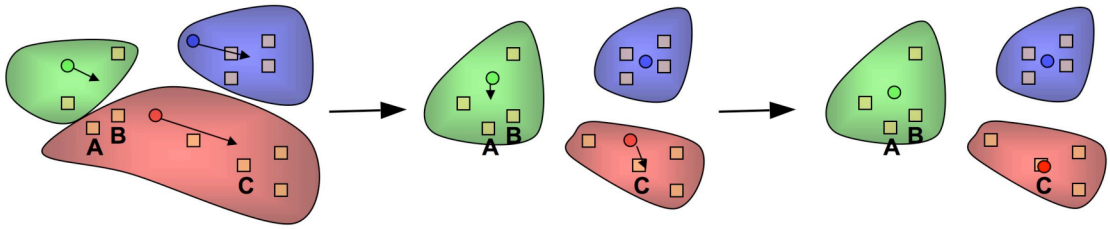


Figure 7.1 – Schematic illustration of the K-means algorithm. Circles represent cluster locations, squares represent water locations, and shaded regions represent the allocation of waters to each cluster in a color-coded fashion.

In our example, waters A and B were part of the red cluster, initially, but eventually were allocated to the green cluster allowing for each cluster to represent the target number of four waters for each bead. Potential concern is raised if waters A and B are closer to water C, initially, resulting in allocation to the red cluster, and, ultimately, two waters would be mapped to the green cluster and six to the red cluster. We observe this behavior in the mapping of our water simulations, but the ensemble average dictates there is an average of four waters to each bead with three beads and twelve waters. In addition, with a different initial guess, the allocations would be different. If analyzing one frame of a trajectory, this would cause a problem, but each new frame in a trajectory contributes to the ensemble average of the target RDF. From different initial guesses, the output target RDF's were found to be essentially identical, indicating the target RDF is relatively independent of the initial guess.

Using the K-means algorithm, a user-defined number of water molecules can be assigned to each bead. Thus, the goal of this work is to determine the degree of coarse-graining that provides an optimum balance between accuracy and computational

efficiency. Water models containing 1, 3 - 6, 8, and 9 atomistic waters mapped to each bead (denoted H2OX, where X is the number of waters mapped to each bead) were parameterized and studied through simulations of pure water and simple mixtures of water and amphiphilic solutes. 1-pentanol was chosen as a representative solute because of its size and simplicity and hexadecanoic acid chosen as a model lipid. The CG mapping scheme for both molecules is shown in Figure 7.2. Pentanol is mapped to two equally sized beads with PALC representing the hydrophilic region of the molecule containing the alcohol group and ALK the hydrophobic alkane portion. For hexadecanoic acid, three bead types are used: HEAD to represent the acid head group, TAIL for the beads in the hydrocarbon tail, and TRM2 the terminal bead that contains the last 2 carbons in the hydrocarbon tail. It is anticipated that if the water model can properly solvate these coarse-grained molecules as mapped in Figure 7.2, it should provide the correct solvation and necessary driving forces for self-assembly in more complex biological systems.

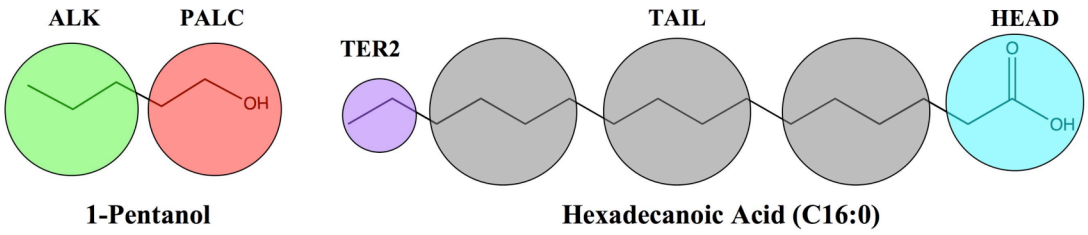


Figure 7.2 – Mapping of the 1-pentanol (left) and C16:0 (right) to the CG level.

7.2 Simulation Details

All simulations, atomistic and coarse-grained, were performed within the open source molecular dynamics program, DL_POLY 2.14.[92] The CHARMM force field[88] was

utilized for the solutes due to its accuracy with respect to biological molecules and TIP3P[120] was used as the model for water since solvation within the CHARMM force field is based on this water model. The Nosé-Hoover[99] algorithm was used for temperature and pressure control as needed.

The pure TIP3P water simulations initially contained 901 water molecules and the pure solute systems contained 100 molecules, each at a density comparable to their experimental values. The mixture of 1-pentanol and water contained 75 solute molecules and 452 solvent molecules and the acid mixture contained 50 solute molecules and 833 waters. For all simulations, the solute molecules were placed on a grid in the initial configuration and water was added as needed. The pure water system was equilibrated for 500 ps and subsequently ran at temperatures of 298 K and 333 K and a pressure of 1.0 bar in an NVT ensemble. For both pure solute simulations and solvated simulations, the molecules were equilibrated for 1.0 ns to randomize the configuration of the molecules prior to the 1.0 ns simulation used to gather results. The pure alcohol and alcohol/water systems were run at 298 K and the simulations involving hexadecanoic acid were performed at 333 K in an NVT ensemble. In all of the coarse-grained simulations, the simulation conditions matched those implemented on the atomistic level.

7.3 Methodology

The non-bonded potential for both the water and the solutes develops from an iterative optimization to fit the RDF's from the atomistic target trajectory mapped to the CG level using the RPM method as discussed in detail in chapter IV. The force field for the solute molecules also requires intramolecular potential parameters be derived from the target

trajectory (also discussed in detail in chapter IV). The potentials for the pure interactions and cross-interactions are optimized for the water models, 1-pentanol, and C16:0 in the manner detailed in chapter IV.

To determine the most appropriate degree of coarse-graining for the water model, several water models were optimized and their individual accuracy assessed through a comparison of the results for different properties as described below. An important property to reproduce on the CG level is the density of pure water. By reproducing the density, the water in mixed systems will describe the density and structure of water in the bulk region correctly. To judge the accuracy of the models in this respect, the percentage difference between the atomistic TIP3P water density and that from the CG water models was determined from constant pressure simulations.

Another important feature of any water model is the correct solvation of solute molecules. The solvating capabilities of each water model was verified by comparing the difference between the RDF for two ALK beads from the coarse-grained mixture simulation to the target RDF via the merit function found in chapter III, equation (3.11), with the weighting function defined in equation (4.12).

Computational efficiency is another key factor and was measured by the reduction in simulation time for the pure coarse-grained water models system compared to the atomistic simulation. The computational gain is represented as a ratio of the coarse-grained simulation time to the simulation time of the pure atomistic water system as shown in equation (7.2),

$$Speed\ Scale = \frac{CG\ Time}{Atomistic\ Time}. \qquad (7.2)$$

These three criteria for assessing the most appropriate level of coarse-graining water were equally weighted and combined into a scoring function described by,

$$S_{fxn} = \frac{Speed\ Scale}{\%diff(\rho_{pure}) * f}, \tag{7.3}$$

where f is described by equation (3.11). The model with the highest value of S_{fxn} will be utilized in a mixture of hexadecanoic acid and water and potentials developed for use at a temperature of 333 K.

7.4 Results and Discussion

We first discuss the results from the development of the non-bonded potential for the different pure water models studied followed by the coarse-grained model development for the bonded and non-bonded potential for the solute molecules.

We first discuss the results from the development of the non-bonded potential for the different pure water models studied followed by the coarse-grained model development for the solute molecules. We then consider the results for the solute-solvent mixtures studied.

The RPM method is found to fit the target RDF for each water model to within line thickness. As an example the water - water RDF for the H2O1, H2O4, H2O6, and H2O8 models is presented in Figure 7.3. Similar RDF's for the remaining water models are not shown as they exhibit the same general shape and achieve the same degree of fit with respect to the target RDF's as seen in Figure 7.3. From Figure 7.3, we note the H2O1 model produces a much thinner and taller RDF as seen in Figure 7.3a, indicating a much smaller bead and higher degree of structure. The high degree of structure most likely comes from the retention of hydrogen bonds between the water sites, whereas in

the multiple water models, the hydrogen bonding is within the boundaries of the water bead and so the hydrogen-bonding between beads is of less significance compared to the non-bonded interactions between beads i.e., on the atomistic level two water molecules on the coarse-grained level assigned to two different four-water beads may be hydrogen bonded, but the potential between these two molecules is negligible compared to the potential shared between the beads representing the clusters of four water molecules. Wang *et al*.[131] observed similar behavior when they applied the RPM method to the TIP3P, SPC, and SCP/E water models to develop coarse-grained models with a single water molecule per coarse-grained bead. They found that a one-water bead with an isotropic potential was not as accurate as models where orientation is incorporated into the coarse-grained model representation, such as the PM[123] or Mercedes-Benz water models,[124,125] and concluded that this was due to the isotropic nature of the interactions in one water models which do not allow for the orientation specific hydrogen-bonding observed atomistically.[131] If the tetrahedral packing is fitted for water to eliminate the isotropic nature of the potentials, the existence of the hydrogen bonds is reinforced, and, as a result, the RDF peak becomes even taller and thinner than its atomistic target.

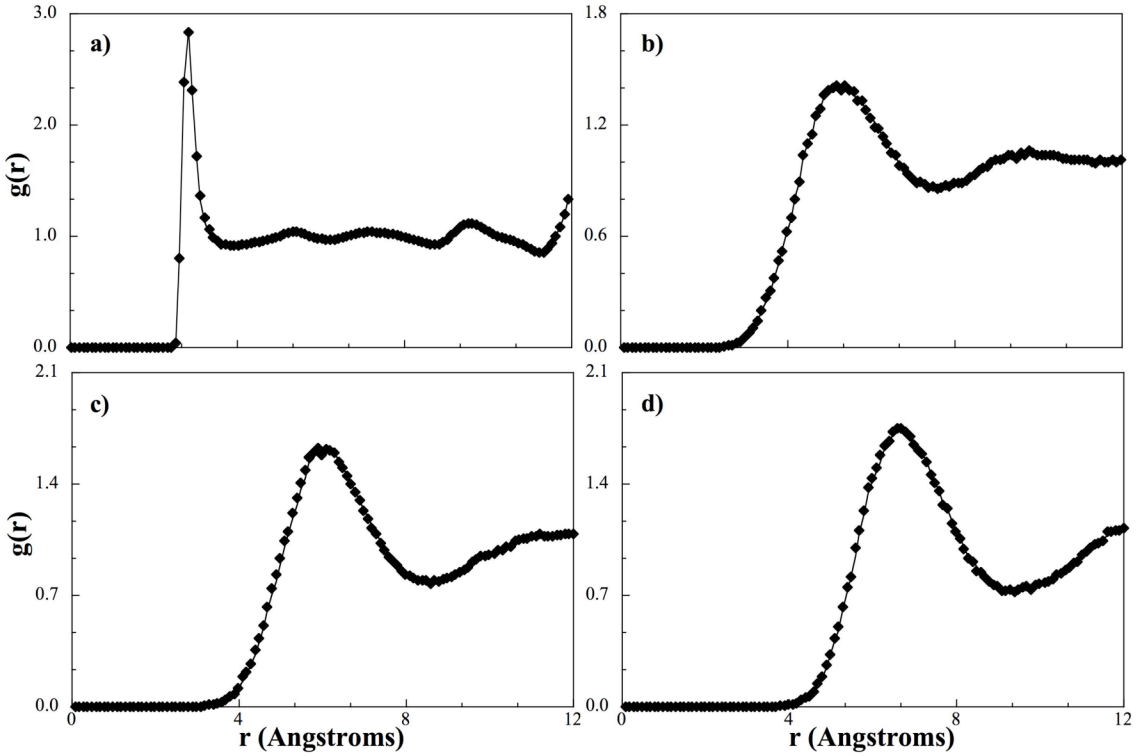


Figure 7.3 – Radial distribution function between a) one-water beads (H2O1-H2O1), b) four-water beads (H2O4-H2O4), c) six-water beads (H2O6-H2O6), and d) eight-water beads (H2O8-H2O8) from a coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

The increase in speed for the coarse-grained water models compared to the atomistic model is reported in Table 7.1, from which we see, for example, that H2O9 provides a factor of 1234 speed-up. However, the speed scale plateaus with increasing coarse-graining, indicating that the gain in calculation efficiency is greater between H2O4 and H2O3 (a ratio of 1.67) than between H2O9 and H2O8 (ratio of 1.26). Also reported in Table 7.1 is the accuracy of each model in reproducing the density of pure water. H2O4 is found to provide the most accurate water density with a difference of 0.1% compared to the atomistic value, which is reasonable if one takes into account the fact that the coordination number of water is 4.35.[132]

Table 7.1 - Comparison of the different water models studied.

Waters/bead	Speed scale	RDF merit function	Density of pure water	% difference in density	Scoring function
Atomistic	1.00	1.88E-03	0.9983	0.0%	N/A
1	16.7	1.86E-02	0.9741	-2.4%	3.76E+02
3	152	3.38E-03	0.9343	-6.4%	7.04E+03
4	254	7.75E-03	0.9996	0.1%	3.28E+05
5	401	2.35E-02	0.9815	-1.7%	1.01E+04
6	562	5.00E-02	1.0942	9.6%	1.17E+03
8	979	5.18E-02	1.0190	2.1%	9.00E+03
9	1234	5.35E-02	0.9412	5.7%	4.05E+03

Coarse-grained solute-solvent systems are simulated to test the different CG water models. With the solvent potentials optimized, the CG models for the solutes need to be developed. The bonded potential for hexadecanoic acid was derived in chapter V, and therefore, will not be discussed in detail. The alcohol contains one bond on the CG level, and the distance distribution, along with its Gaussian fit, is displayed in Figure 7.4. The molecule is flexible on the atomistic level, and as a result, produces a very wide distribution resulting in a fairly low force constant, 35.6 kJ/Å^2/mol.

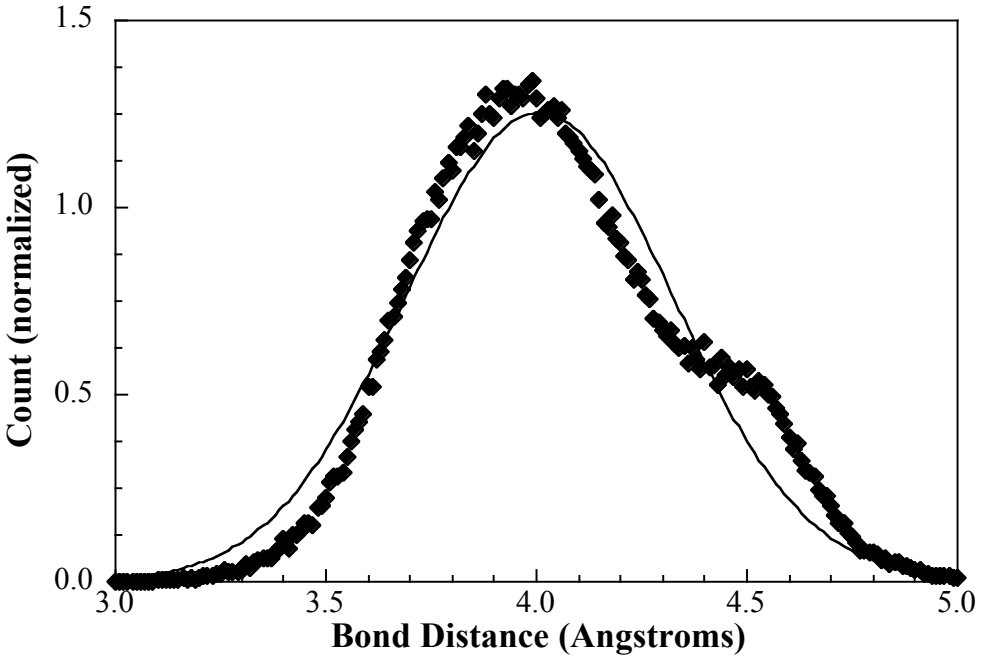


Figure 7.4 – Probability distribution for a PALC-ALK bond length from an atomistic trajectory (diamonds) and fitted by a Gaussian curve (solid line).

The potential for the three interactions of 1-pentanol (ALK-ALK, PALC-PALC, and ALK-PALC) were optimized and later transferred for use in the water mixture simulations. Due to the fact these potentials are fitted, only the RDF between ALK beads will be shown in Figure 7.5 as a representative result. To simulate a CG system of biological interest, the non-bonded interactions of C16:0 needed to be optimized for a temperature of 333 K.

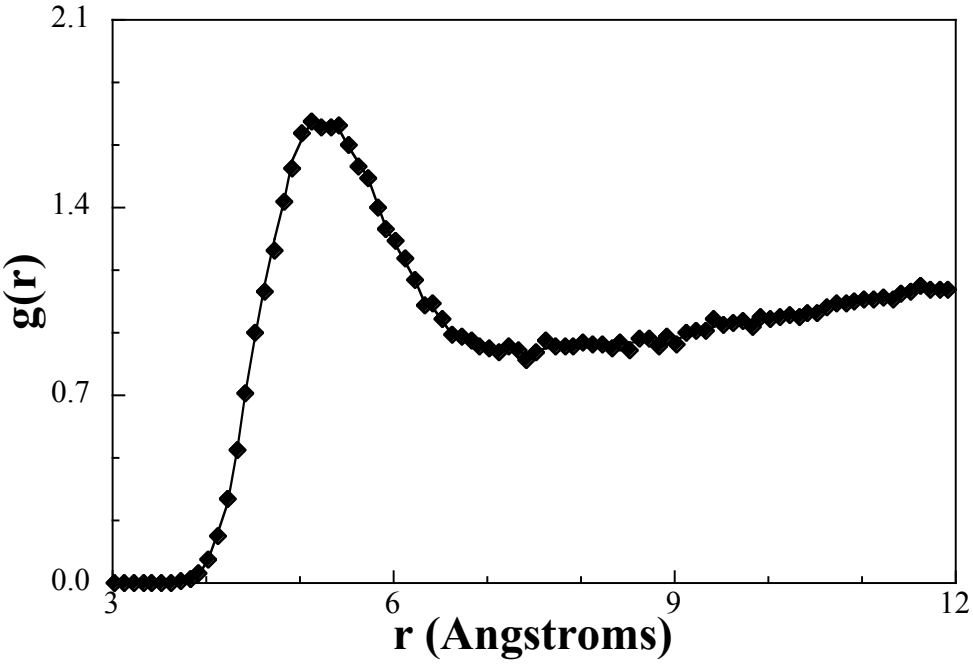


Figure 7.5 – Radial distribution function between two hydrophobic beads (ALK-ALK) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) of 1-pentanol.

We now consider the coarse-grained pentanol-solvent simulations to further test and evaluate the different coarse-grained water models. As described above, the ALK-ALK and PALC-PALC potentials used in the water-pentanol simulations were taken from the simulation of pure pentanol and used in a transferable fashion, while the cross-interactions were fitted. In Figure 7.6a and 7.6b, the RDF between the ALK beads of pentanol is compared from simulations using each of the different water models and the atomistic target. From the figures we see the agreement between the RDF and its target generally deteriorates with an increase in the coarse-graining of the water. In other words, the highest detail model (atomistic TIP3P) provides the best solvation for the coarse-grained solute, and the H2O9 model does not properly solvate the 1-pentanol. Although not shown, this trend is seen in the comparative RDF's for the PALC-PALC

and ALK-PALC interactions. An exception is seen for the pentanol-H2O1 system shown in Figure 7.6a, where more coarse-grained models provide better agreement. We believe this also reflects the differences in hydrogen bonding for the H2O1 and multi-water models, as discussed above. When hydrogen bonding is explicitly accounted for on the single water scale (like with atomistic water), the structure of the hydrophobic beads is retained.

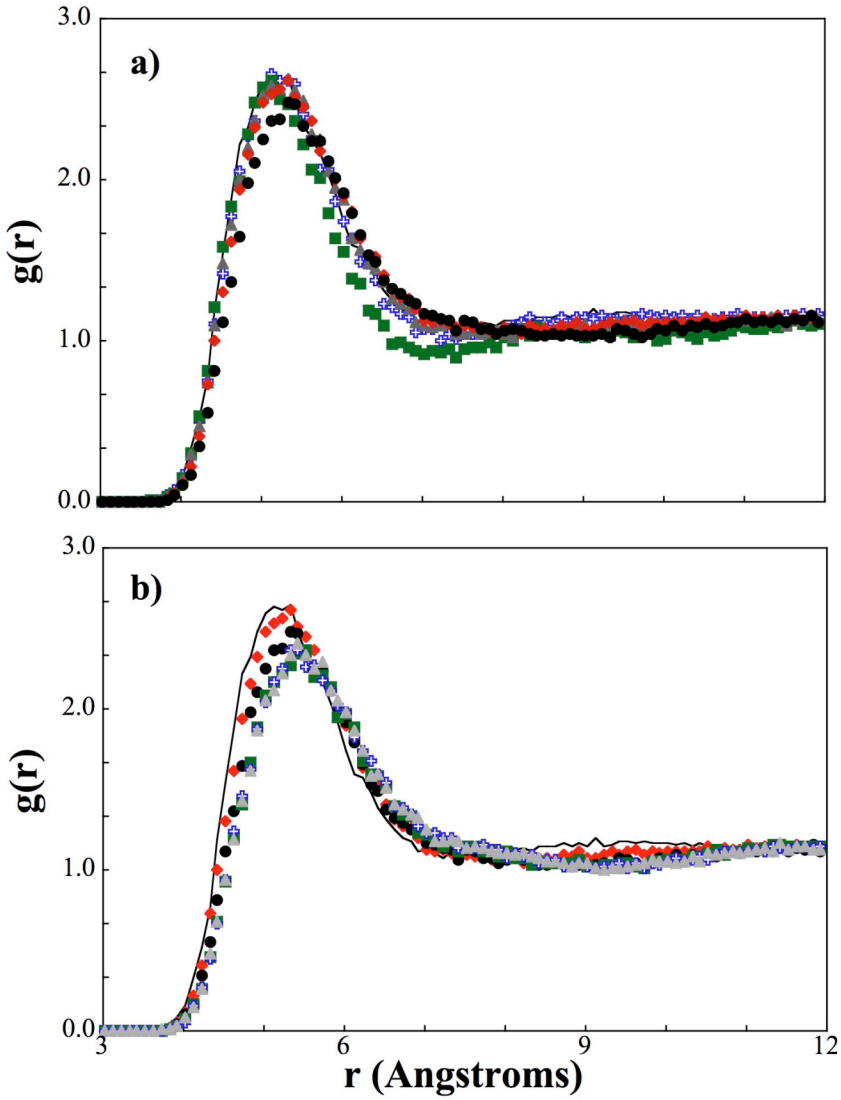


Figure 7.6 – Radial distribution function between two hydrophobic beads (ALK-ALK) within a mixture of 1-pentanol with a) atomistic water (crosses), H2O1 (squares), H2O3 (triangles), H2O4 (diamonds), and H2O5 (circles) compared to the target atomistic simulation (solid line). In addition, radial distribution function between two hydrophobic beads (ALK-ALK) within a mixture of 1-pentanol with b) H2O4 (diamonds), H2O5 (circles), H2O6 (squares), H2O8 (crosses), and H2O9 (triangles) compared to the target atomistic simulation (solid line).

In Figure 7.7, we compare the water-water RDF from simulations of the water-pentanol mixture for a multi-water coarse-grained model (H2O4) and, in Figure 7.8, a

single-water coarse-grained model (H2O1) with the atomistic target data. From the figure we can see that the level of interaction between H2O4 beads is in good agreement with the target data, as indicated by the agreement in the RDF's, while the pure H2O1 potential is lower than the target value. Although the RDF in the pure water case for H2O1 is much taller than the other water models because of its hydrogen bonding, it is not structured enough to produce an appropriately strong interaction in the mixed system. The multi-water model on the other hand has the necessary water-water interaction strength, since it does not require the pure water potential to account for hydrogen bonding, i.e., the hydrogen bonding is predominately found within the coarse-grained bead rather than between beads. As a result an isotropic interaction accurately accounts for the water-water structure in the multiple-water coarse-grained models.

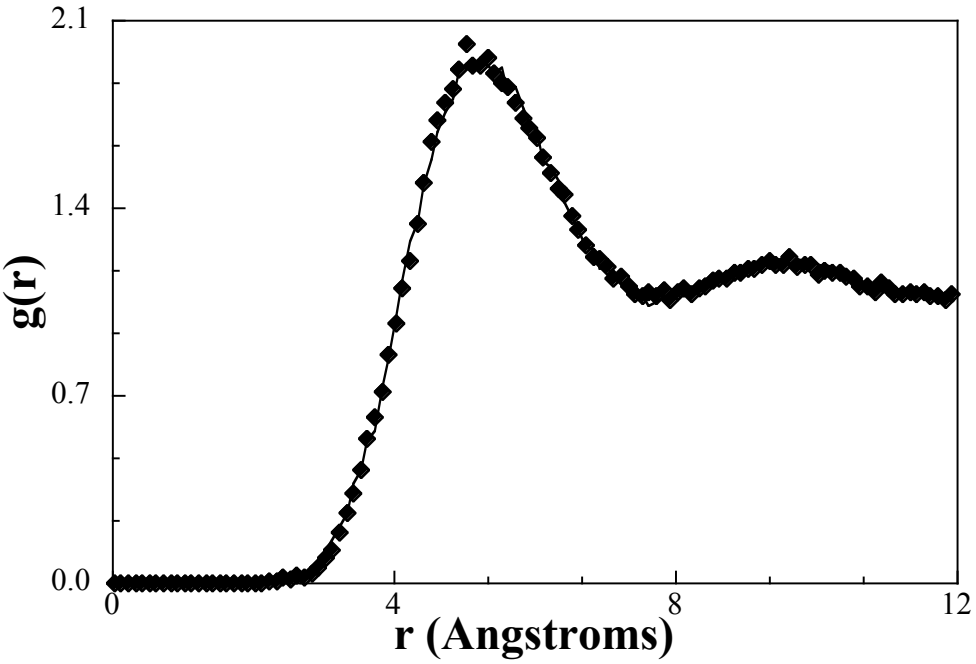


Figure 7.7 – Radial distribution function between four-water beads (H2O4-H2O4) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) from the water/1-pentanol mixture.

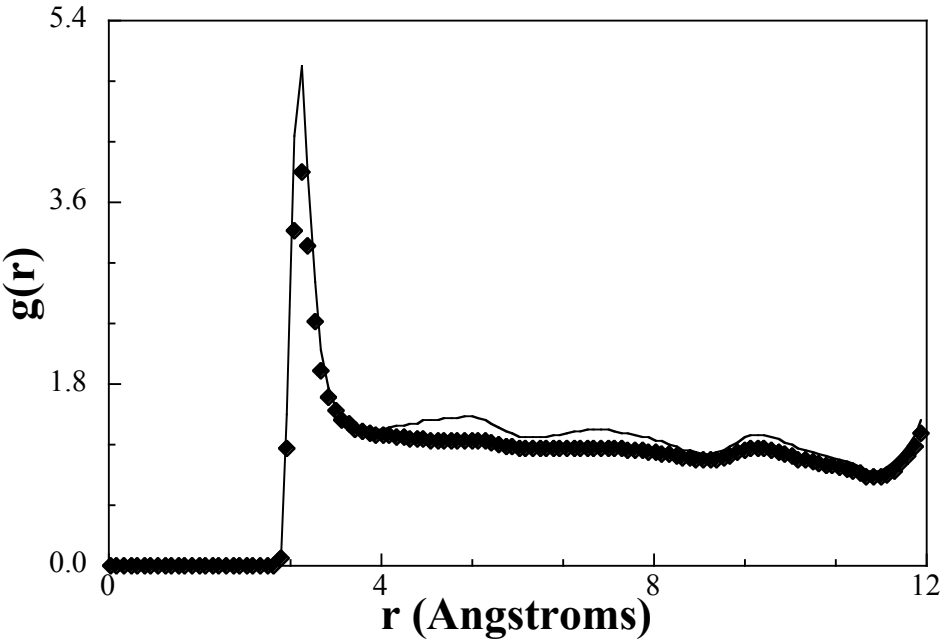


Figure 7.8 – Radial distribution function between one-water beads (H2O1-H2O1) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) from the water/1-pentanol mixture.

Wang *et al*., also observed a lack of structure in one-water coarse-grained models for atomistic TIP3P, SPC, and SPC/E water,[131] finding that if RDF's were fitted via the RPM method the tetrahedral packing could not be reproduced. Conversely, if the potential was modified to reproduce the tetrahedral packing, the resulting RDF exhibited a much higher degree of structure. Based on this work, better agreement for the water RDF in the pentanol-H2O1 mixture could likely be achieved by re-optimizing the H2O1 model to reproduce the tetrahedral packing in pure water; however, a discrepancy in the pure RDF would result.

Based on the results from the pure and mixed systems studied and reported in Table 7.1, we can determine the scoring function for each water model based on the RDF

merit function, the percentage difference in the density for the coarse-grained models compared to the atomistic model, and the speed scale (equation 7.3). From the table, we note that the H2O4 model has the highest scoring function, indicating it has the optimal trade-off between speed and accuracy, even though H2O3 is the most accurate with respect to the transferred RDF.

The four waters-per-bead model (H2O4) was, therefore, chosen as the coarse-grained water model to study the water-hexadecanoic acid system. In this preliminary study of a solvated lipid system, the cross-interactions between water and hexadecanoic acid (H2O4-HEAD, H2O4-TAIL, and H2O4-TER2), were optimized at 298 K and 1.0 bar. After optimization, the RDF of the transferred pure TAIL-TAIL interaction was compared to its target as shown in Figure 7.9a. The agreement between the coarse-grained and target RDF is good with the location of the peak indicating the hydrocarbon tails are structuring themselves on the coarse-grained level in the same way as on the atomistic level. The height of the coarse-grained RDF is somewhat higher than its target, indicating a higher degree of clustering of the tail beads in the coarse-grained model compared to the target. In a bilayer system, this behavior could lead to a stronger tendency to phase separate from water and induce structure in the hydrophobic region of the bilayer. In Figure 7.9b, we see a similar trend for the H2O4-H2O4 RDF using the water interaction transferred from the pure simulation, indicating structure of water in the bulk phase is retained on the coarse-grained level. In the RDF of the transferred HEAD-HEAD interaction (Figure 7.10) the target is found to be significantly lower than the RDF seen in the coarse-grained simulation.

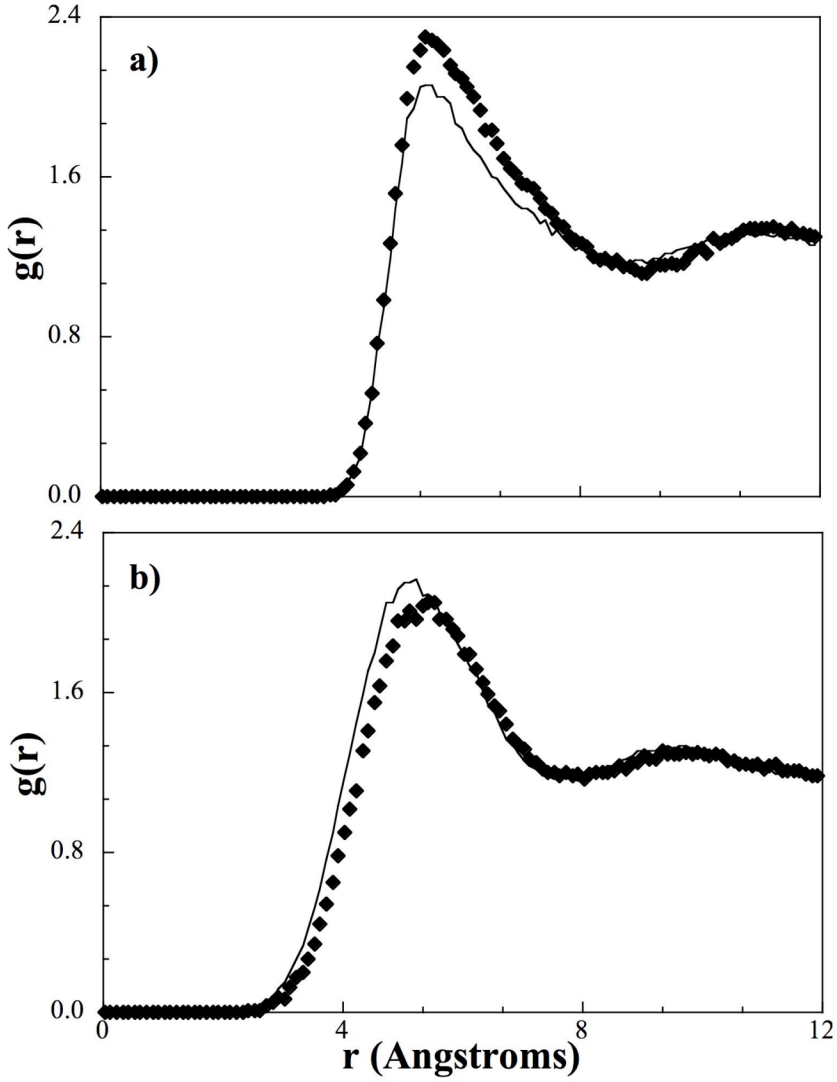


Figure 7.9 – Radial distribution function between a) tail beads (TAIL-TAIL) and b) four-water beads (H2O4-H2O4) for acid/water mixture from a coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

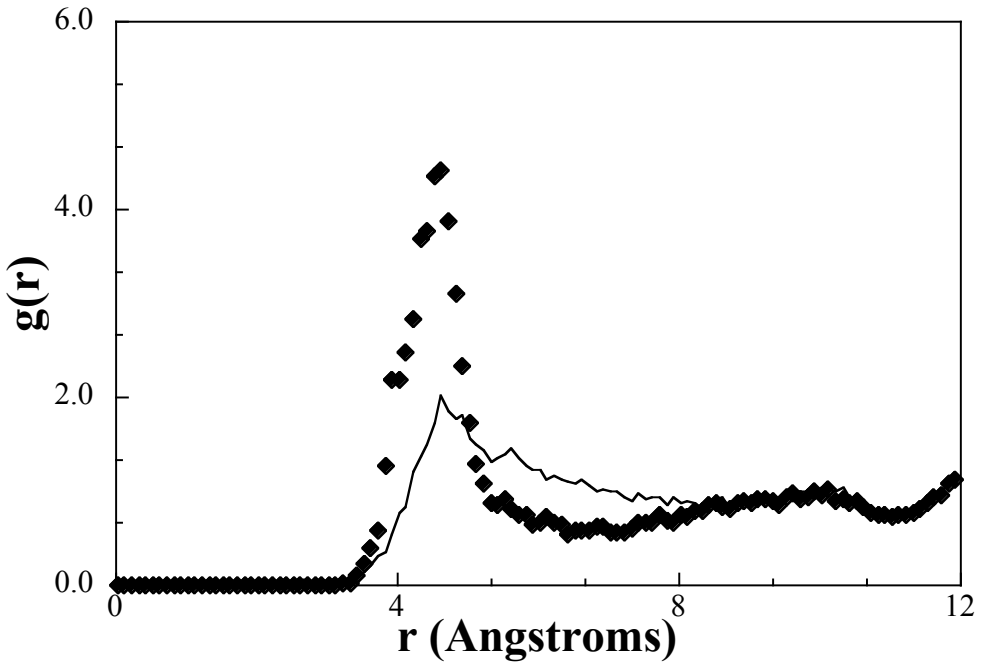


Figure 7.10 – Radial distribution function between head beads (HEAD-HEAD) for acid/water mixture from a coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

To investigate why the HEAD-HEAD RDF does not match its target, we reverted back to the simpler (and computationally cheaper) alcohol system. From Figure 7.11, we can see that the PALC-PALC interaction is also higher than the target for both the mixture with H2O4 and atomistic water. This indicates that the discrepancy is independent of the coarse-grained water model and is mostly likely dependant on the solute potential. To investigate the effect of the PALC potentials, all of the interactions (pure and mixed) for the coarse-grained H2O4-pentanol system were re-optimized against the target RDF's from the atomistic mixture system. The result of the fitting for the PALC-PALC interaction is also presented in Figure 7.11, where the RDF can be fitted within line thickness using the RPM method. If the re-optimized PALC-PALC, PALC-ALK and ALK-ALK interactions are transferred to a pure pentanol simulation, we

achieve the results seen in Figure 7.12, in which the coarse-grained peak is much lower than the target value.

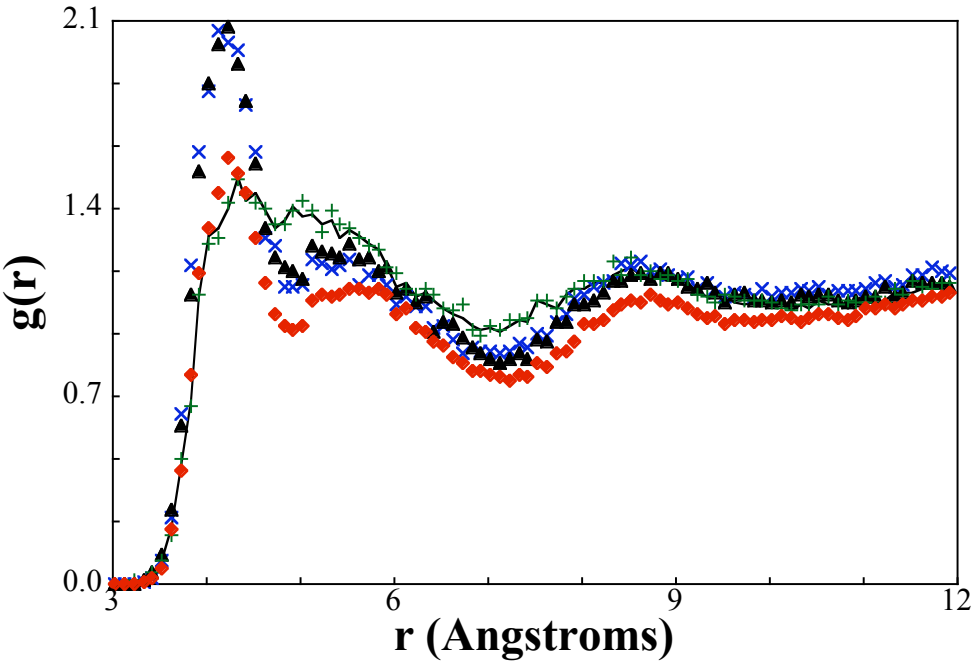


Figure 7.11 – Radial distribution function between a) alcohol beads (PALC-PALC) from a coarse-grained simulation of atomistic water mixed with pentanol (x's), the H2O4 potential transferred to the water/pentanol mixture (triangles), the H2O4 pentanol mixture with all potentials fitted to the mixture RDF's (+'s), and the H2O4 mixture with an attractive ALK-H2O4 (diamonds) potential compared to the target atomistic simulation (solid line) from the water/1-pentanol mixture.

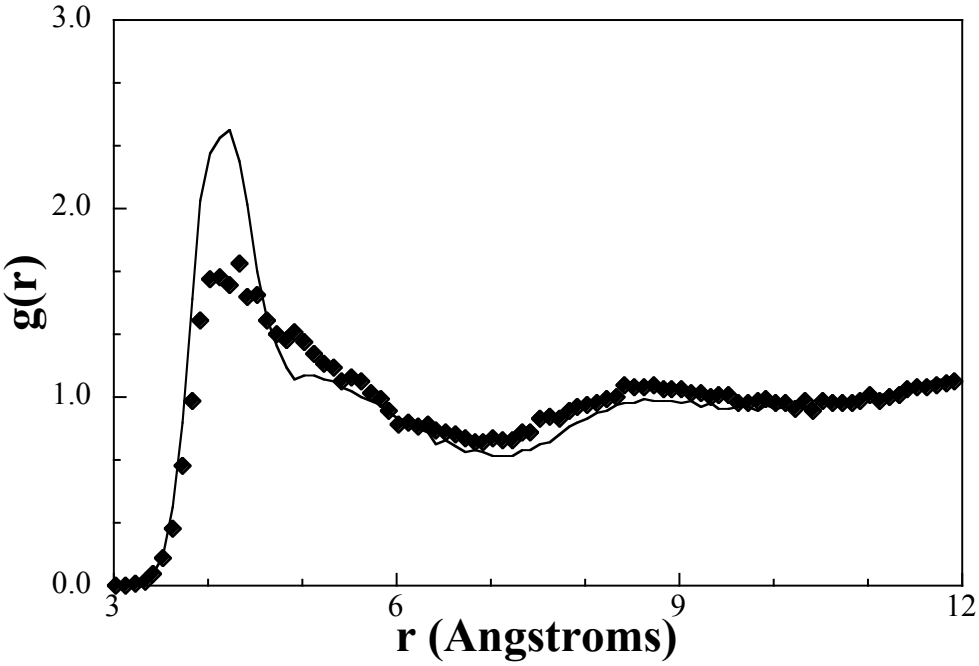


Figure 7.12 – Radial distribution function between alcohol beads (PALC-PALC) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) from pure pentanol with the potential transferred from the mixture fit.

Izvekov and Voth[109] observed similar results in their study of cholesterol within a DPPC bilayer using coarse-grained force fields derived through the force-matching procedure, where they found that the cholesterol head group, whilst having the same forces on both the atomistic and coarse-grained level, produced a strong peak in the coarse-grained RDF while only a weak peak was found on the atomistic level. As a result the interaction between hydrophobic sites and water was repulsive compared to the interaction between the cholesterol head beads and the interaction between the cholesterol head bead and water. The authors concluded that the discrepancy in RDF's was due to the fact that the forces are being derived from a system where the hydrophobic beads maintain a larger average separation distance from the water beads than the

hydrophilic beads. Based on their work we hypothesize that the observed disagreement in the HEAD-HEAD RDF is independent of our coarse-graining methodology and also due to sampling issues. The RDF between the ALK and H2O4 bead, as shown in Figure 7.13, indicates that the fitted potential will be repulsive to further support the fact. If the ALK-H2O4 interaction is replaced with a weakly attractive interaction (PALC-ALK), the coarse-grained model produces an RDF, which matches the target to a reasonable degree as seen in Figure 7.11.

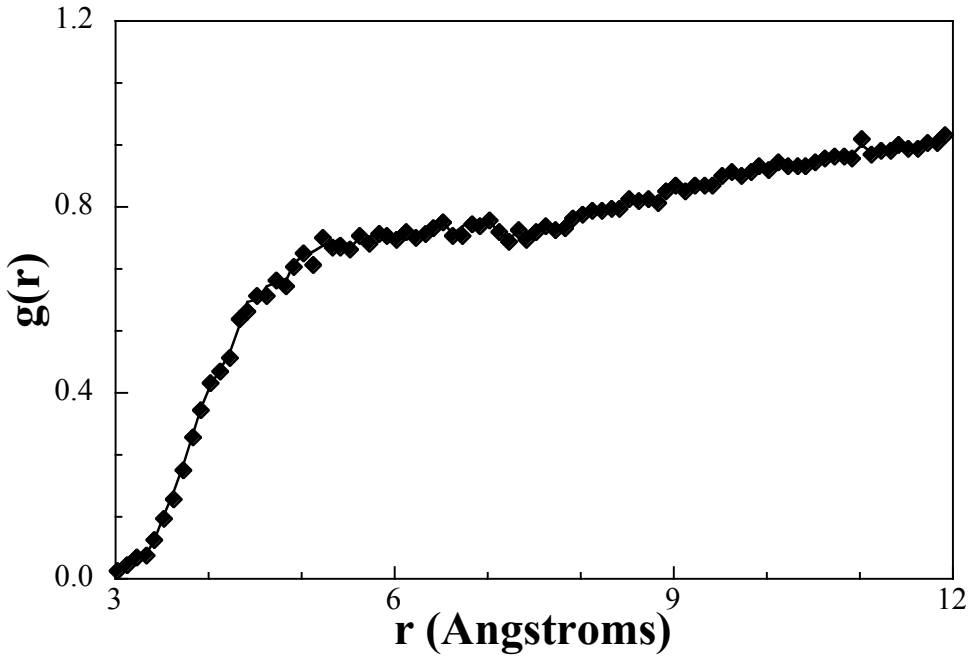


Figure 7.13 – Radial distribution function between the hydrophobic section of the pentanol and a 4-water bead (ALK-H2O4) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) within the water/1-pentanol mixture.

If the PALC-PALC interaction was fitted from the mixture RDF's, the interaction strength is too low to produce the RDF peak found in the pure pentanol system. In

addition, if the ALK-H2O4 interaction is replaced by a weakly attractive potential, the RDF of the PALC-PALC interaction transferred from the pure system to the mixture is similar to the target. This indicates that the hydrophilic head group exhibits two priorities: to hydrogen bond with itself and to provide hydrophobic shielding between the water and the carbons in the hydrocarbon tail. Because the hydrophobic sites maintain minimal contact with water, which the coarse-graining interprets as a repulsive interaction, the ALK-H2O4 interaction is repulsive. If the ALK-H2O4 interaction is repulsive the PALC beads do not need to shield ALK beads from water and are free to associate with each other, resulting in a high RDF peak. Atomistically, there is a level of attraction between the atoms of the hydrophilic bead and the water, so the atoms of the hydrophilic bead must sacrifice optimal hydrogen bonding to shield the water from the rest of the hydrophobic solute.

Another possible explanation for the discrepancy found for the pure hydrophilic RDF's in a mixture involves the isotropic nature of the hydrogen bonding. Atomistically, the alcohol group is constantly changing its hydrogen bonding pair. As such, the distance between alcohol groups has a large variance, which results in a very broad radial distribution function. Because the coarse-grained interaction is isotropic, a hydrogen bond is established as long as the beads are at a set distance from each other. This feature of the coarse-grained bead allows it to hydrogen bond with multiple neighbors, assuming they are all at the optimal hydrogen bonding distance from each other, resulting in a very small variance in bead spacing and sharp RDF peaks. In other words, to maintain an optimal hydrogen-bonding network, the orientation-dependent atomistic alcohol is continually changing its interaction distance with all of its alcohol neighbors, but the

isotropic CG interaction is at an energy minimum by sustaining its position relative to its alcohol neighbors.

Independent of the reasoning for the discrepancy in the RDF, we believe it is an acceptable anomaly. The strong clustering of the head groups suggest a strong association and binding between hydrogen-bonding lipid sites. As a result, these hydrophilic groups will be encouraged to induce structure in a bilayer phase. If the RDF from the mixture system was fitted, we predict the HEAD beads would be well dispersed throughout the lipid phase and not localized (i.e. clustered) as an interface between the water and the hydrophobic beads.

7.5 Conclusions

The K-means algorithm is designed to find the optimal clustering of data points, allowing for successful dynamic mapping of multiple waters from the atomistic trajectory to single coarse-grained beads. Coarse-grained water models representing one to nine waters per bead were optimized and studied in both the pure state and in mixtures with 1-pentanol, a representative amphiphilic solute. Based on the ability to reproduce the density of pure water, solvate the solute properly (measured by the RDF of the key solute potential), and the increase in computational efficiency, the 4-water model, H2O4, was chosen as the optimal water model for use in simulating biological coarse-grained phenomena like self-assembly.

The H2O4 model was subsequently used in simulations of a water-hexadecanoic acid binary mixture. The solvation properties of the 4-water model provided for the correct structuring of the hydrophobic component of hexadecanoic acid as shown by the

RDF between the TAIL beads derived from the mixture simulation. The RDF between HEAD beads appeared to be more attractive and structured than the atomistic target. Although in a simple mixture this may not be acceptable, we believe this clustering of the HEAD beads will be the driving force for structuring in self-assembly simulation.

Chapter VIII

SELF-ASSEMBLY OF COARSE-GRAINED SKIN LIPIDS INTO AN EXPERIMENTALLY OBSERVED BILAYER

Although the coarse-grained (CG) models presented in the previous chapters can accurately reproduce crystal structures and hydrogen-bonding networks in pure systems and simple binary mixtures, self-assembly has not yet been tested. To test the ability of the CG models currently developed for cholesterol (CHOL) and free fatty acids (FFA) to ultimately self-assemble in water into the structures observed in the SC, a simple mixture of CHOL, FFA, and water has been studied. Observing self-assembly in this simple mixture into a bilayer that matches an experimentally observed structure would support the validity of our model as a predictive tool and its appropriateness for the study of self-assembly in SC-like systems.

Many self-assembled structures of SC lipids exist in the literature as discussed in chapter II. However, many of these systems involve at least three different lipid species. We have identified a simple mixture involving only CHOL, a single free fatty acid, and water in order to test the models developed in a systematic fashion and build up to more complex systems. The best strategy for testing our models is to compare structures from the simplest mixture of SC lipids. Currently, the simplest experimental mixture capable of forming a bilayer is the C16:0/CHOL system studied by Lafleur and coworkers.[12,54] They have explored the experimental phase behavior of cholesterol (CHOL) and free fatty acids (FFA) in well-hydrated mixtures by examining the effects of temperature,

lipid composition, and acid chain length. The authors found CHOL and FFA exist in completely separate crystal phases at low temperatures; however, if the system is heated up to a temperature near the melting point of the FFA, the lipids mix and form a low-order bilayer. If the temperature is further raised, the system phase separates into separate liquid phases above the melting temperature of CHOL. At temperatures above the melting point of the fatty acid and below the melting point of cholesterol, the experimental data suggest high concentrations of cholesterol exist locally within the bilayer in cholesterol-rich domains. The structures formed vary only slightly with large changes in lipid composition from 25% to 65% cholesterol. In related work, Ouimet *et al.* observed bilayer formation with C14:0 and C18:0 mixed with CHOL, but not for mixtures containing FFA with tails longer than 18 carbons or shorter than 14 carbons.[54] In the latter systems, at all temperatures above the FFA melting point, the FFA crystal melts but remains phase separated from cholesterol.

From these results, the authors postulated criteria for mixtures of FFA and cholesterol to self-assemble into a bilayer.[12,54] At low temperatures, separate crystal phases for FFA and CHOL, respectively, are more stable than an integrated bilayer phase. At temperatures above the FFA melting point, a liquid FFA phase is at a higher energy state than FFA found in a low-order bilayer phase. Optimal van der Waals (VDW) contacts are established between the two lipids and the hydrophilic groups provide shielding from the water for the hydrophobic components of the lipids. When the FFA integrates into the cholesterol crystal, a symbiotic relationship is believed to exist between the two lipids in the bilayer phase to minimize the contact between the hydrophobic portions of the lipids and water. Cholesterol cannot maintain a bilayer

without FFA because the hydrophilic nature of the alcohol group is not strong enough to drive self-assembly. However, the hydrophilic acid group provides shielding between hydrophobic groups and water allowing the lipids to stabilize in a bilayer phase. Also, fatty acids do not form bilayers without cholesterol; rather, they form a separate crystal phase at low temperatures or a separate liquid phase at high temperatures. At high temperatures, the high-energy tails are too disordered and stay in an amorphous liquid phase versus a pure fatty acid bilayer. When mixed with cholesterol, the rigidifying nature of cholesterol is believed to induce order in the acid tail structure allowing both lipids to maintain a bilayer phase.

Ouimet *et al.* also hypothesize fatty acids with short or long tails cannot form a bilayer with cholesterol because of the hydrophobic mismatch between the two molecules.[54] C16:0 and cholesterol are approximately the same length, so they can maintain the appropriate VDW contacts. In C12:0/CHOL mixtures, the tails are too short and the cholesterol sacrifices too many VDW contacts with itself when mixing with the FFA. When the fatty acid tail is longer than 18 carbons, the fatty acids cannot constrict enough to make optimal VDW contacts with cholesterol, so they stay in a separate liquid state.

The relative simplicity of this experimental system serves as an appropriate test of the CG models developed in earlier chapters. In addition, the specific criteria for bilayer formation (i.e. fatty acid tail length) can be readily verified through simulations. Atomistic simulations have previously been performed by Holtje *et al.* for bilayers containing fatty acids and cholesterol similar to the experimental work above.[133] However, bilayers with both C16:0 and C18:0 mixed with cholesterol were studied and

bilayers with fatty acids of differing lengths could exhibit different characteristics like bilayer thickness when compared to a bilayer with a fatty acid with a uniform tail length. Also, they simulated the bilayer at temperatures well below the melting point of the fatty acids. The implications of this will be discussed in the results section, but their results cannot be used to validate the experimental findings because the simulation conditions do not match the experiments. To support the experimental findings, we performed atomistic simulations of bilayers matching the experiments as close as possible by simulating bilayers with only C16:0 mixed with cholesterol at temperatures above the melting point of C16:0. In addition to validating the experimental findings, the atomistic simulations can provide additional validation for the CG models of CHOL, FFA, and water.

Within this chapter, the simulation details specific to bilayer simulations are summarized, followed by results from atomistic simulations of the bilayer system. The development of the CG model is then presented and general heuristics on performing CG self-assembly simulations of SC lipids with respect to the effects of different initial configurations and hydration levels are discussed. Finally, results from the CG self-assembly simulations are compared to those from experiments and from our atomistic simulations.

8.1 Simulation Details

Atomistic simulations were performed on a preassembled bilayer with 144 molecules in each leaflet and containing an equimolar ratio between cholesterol and C16:0. Different initial configurations were built for the CG self-assembly simulations, but will be

discussed in a later section. The molecular dynamics (MD) simulations for the bilayer system were run using the current LAMMPS MD code[93] at a temperature of 333 K and a pressure of 1.0 bar.

To maintain consistency amongst the CG model development, simple mixtures between the components of the bilayer were simulated on the atomistic level in DL_POLY 2.14[92] at temperatures of 333 K and 1.0 bar in order to determine missing cross-interactions. As discussed in chapter IV, radial distribution functions (RDF) were measured from these atomistic trajectories to serve as the optimization targets for the cross-interactions between the CG lipids and CG water. In general, equimolar simulations of cholesterol and C16:0 and from equivolume simulations of the lipids in water were performed atomistically. The CG simulations for the CG potential optimization were also performed with the DL_POLY code.

8.2 Atomistic Simulations

In order to determine if the initial configuration of the bilayer would affect the results from the atomistic simulations, two atomistic bilayers with different initial configurations were built, and subsequently, the different configurations were equilibrated and run in the same manner as each other. In, the first configuration, the molecules were placed on an alternating grid, i.e., the nearest molecule to each cholesterol was a fatty acid and vice-versa. The second configuration involved a random replacement of cholesterols in a fatty acid bilayer while keeping each leaflet equimolar between the two lipids.

After equilibrating the two systems for approximately 500 ps in a fashion similar to the crystal equilibration discussed in chapter IV, the bilayers were run for an additional

30 ns in an NPT ensemble in which each axis of the simulation box is changed independently (similar to the NσT ensemble) but the system is kept orthorhombic. Throughout the simulation, the bilayer kept its shape and remained intact establishing a bilayer structure is stable for a mixture of cholesterol and fatty acids. Independent of the initial configuration, both simulations produced the same structures and features. In addition, migration of the cholesterols was observed in the "grid" configuration until it had a structure similar to the "random" configuration as confirmed by the lateral radial distribution functions (RDF) between the head groups (shown in later section). In both configurations, the cholesterol-rich regimes form within the bilayer, as seen in Figure 8.1. Figure 8.1 provides further support the existence of cholesterol as suggested by the experimental findings.[12]

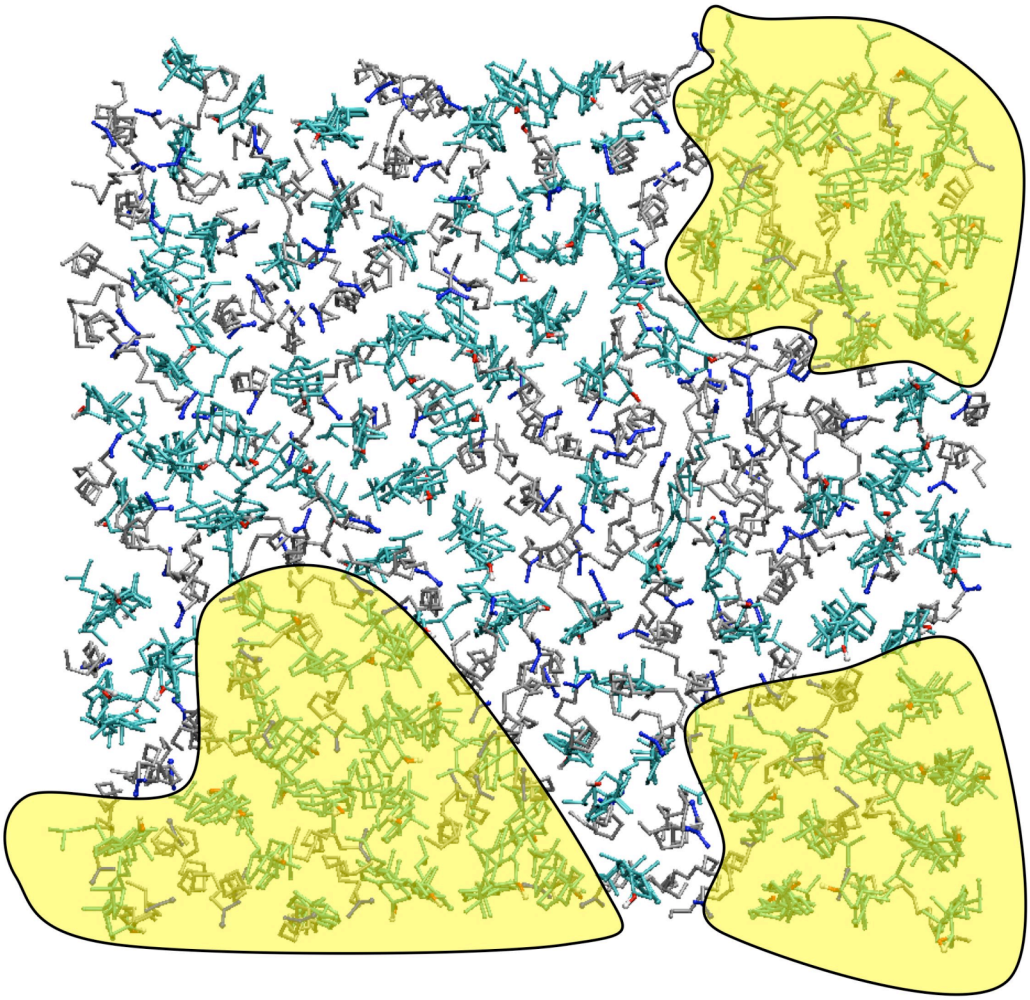


Figure 8.1 – Top down view of atomistic bilayer with cholesterol rich domains highlighted (yellow regions). In the figure, the cyan/red molecules are cholesterols and the silver/blue are free fatty acids.

The average height of the bilayer can be determined by calculating the average distance between the edges of each leaflet and was found to be 31.95 Å, which is in good agreement with the experimental value of 32.00 Å. This degree of accuracy indicates other properties (such as the average area per lipid), measured from the atomistic simulation, can be used to validate the results from coarse-grained simulations. Pare *et al*. believes the bilayer should have a height of 32.00 Å when considering the average

lengths of C16:0 and CHOL are both approximately 16 Å.[12] Holtje *et al.* reported a bilayer height of 33.1 Å, but the inclusion of C18:0 would likely increase the thickness of the bilayer.[133]

The order parameter of the lipid tails can also be determined and compared to experimental data. The order parameter is a measure of how well trans-gauche conformations are adopted within the tails as measured from

$$P_2 = \left\langle \tfrac{3}{2}\cos^2\theta - \tfrac{1}{2} \right\rangle, \tag{8.1}$$

where the brackets represent an ensemble average and θ represents the angle calculated between the bilayer normal and the vector measured between two bonded carbons. In equation (8.1), a value of 1.0 indicates the bond is completely parallel with the axis of the molecule, a value of 0.0 represents complete disorder in the tail, and a value of -0.5 means the bond is perpendicular to the reference axis. Figure 8.2 plots the order parameter measured from our simulations and compares them to those measured experimentally. The values measured experimentally follow a trend commonly seen for bilayers containing cholesterol, namely that the carbon bonds near the ring structure have a higher order parameter than those near the flexible tail of cholesterol.[66,68-70,106] The order parameters measured in the bilayers exhibit the odd-even effect as detailed by Douliez *et al*., where the order parameter for an even-numbered carbon was consistently less than that for subsequent odd-numbered carbons at temperatures above the gel-fluid transition temperature.[134] The oscillating behavior found in Figure 8.2 therefore suggests the bilayer is in a more fluid-like state than seen experimentally. However, if the tails were in a fluid-like state, we infer the bilayer height would be much smaller than the experimental value measured. A fatty acid with each bond in a trans-gauche

conformation spans a larger distance than one with a tail containing kinks and random conformations. Therefore, although the order parameter does not reproduce the experimental values, we believe the atomistic results are still valid for comparison to a CG system.

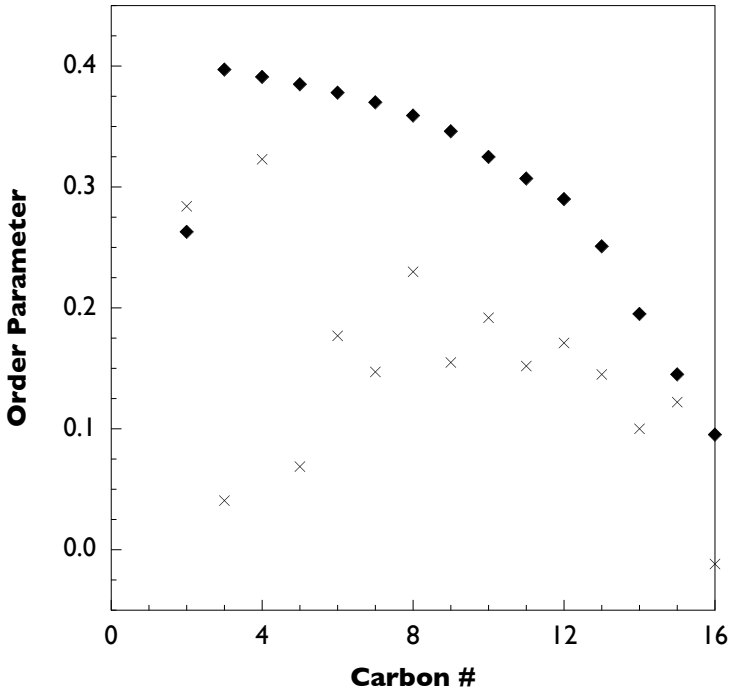


Figure 8.2 – Experimental (diamonds) and atomistic (crosses) order-parameter of fatty acid tail.

If the crystal phase is more stable for the FFA at lower temperatures, the bilayer could be assumed to collapse and the lipids phase separate upon cooling of the system. When the temperature of the bilayer was lowered to a temperature of 298 K, the density of the system increased and the bilayer retained its structure. The bilayers studied by Holtje *et al*. also retained their structure and the density of the system increased upon cooling.[133] This behavior could be analogous to a super-cooling effect, where the bilayer is stabilized before cooling and the energy barrier to disrupt the bilayer structure and

reform into phase-separated crystals is too high to overcome on the timescale of an atomistic simulation. In the experimental work, the system was studied at several different temperatures, but Lafleur and coworkers do not state if they cooled down their system after a bilayer was formed or if it phase-separated upon cooling of the mixture.[12,53,54]

The atomistic simulations were able to reproduce the average bilayer height and provide further evidence for the existence of cholesterol-rich domains within the bilayer. In addition, the atomistic results were independent of the initial configuration used for the simulation. From this, we believe the results from the atomistic simulations are valid, and additional measurements can be taken from these simulations to further verify the results of the CG self-assembly simulations. The measurements from the atomistic trajectory (i.e. average area per lipid and lateral radial distribution functions) will be presented alongside the CG results.

8.3 CG Model Development

With a CG force field already developed for the pure CHOL, C16:0, and water in previous chapters, the next step to developing the CG force field is to optimize the cross-interactions between bead types of different lipids. The general idea is to optimize the cross-interactions via the RPM method in a fashion similar to optimizing the interactions in pure lipids, while using the pure interactions transferred from the pure systems. For the bilayer system, we needed cross-interactions from simulations of CHOL/C16:0, CHOL/water, and C16:0/water.

By combining the potentials from pure lipid systems and the cross-interaction potentials from simple binary mixed systems, the force field for our SC lipid CG model can then be used to study self-assembly between fatty acids and cholesterol. To test the force field, 400 CHOL, 400 C16:0, and 7,000 water beads (35 waters/lipid) were placed in a simulation box and simulated until self-assembly was achieved. In one configuration, the lipids and water were "well-mixed" by duplicating a simple unit cell with one cholesterol and one acid hydrated by an appropriate amount of water, while a second configuration represented an immiscible system in which the acids were phase-separated from the cholesterols. In both configurations, lipid aggregation was achieved within a short simulation time, but a definitive structure never formed.

The validity of the force field was studied by "screening" the model as applied to a simulation of a preassembled bilayer. Assuming the CG force field was appropriate, a preassembled bilayer in an equilibrated state would retain its structure on the CG level, and the validity of the force field could be assessed from the final configuration of the lipids. From the screening simulation, the initial approach was determined to be inadequate as can be seen from the snapshot of the simulation found in Figure 8.3. Although the lipids stay phase-separated from water, the lipids do not remain in a bilayer phase or reform a bilayer structure.

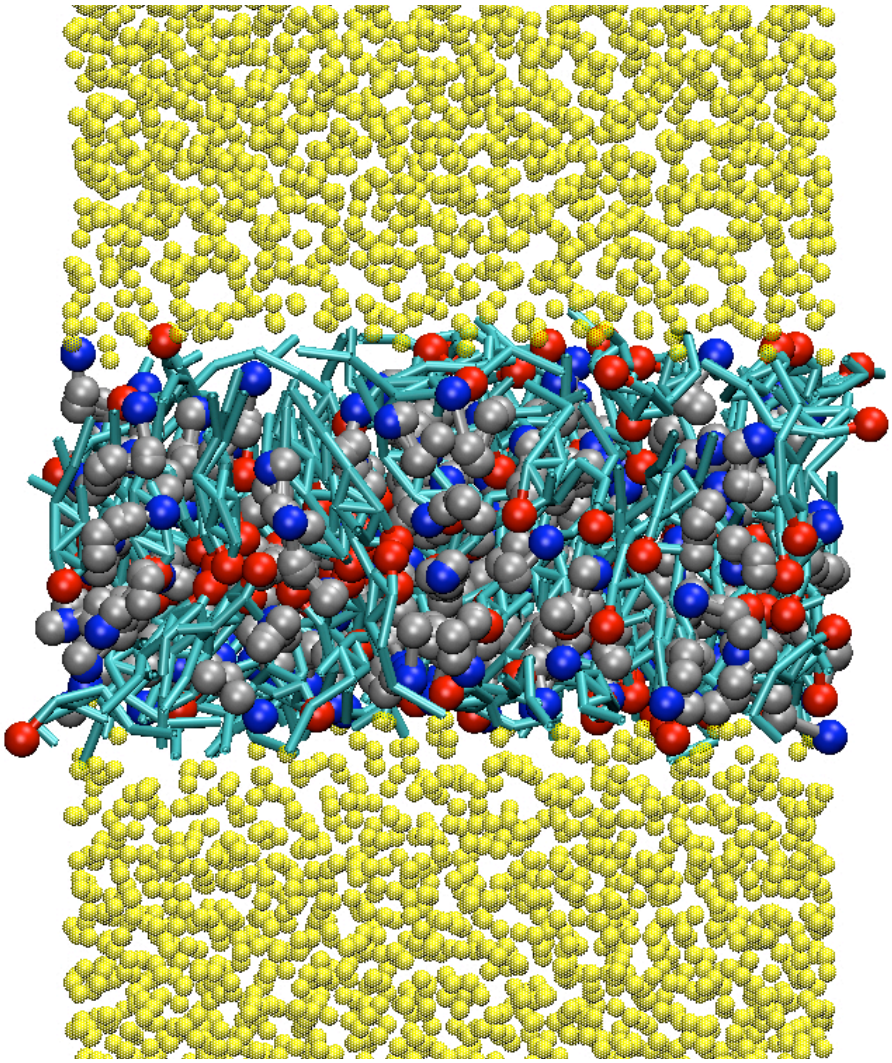


Figure 8.3 – Snapshot of bilayer in screening simulations after 5.0 ns using original force field. In the figure, cholesterol (cyan/red) and the fatty acids (silver/blue) stay phase-separate from water (yellow).

Although our initial approach for finding the cross-interaction potentials was simple and had worked well for fatty acid mixtures (as seen in chapter V), the force field developed in this way was not appropriate for studying the self-assembly of fatty acid and cholesterol in water. Alternative strategies therefore need to be tested to determine the appropriate CG potentials for cross interactions. These alternate strategies are briefly described below. The first alternative involved optimizing both pure interactions and

cross-interactions from the simple mixture systems (CHOL/water, FFA/water, and CHOL/FFA). For clarity, these potentials will be referred to as the mixture potentials and the potentials of the original force field will be referred to as the simple potentials. The alternate potential seemed to be a reasonable approach, when one considers the peculiar behavior observed in the HEAD-HEAD RDF's of chapter VII for C16:0 mixed with water. Although the tails were structuring themselves in the same manner as their atomistic counterparts (Figure 7.9a) , the HEAD beads had very sharp peaks compared to the target RDF (Figure 7.10).

If the cross-interactions from the mixture potentials are combined with the pure interactions from the simple potentials, the bilayer in the screening simulation forms a structure like the system seen in Figure 8.3. If the pure interactions are taken from the mixture potentials, two different sets of interactions can be used. For example, the H2O4-H2O4 interaction can come from the mixture potential obtained from the CHOL/water simulation or the FFA/water simulation. As a result, eight separate combinations of mixture potentials for the pure and cross-interactions were tested. None of these combinations were found to be capable of retaining a bilayer structure; rather, they either formed a phase-separated lipid region, as in Figure 8.3, or the cholesterol phase separated from the fatty acids, as seen in Figure 8.4. As can be seen from the snapshot, the cholesterol is completely miscible in water and has no driving force to aggregate or form a bilayer. With the hydrophilic interactions coming from the mixture potential, there is minimal attraction between the hydrophilic beads as compared to the other interactions, and cholesterol freely mixes with water. In addition, the fatty acids stay separated from the water, but also separate from the cholesterol.

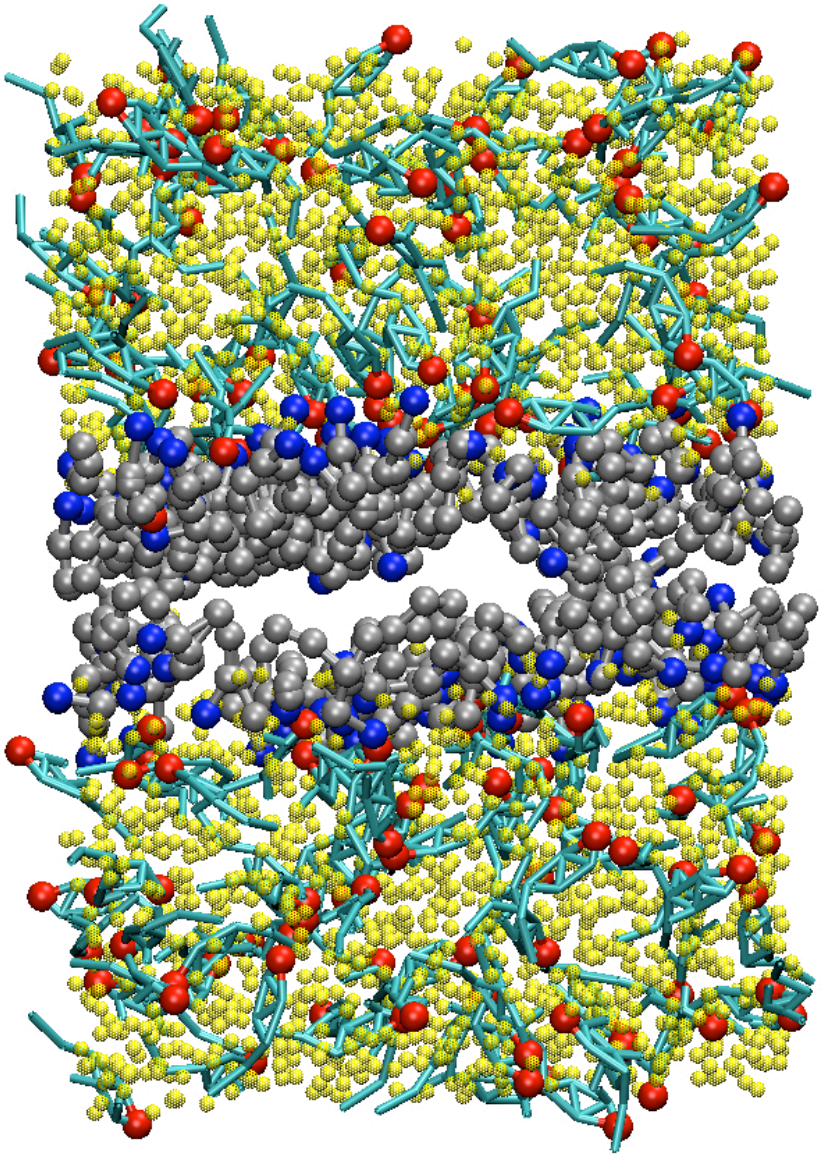


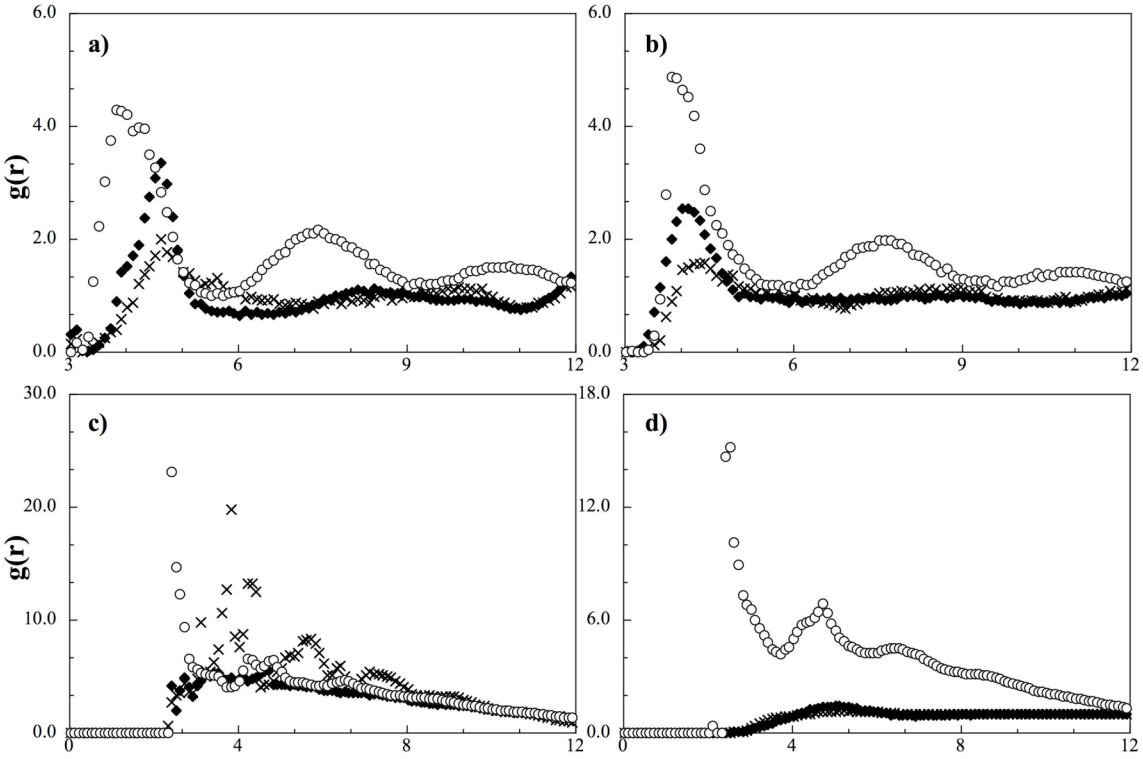
Figure 8.4 – Snapshot of bilayer in screening simulations after 5.0 ns using force field with all interactions optimized from water mixture simulations. In the figure, cholesterol (cyan/red) is well mixed in water (yellow), but is phase-separated from the fatty acids (silver/blue).

This evaluation of the shows deriving the potentials strictly from pure systems and simple mixtures does not reproduce the behavior or structures seen in a bilayer. In other words, the structures present in simple mixtures are not representative of what is found in the bilayer. Considering a bilayer can only form in a mixture of the two lipids in

the presence of water, this conclusion seems to be valid. To assess the changes that are required in the force field, all of the interactions were fitted to RDF's obtained from the preassembled atomistic bilayer simulation. The potentials were optimized from simulations of the preassembled bilayer mapped to the CG level consisted of 64 lipids per leaflet (as opposed to the original 144 lipids per leaflet). Again, for clarity, these interactions will be referred to as the bilayer potentials. By optimizing all of the interactions of the lipids in a bilayer state, the potentials can be compared and the requirements for a CG model capable of self-assembling assessed. In any system with multiple interaction types, the potentials have a coupling effect where changes in one potential affect the structuring of other bead types. For example, if TAIL-RING interaction is made more attractive, the acid will remain closer to the cholesterol molecules, and the HEAD-ALC RDF will likely have a larger peak then compared to before the change. The comparison between each potential was achieved by isolating a specific interaction to avoid the coupling effect amongst the different interactions in a system. Some potentials are vastly different in comparison but others have subtle differences making a clear assessment of what needs to be altered in the original force field difficult. As a result, a simulation of 225 beads in a cubic box of 30 Å (corresponding to the density of CG water) was modeled with only the interaction of interest studied. From the single interaction simulation, the measured RDF for the simple potential, the mixture potential, and the bilayer potential can be compared, and the qualitative differences between the potentials established. Specifically, the differences between the bilayer potential and the other potentials provide insight into how to alter the force field for accurate self-assembly simulations.

In the CG force field, we considered the HEAD, TAIL, ALC, RING, and H2O4 beads to be most important, because of their large presence in the simulations. For each pure interaction and cross-interaction between these beads, we compared the single bead simulation RDF's between each optimization strategy.

For pure HEAD bead interactions, the differences seen between the different potentials are shown in Figure 8.5a. From this figure, the bilayer potential is shown to have the strongest HEAD-HEAD interaction and the mixture potential has the weakest, with the simple potential lying in between. This indicates a potential capable of matching the RDF between HEAD beads in a FFA/water mixture does not have the strength to drive self-assembly. We obtain a similar trend for ALC-ALC interactions as seen in Figure 8.5b; the bilayer potential has the highest peak, and the mixture potential has the lowest.



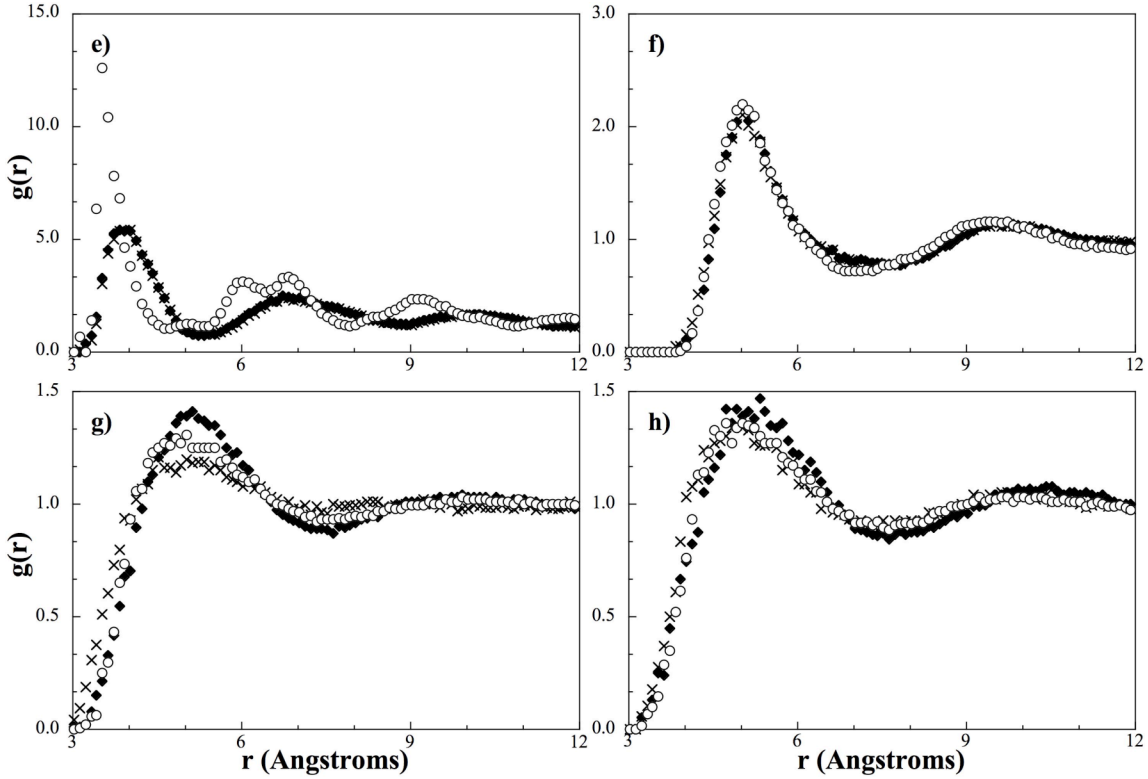


Figure 8.5 – Qualitative comparison of RDF's from a pure single bead system with different interaction potentials from the simple potential (diamonds), the mixture potential (crosses), and the bilayer potential (circles). The potential examined is a) HEAD-HEAD, b) ALC-ALC, c) HEAD-H2O4, d) ALC-H2O4, e) HEAD-ALC, f) RING-RING, g) RING-H2O4, and h) TAIL-H2O4.

A slightly different trend is seen between HEAD beads and water as seen in Figure 8.5c, while it is obvious the bilayer potential is the strongest, the interaction from the mixture potential is stronger than the simple potential. In the mixture potentials, the interaction strength between HEAD beads was decreased when compared to the simple potential. As a result, the HEAD-H2O4 interaction must compensate by becoming stronger. In addition, although the HEAD-H2O4 interaction is strong, the interaction from the bilayer potential is significantly stronger.

The interaction between the alcohol group of cholesterol and water has similar RDF's between the simple potential and the mixture potential, but the bilayer potential is completely different as seen in Figure 8.5d. The ALC-H2O4 mixture potential seems to be relatively weak compared to the mixture potentials from Figures 8.5a-c. When considering the hydrogen bonding capabilities between water and an alcohol compared to the hydrogen bonding of the other systems, this trend is expected; i.e., an acid group will have a stronger attraction to water than alcohol. If the interactions were of the same magnitude, the alcohol group of cholesterol would provide ample hydrophobic shielding and would not require fatty acids to maintain a bilayer structure. Also, the ALC-H2O4 RDF and the HEAD-H2O4 RDF from the bilayer potential exhibit a similar shape and magnitude, indicating there should be minimal preference between water and a specific hydrophilic lipid bead in the bilayer phase.

When optimizing the cross-interaction between the hydrophilic groups of the two different lipids, little difference was seen between the mixture potential and the simple potential as shown in Figure 8.5e; however, there is a significant difference between these two force fields and the bilayer potential. Additionally, the lack of similarity in the RDF's stems from the differences in structure between the simple mixture and the bilayer system. Although clustering of the different hydrophilic groups occurs, they are not as ordered or as densely packed in an amorphous configuration as in the bilayer. We hypothesize this interaction is why no definitive structure is found within the lipid region using the simple potentials. The lipids are arranged in the same manner in the lipid region with water as they are in the simple mixture between the two lipids as shown in Figure 8.6. The structure of HEAD and ALC beads in a simple mixture is not

representative of the behavior seen in a bilayer, and so a bilayer does not form with potentials optimized from simple mixtures.

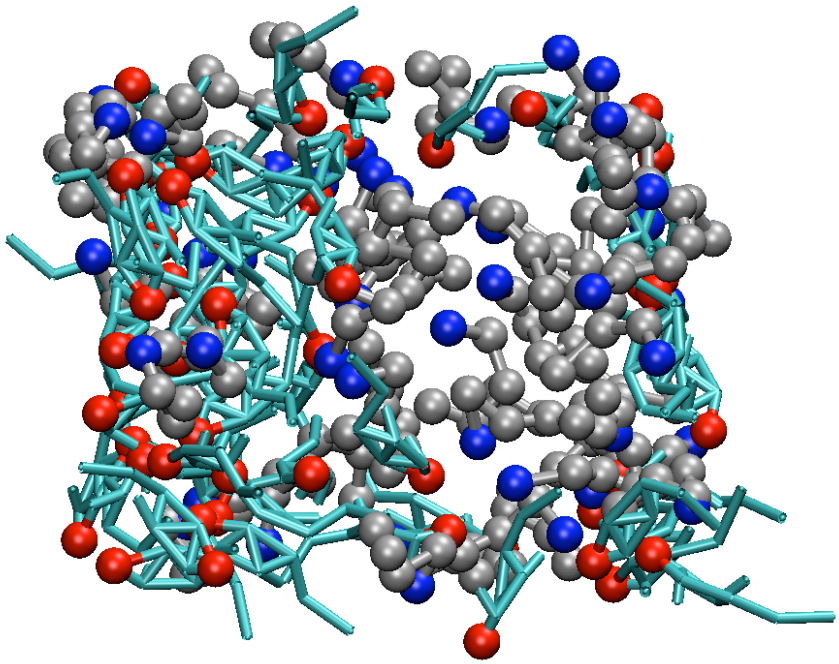


Figure 8.6 – Snapshot of simple mixture between cholesterol (cyan/red) and fatty acids (silver/blue) with a similar structure as Figure 8.3.

For interactions involving hydrophobic beads we find negligible differences between the different potentials (for an example, refer to the RING-RING RDF in Figure 8.5f). This indicates the hydrophobic groups are structuring themselves independent of the hydrophilic interactions. We hypothesize this phenomenon stems from the fact that all of the hydrophobic beads have a very similar chemical makeup (carbons and hydrogens) and from the beads' weak electrostatic interactions and negligent dipoles. Without hydrogen bonding, the atomistic counterparts will interact with themselves with the same interaction strength in most orientations. When atomistic components hydrogen

bond, the relative interactions between sites can vary from system to system because of the dependence upon orientation of hydrogen bonds, as can be inferred from Figures 8.5a-e.

The exception to the above generalization is the interactions between water and the hydrophobic beads. Although not as different as the hydrophilic RDF's, a small difference does exist between the three potentials as can be seen in Figure 8.5g for the RING-H2O4 interaction. The interaction strength between RING and H2O4 beads is slightly higher in the simple potential compared to the bilayer and mixture potential. In the simple potential, the strong interaction between head groups and the ALC-H2O4 interaction causes the ring beads to be too isolated from the bulk water. As a result, the RING-H2O4 interaction needed to be stronger to match the RDF, or else the CG RDF would be much lower than its target. The weaker ALC-ALC interaction in the mixture force field allows an appropriate mixing between the hydrophobic beads and water with a weaker RING-H2O4 interaction. With respect to the bilayer potential, water has minimal contact with the RING beads and the potential must reflect this behavior by making the interaction near repulsive. These conclusions also hold true for the TAIL-H2O4 interactions when considering the RDF's seen in Figure 8.5h.

The results of the above qualitative analysis directed the necessary changes to the original CG force field in order to develop a model capable of inducing self-assembly of the lipids in water. Altering the pure interactions was not considered to maintain a level of consistency between different optimizations for deriving cross interactions. For example, if the pure HEAD interaction and the pure ALC interaction were made stronger, the interactions between RING and TAIL beads may be affected in a negative way. The

HEAD-H2O4 potential is too weak in its current form, so the strength of the attraction between the two beads was increased in order to match the behavior seen in Figure 8.5c. Because of the similarity in RDF's from Figures 8.5c and 8.5d for the bilayer potential, we also altered the ALC-H2O4 interaction to match the strength of the HEAD-H2O4 interaction. For the hydrophilic components of the lipids to form structures as seen in a bilayer versus amorphous aggregation, the potential between HEAD and ALC beads contained a deeper well compared to the original force field. Finally, the affinity between water and the key hydrophobic beads (RING and TAIL beads) was decreased to the same degree as the difference in RDF's of Figures 8.5g and 8.5h. To achieve these changes, the interactions mentioned above were optimized in the sample bilayer system to make up what we refer to as the "updated" potential.

From the updated potential, self-assembly was achieved, but a vesicle was formed versus a bilayer as shown in Figure 8.7. Molecules capable of self-assembly either form a spherical, cylindrical, bilayer, or vesicle shape based on the area/volume ratio of the molecule. Assuming the volume cannot be changed and the area is dependent upon the HEAD-ALC interaction distance between beads, the potential was altered to exhibit a larger interaction diameter by moving the potential minimum to a larger distance. By altering the potential this way, the optimization is reinitialized to achieve a more accurate local minima. After the altered HEAD-ALC potential was reoptimized a final time, RDF's from the single interaction simulation were gathered to confirm the new optimization resulted in a potential with a larger interaction diameter. From Figure 8.8, we can see the peak has clearly shifted to a greater value indicating the interaction has a

larger diameter. As a test, the final force field was screened using the preassembled bilayer and the bilayer retained its structure.

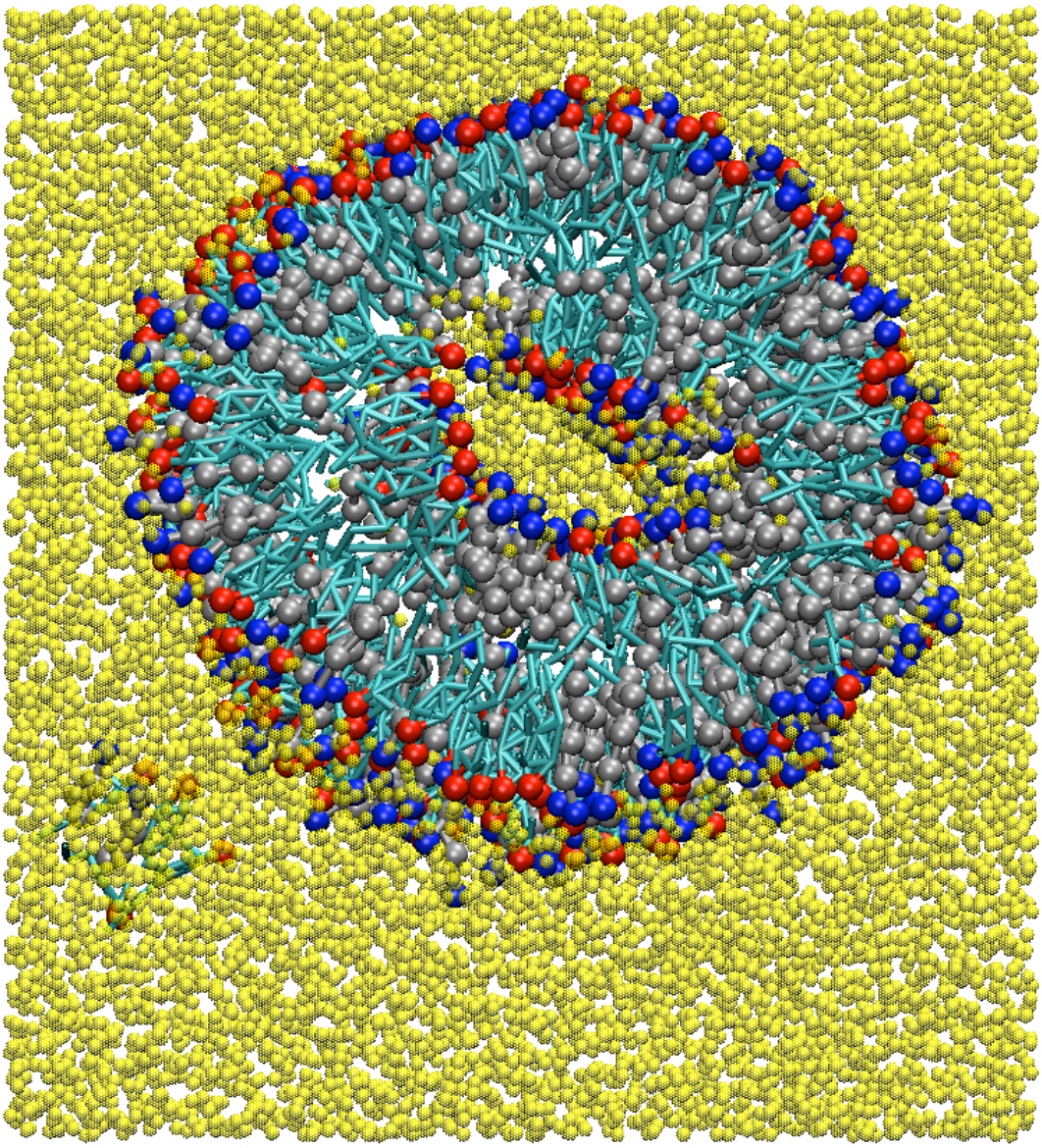


Figure 8.7 – Snapshot of self-assembly simulation forming a vesicle between cholesterol (cyan/red) and fatty acids (silver/blue) while surrounded by water (yellow).

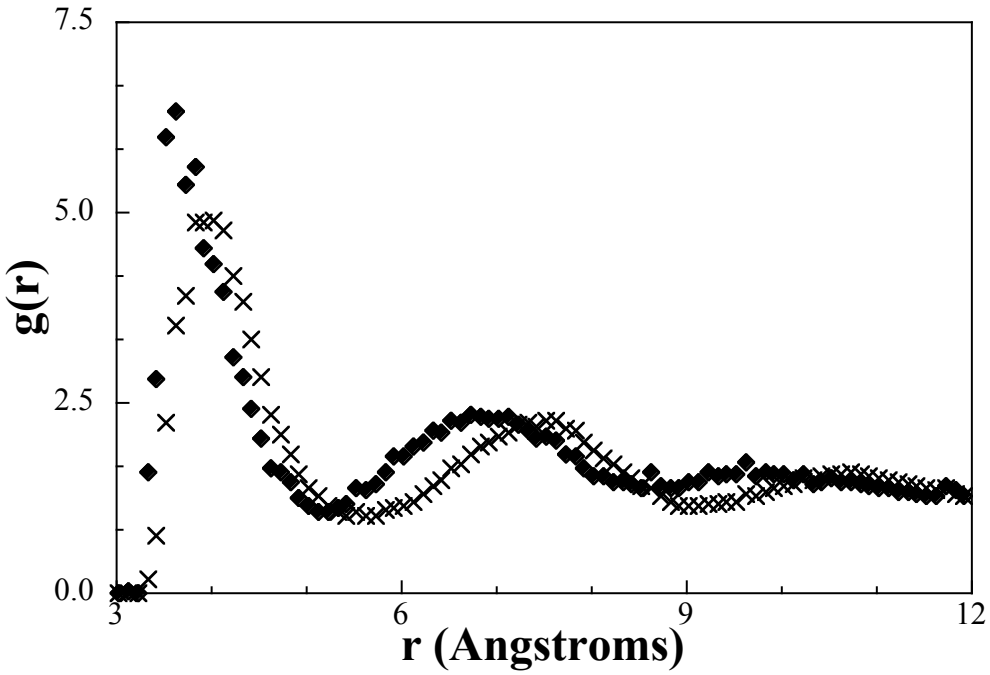


Figure 8.8 – Qualitative comparison RDF from single-bead simulation using the HEAD-ALC potential from the bilayer optimization (crosses) and a potential, which forms vesicles (diamonds).

8.4 CG Self-Assembly Heuristics

With a CG force field capable of retaining a bilayer structure, certain self-assembly heuristics were developed. Simulation conditions like temperature, lipid composition, and pressure are chosen to match the experimental conditions. The total number of lipids, the amount of water beads per lipid, the initial configuration, and the equilibration scheme all affect self-assembly simulations, but the appropriate choice for these variables is unclear and has been investigated as a result.

Initially, a system of 800 lipids with a 50/50 mixture between cholesterol and C16:0 and 35 waters per lipid was studied. Four separate initial configurations (shown in Figure 8.9) were built and simulated using the CG force field. Configurations (a.) and (b.) were built to represent an initially immiscible state. To imitate this state, the final

configurations from the CHOL/water and FFA/water simulations (which are phase-separated) were duplicated on a grid. In both, fatty acids were placed on the left and cholesterol was placed on the right, but (a) contained a 2x2x2 duplication grid and (b) contained a 4x1x2 duplication grid. The (c) configuration represents a "well-mixed" configuration as discussed earlier. The final configuration studied involved building a cholesterol crystal and a fatty acid crystal in a cell containing the appropriate amount of water as seen in 8.9d.

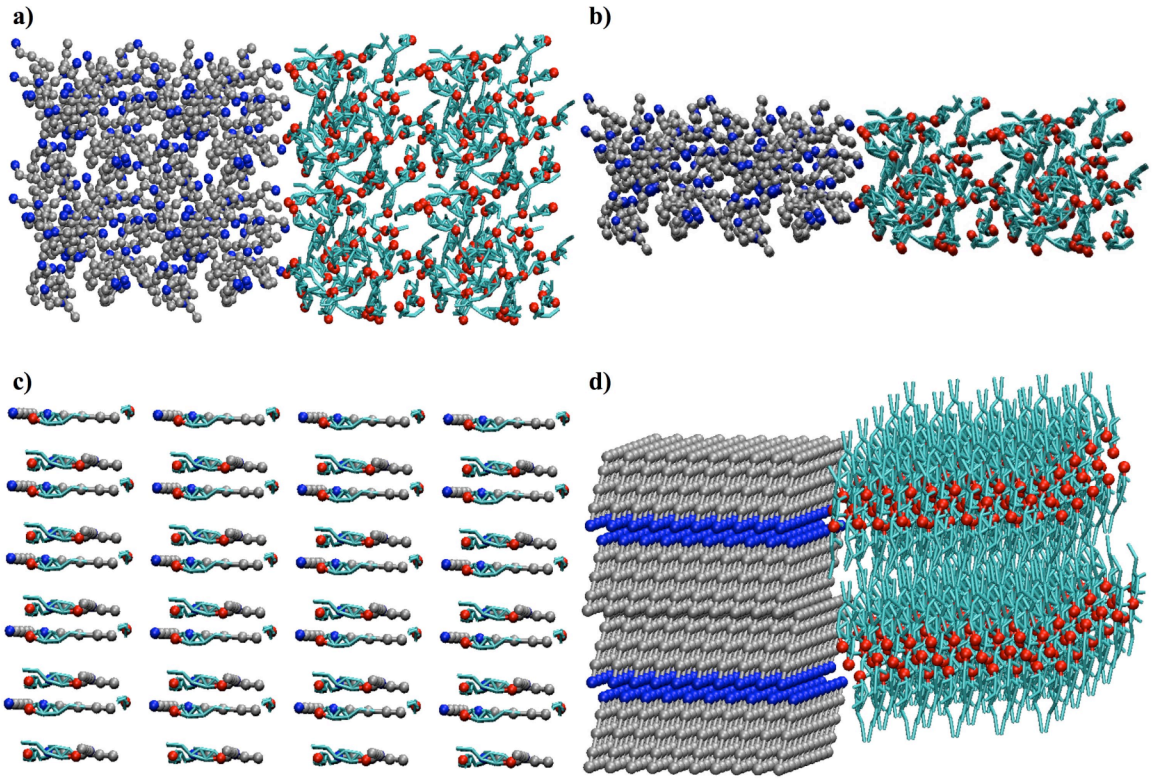


Figure 8.9 – Initial configurations of cholesterol (cyan/red) and fatty acids (silver/blue) in a) an immiscible state, b) an alternative immiscible state, c) a well-mixed state, and d) separate crystal states. Water not shown for clarity.

In all of the configurations, the lipids quickly aggregated and formed some sort of structure, with most forming a bilayer spanning the periodic boundary conditions (PBC) at an angle as seen in Figure 8.10. Independent of how long the simulations were simulated, the "spanned" bilayer was too stable to reform into a continuous bilayer in which appropriate analysis could be performed. The amount of water in the simulation was not large enough because the lipids did not have sufficient working volume to form a single bilayer within the cell. Essentially, the bilayer was stabilized prematurely by connecting pieces of a bilayer across the reflections from the PBC. By adding more water (50 waters per lipid), the lipids have additional room to move around without encountering the cell boundary. With more water, a continuous bilayer formed only in configuration (b), and branched bilayer formed from configurations (a), (c), and (d) as seen in Figure 8.11. In configuration (b), the lipids were initially in the same plane at a thickness similar to the final bilayer height allowing for self-assembly in a short amount of time. The branched portion of the bilayer in the other configurations was too stable to reform into a uniform bilayer. However, the bilayers formed in all of the configurations did not span across the PBC, as seen in Figure 8.10, so 50 waters per lipid appeared to be appropriate.

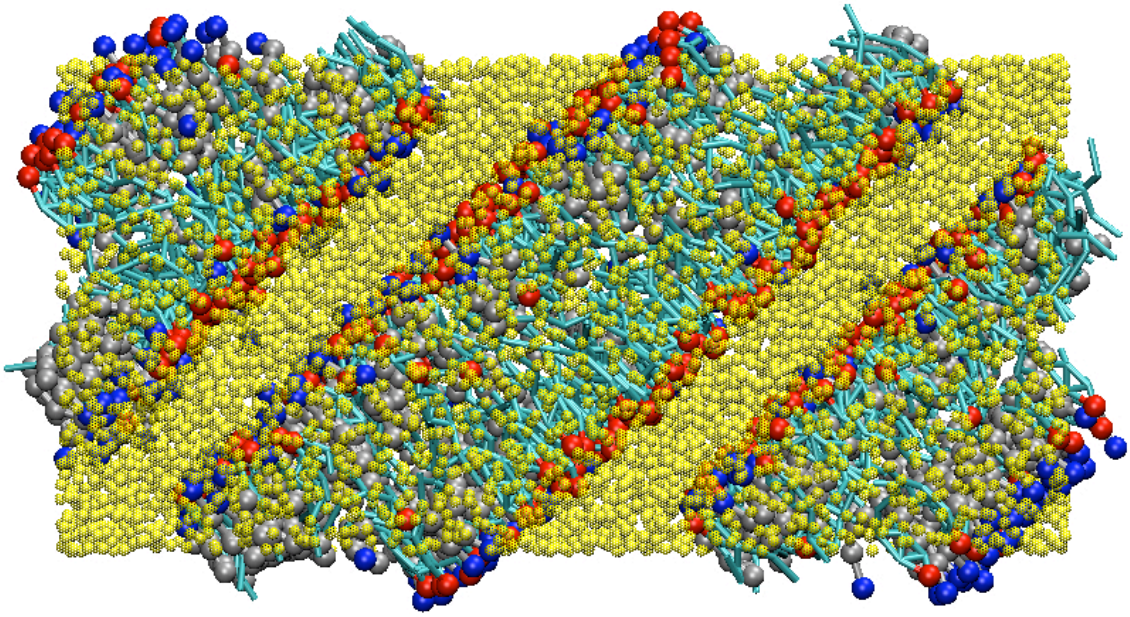


Figure 8.10 – Snapshot of self-assembly simulation where bilayer of cholesterol (cyan/red) and fatty acids (silver/blue) solvated by water (yellow) bridges across PBC.

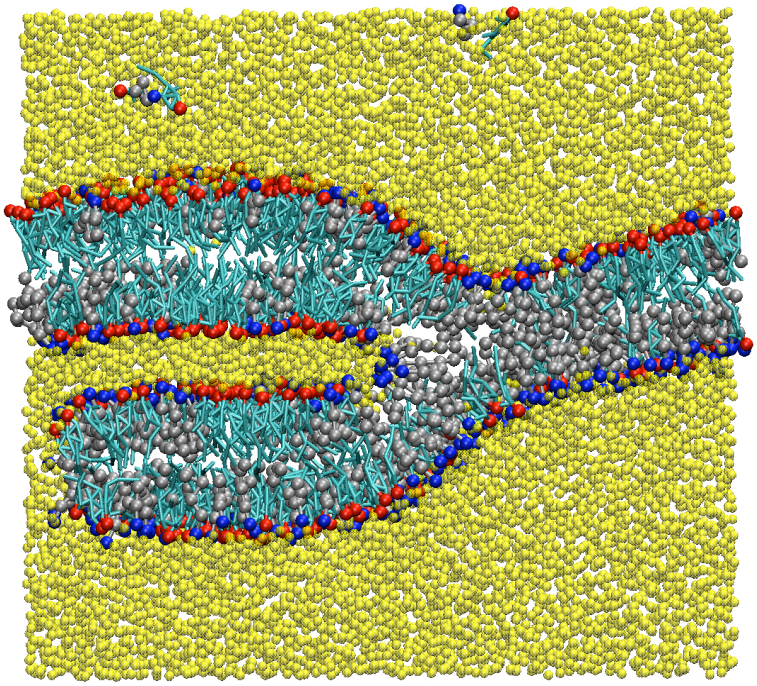


Figure 8.11 – Snapshot of self-assembly simulation where bilayer of cholesterol (cyan/red) and fatty acids (silver/blue) solvated by water (yellow) forms branches.

These results suggest the initial configuration may be biasing the form of the structure. As part of the heuristics development, it was desirable to develop a strategy for how to build an initial configuration and have it self-assemble into a bilayer independent of the initial structure. From our initial self-assembly simulations discussed in the previous section, the lipids phase separated from water and aggregated into a lipid region independent of the initial configuration, but when the force field was updated to the new force field in the simulations, bilayers formed from the aggregated lipid phase. These two facts indicate the hydrophobic interactions drive lipid aggregation in water, but the hydrophilic interactions induce structure and order into the lipid region to form a bilayer. The equilibration scheme was therefore updated to isolate these phenomena in a step-wise fashion by initially using the original force field to drive lipid aggregation and then switch to the new force field to induce structure into the aggregated lipid phase.

In Figure 8.12, a series of snapshots from the simulation of configuration (d) using the new equilibration scheme are preserved. From these figures, we can see the two lipid crystals combine and their components start to integrate into one continuous lipid phase. After a homogenous lipid phase with a constant thickness is formed (frame (b)), the hydrophilic interactions are replaced with the new interactions (discussed in section 8.3), and the lipids orientate themselves with the hydrophilic beads isolating the hydrophobic beads from the bulk water. Throughout the equilibration defects form and undulations occur as seen in Figure 8.12c. After some time, more lipids diffuse from the interior of the lipid phase and continue building the bilayer along the x-axis until the system achieves a uniform thickness (Figure 8.12d).

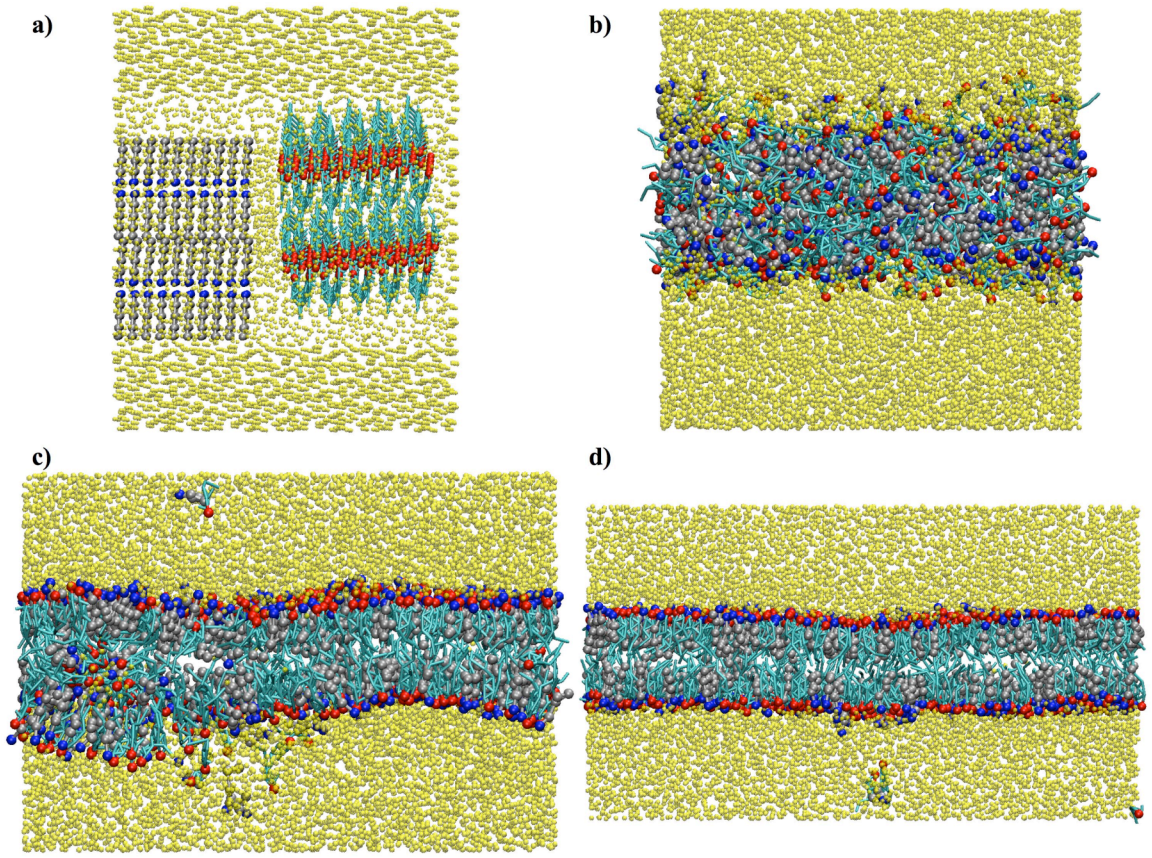


Figure 8.12 – Series of snapshot from self-assembly simulation with cholesterol (cyan/red) and fatty acids (silver/blue) solvated by water (yellow). The simulation a) starts as a crystal, b) the lipids aggregate, c) the force field is switch, and d) the defects are eliminated in the final configuration.

When 50 waters/lipid are used to solvate the configurations found in Figure 8.9, a bilayer is eventually formed in every case using the equilibration strategy discussed. However, only configuration B forms a bilayer when 35 waters/lipid are used. A similar problem as previously discussed with regard to Figure 8.10 is encountered, where the lipids span each axis to stabilize false structures as shown in Figure 8.13. The "column" spanning the y-axis never disappears throughout the simulation. As such, we conclude 50 waters/lipid is necessary for bilayer self-assembly with the CG models developed.

Although self-assembly can occur with less water, the PBC and limited volume prevent self-assembly into a bilayer.

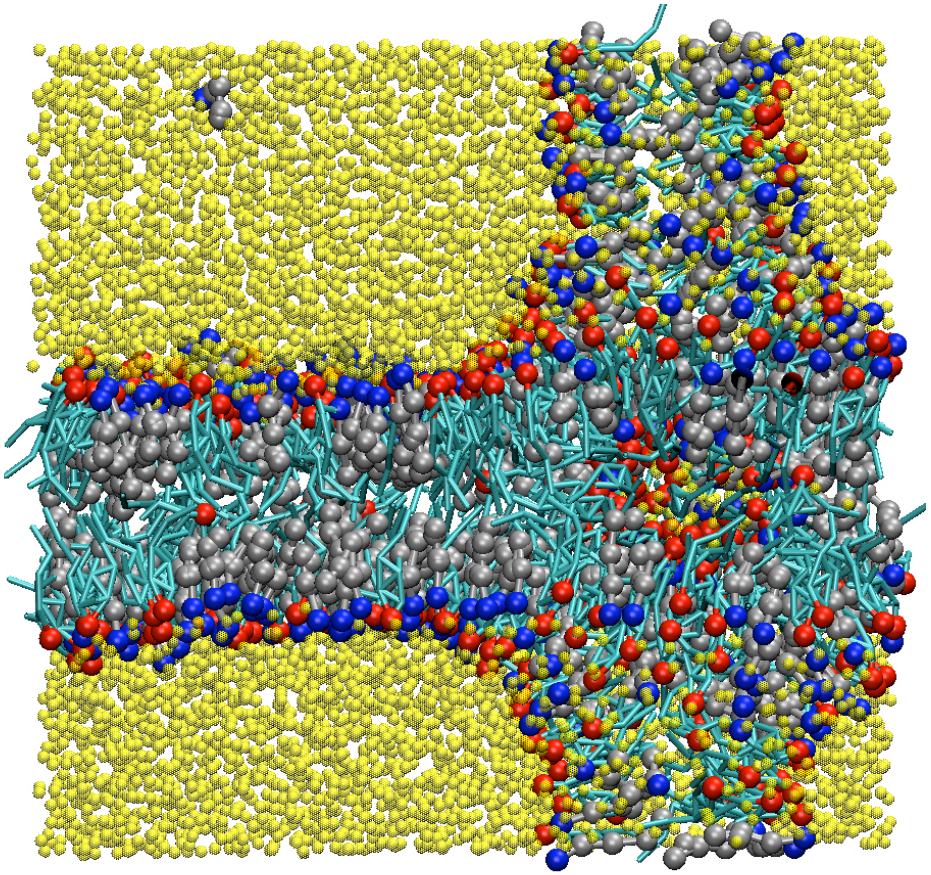


Figure 8.13 – Snapshot of self-assembly simulation where cholesterol (cyan/red) and fatty acids (silver/blue) form a bilayer, but a column bridges the PBC in the bilayer normal while surround by water (yellow).

8.5 Coarse-Grained Bilayer Results

With a self-assembled structure observed, a robust comparison between the CG model and atomistic simulations can be made. Initially, the presence of cholesterol-rich domains in the bilayer was confirmed from a snapshot of the bilayer as seen in Figure 8.14.

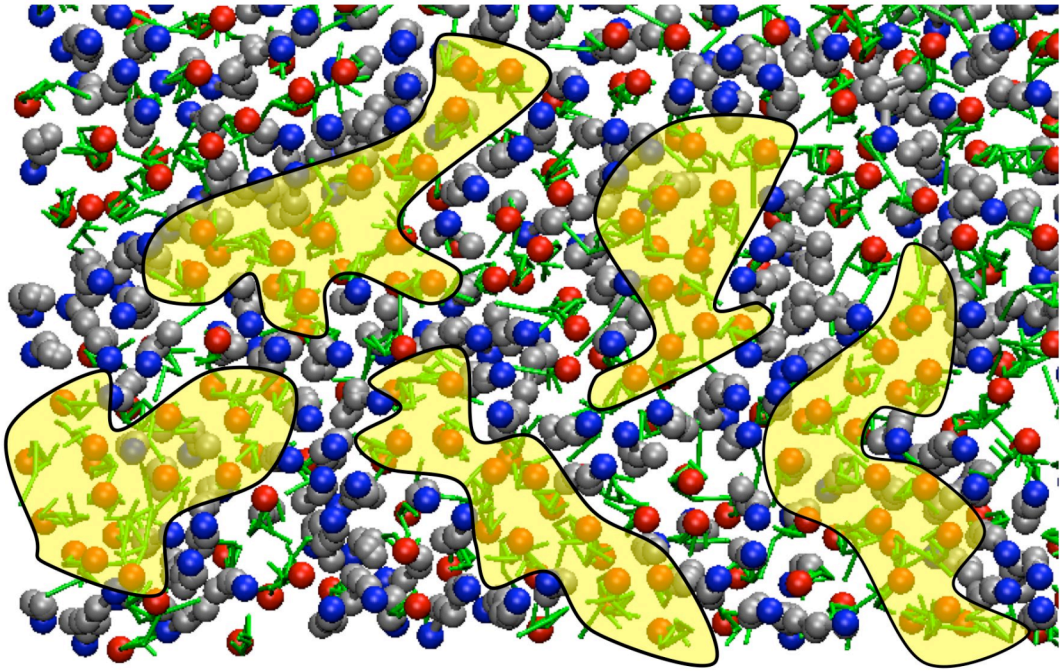


Figure 8.14 – Overhead view of CG bilayer with cholesterol rich domains highlighted (yellow regions). In the figure, the green/red molecules are CHOL and the silver/blue are FFA.

The average bilayer height was determined to be 30.0 Å and is ~6% different from the atomistic and experimental value of 32.0 Å. Although not available experimentally, the average area per lipid can be calculated in both the atomistic and the CG bilayer and are 33.1 $Å^2$ and 33.3 $Å^2$, respectively, which equates to less than a 1% difference. A CG bilayer with an accurate area and a smaller than expected thickness indicates the water might be compressing the bilayer. The average density of the bulk water phase was calculated and compared between the coarse-grained and atomistic systems. From atomistic simulations, the density is 0.984 g/cc for the bulk water phase at 333 K and is 0.987 g/cc at a temperature of 333 K in a pure water system, but from the coarse-grained simulations, the density was estimated to be 1.345 g/cc. In the CG water phase, the beads are packing more closely than in the pure water simulations. If the

diameter of the water beads was increased the water phase would be less dense and, the bilayer should be less compressed. Shifting the location of the water potential minimum to a larger distance and running the CG bilayer with the modified water potential will minimize the bilayer compression. As expected, the density of the water phase decreased and the average bilayer height increased. However, the density of the water phase had to be approximately 0.69 g/cc for the bilayer height to be approximately 32 Å. Therefore, the packing of the water must be significantly altered for the bilayer height to match the atomistic results, so the CG force field was not modified.

The long-term implications of the water compressing the bilayer involve measuring distances of the long-periodicity and short-periodicity phases in the SC lipid matrix. When attempting to validate the formation of these phases using simulations, the final coarse-grained configuration must be reverse-mapped to the atomistic level and equilibrated to verify if the distances of the repeat structures match the experimental values. Duplication of the average area per lipid indicates the lateral structure is accurate on the CG level, and this is the main source of ambiguity in determining the molecular level structure in the SC. As a result, we hypothesize the molecules will arrange themselves in the same way as they would experimentally, but the tails of the molecules will be slightly compressed. However, when reverse-mapped, the water would no longer compress the bilayer and the tails would extend to their correct length to reproduce a realistic bilayer thickness.

An objective of this work was to verify the ordering effect of cholesterol by measuring the order parameter with equation (8.1) for coarse-grained systems. The experimental values for the order parameter cannot be reported on the CG level, but the

results from CG simulations can be compared against values measured from atomistic simulations mapped to the CG level as seen in Figure 8.15. Other than the first bond, the CG values match the atomistic parameters well, and the trend of decreasing order at larger bead numbers matches the experimental trend. This trend implies the rigid structure of cholesterol induces order in the tails, and the flexible tail of cholesterol is not as effective at inducing order in the fatty acid tails towards the center of the bilayer at both the atomistic and CG level.

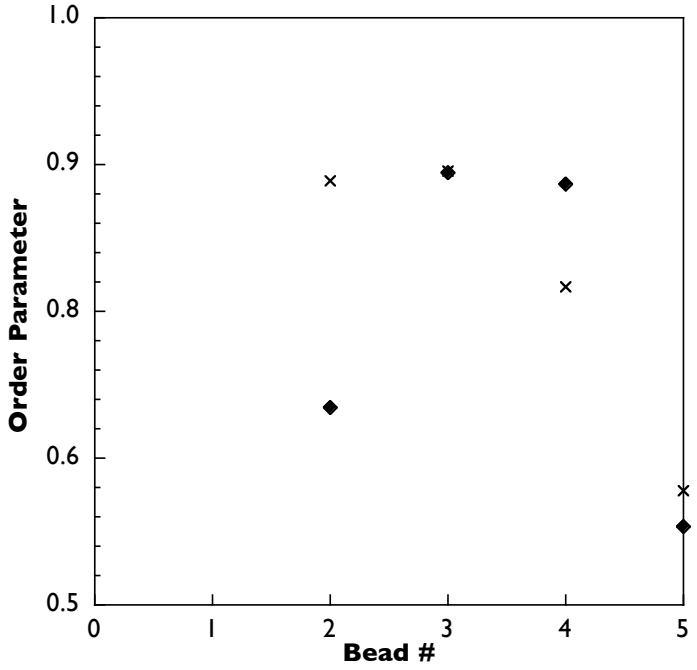


Figure 8.15 –Coarse-grained (crosses) and atomistic (diamonds) order-parameter of CG fatty acid tail.

The final comparison made was between the lateral RDF's of the atomistic bilayer and the CG bilayer. As seen in Figure 8.15, the HEAD-ALC RDF and the ALC-ALC RDF match the target very well and indicates the cholesterol is packing and arranging itself in the same manner on both length scales. The pictures from Figures 8.1 and 8.13

in conjunction with the RDF results for Figures 8.16a and 8.16b provide support for the existence of cholesterol-rich domains within this bilayer system. The HEAD-ALC RDF is slightly higher in the CG system, but is consistently higher at an average value of 0.07. This may be a consequence of the mismatch in the HEAD-HEAD RDF.

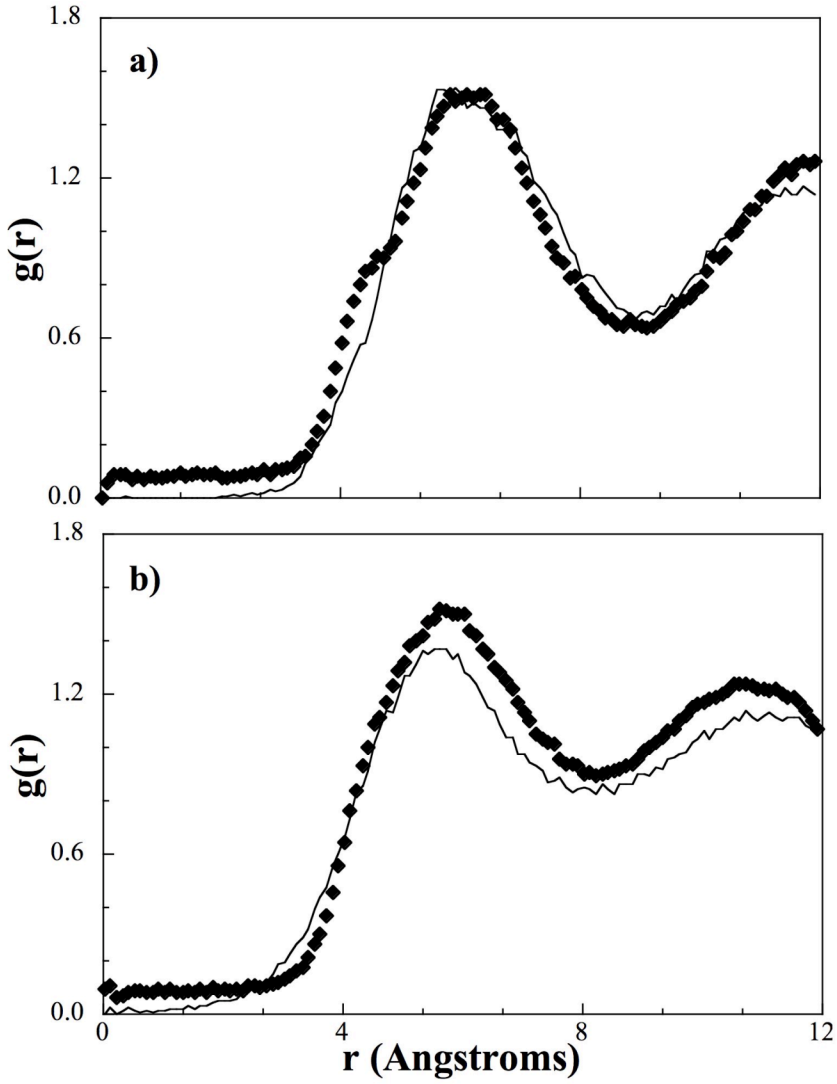


Figure 8.16 –Lateral RDF's from CG simulation (diamonds) as compared to atomistic simulations (solid line) for a) ALC-ALC interaction, b) HEAD-ALC interaction in bilayer system.

In Figure 8.17, the HEAD-HEAD RDF on the CG level has a much higher and thinner peak than the atomistic target. This indicates the CG bilayer varies little with respect to the spacing of the HEAD groups. The location of the RDF peak is reproduced in the CG bilayer supporting the conclusion that the lipids are packing correctly as seen by the match in area per lipid between the CG bilayer and the atomistic simulation. Atomistically, the acid groups hydrogen bond with water, but also hydrogen bond laterally with cholesterol and other acids. The pairs within a hydrogen bond continuously change throughout the simulation, so the distance between the atomistic acid group and a neighboring acid group varies significantly. On the CG level, the hydrogen bonding is implicit and the interaction is isotropic, so the relative positions of interacting (hydrogen bonding) HEAD beads have little variance resulting in a sharp peak in the RDF.

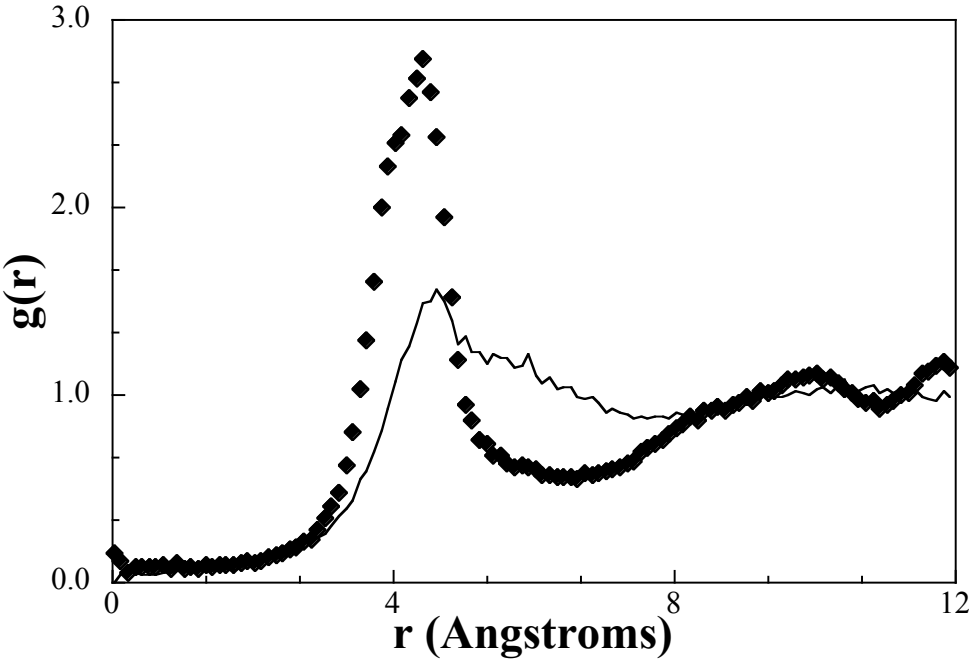


Figure 8.17 –Lateral RDF's from CG simulation (diamonds) compared to atomistic simulations (solid line) for a) ALC-ALC interaction, b) HEAD-ALC interaction in bilayer system.

8.6 Conclusions

As a first step towards developing a coarse-grained (CG) model for all the lipids of the stratum corneum (SC) that is capable of self-assembling into structures seen in native human skin, we have developed a model for cholesterol, free fatty acids, and water to be used to observe self-assembly of these lipids. When developing the cross-interactions, it was found hydrophobic cross-interactions could be taken from simple mixtures, while the hydrophilic interactions must be optimized against radial distribution functions (RDF) obtained from an atomistic bilayer system. We believe this is because the structures seen in simple mixtures are too different from those seen in a bilayer phase. If the structures and phases from simple mixtures better matched those seen in a bilayer phase, fitting RDF's from a bilayer system may not be required. For example, simple mixtures of ceramides and water form bilayers, and RDF's from those mixtures may be appropriate for more complex bilayer simulations involving ceramides. From the final force field development, we found the hydrophobic interactions drive the lipids to phase-separate from water and aggregate into a homogenous lipid phase, but the hydrophilic interactions induce structure and dictate the orientation of the lipids resulting in the formation of a bilayer.

With a valid CG model, bilayer self-assembly heuristics were developed for future simulations involving more complex bilayer systems. We found the lipids must be solvated by 50 waters per lipid to provide the appropriate working volume for the lipids. In addition, if the force field developed from just simple mixtures was used to drive lipid aggregation, and then the force field was switched to the model with correct hydrophilic

interactions, a similar bilayer structure formed independent of the initial configuration used. Within the CG bilayer simulations, the water beads compressed the bilayer, but if the simulation were reversed mapped to the atomistic level, the density of water could lower to the experimental value and the bilayer would equilibrate and reproduce an accurate bilayer height.

From simulations on the atomistic level and on the CG level, the bilayers studied duplicate the behavior and structure observed experimentally with respect to cholesterol rich domains and average bilayer height. In addition, numerical results such as lateral RDF's and the average area per lipid were very similar between the atomistic and CG simulations, further verifying the validity of the CG model developed.

The analysis within this chapter provides evidence a robust CG model can be developed for all of the SC lipids, and that model could self-assemble into structures seen experimentally. Special consideration may be required for any bead capable of hydrogen bonding, but we have developed a robust strategy and set of heuristics to observe self-assembly of SC lipids into bilayers observed experimentally. Also, the compression effect of our water model may result in small bilayer thicknesses, but can be corrected through reverse mapping to the atomistic level. The self-assembled structure could be analyzed to support experimental findings and could elucidate molecular level phenomenon not seen via experiments.

Chapter IX

CONCLUSIONS AND FUTURE WORK

Providing insight into the molecular level structure of the lipid matrix in the stratum corneum (SC) could aid in research concerned with drug delivery through the skin and in treatment strategies for diseased skin. The long term goal of this work is to develop a coarse-grained (CG) model of all of the skin lipids capable of self-assembling into a structure as is seen experimentally using molecular dynamics simulations. Although there is a lot of research to be done to achieve the long-term goals of this work, we have developed a solid foundation for continuing this research. As a major step towards achieving that goal, this work details the development of CG models for free fatty acids, cholesterol, and water.

From the work with free fatty acids, we found crystal structures could be retained on the CG level. In addition, a damping factor was added to the RPM method in order to stabilize the optimization in complex systems like crystals. Structural features like density and hydrogen bonding were well retained for the CG crystal simulations. In addition, we found the potentials developed were not only transferable but worked well in a mixture of all of the free fatty acids found within the SC.

The other major skin lipid model we developed was for cholesterol, where we developed different models to retain the unique structural features of cholesterol. Our unique mapping contained a rigid multi-ring structure with a distinct rough and smooth face on the CG level allowing the CG cholesterol to retain its crystalline structure. Like the fatty acids, the density and hydrogen bonding were retained on the CG level.

Although the model with a unique bead type for each ring was the most accurate for crystalline structures, the model with a homogenized bead potential was able to exhibit similar packing behaviors as the atomistic counterpart. Because of the fitted alcohol potential, we hypothesize the homogenous model will provide the right molecular level structure for bilayers involving skin lipids. Also, the homogenized model will provide for easier optimization of future skin lipid models due to the decreased number of bead types.

Another major contribution to the field of CG simulations is the water model developed in chapter VII. Typically, water must be mapped as one atomistic water to one bead when coarse-graining. However, by applying the k-means algorithm, an optimal clustering of water can be found from the target trajectory, allowing multiple waters to be mapped to single beads for use in any center-of-mass based CG method. From our analysis, we found mapping four waters to a single bead provided the ideal balance between speed and accuracy with respect to bulk water and solvating an amphiphilic solute. In addition, we found the water model was capable of structuring the tails of a fatty acid in a simple water/acid mixture.

With a model for free fatty acids, cholesterol, and water developed, we the optimized the cross-interactions between these components for use in a CG self-assembly simulation. For a cholesterol/free fatty acid bilayer, the cross-interactions optimized from simple mixtures were inappropriate for self-assembly. The state of the simple mixtures is too dissimilar from what is found in a bilayer. By strengthening the hydrophilic cross-interactions and weakening the affinity between water and TAIL and RING beads, respectively, the CG lipids were able to self-assemble and form a bilayer

exhibiting the same structural features as the experimental and atomistic target. In addition, general heuristics were developed for observing self-assembly of skin lipids into an experimentally observed bilayer.

The components necessary for self-assembling into the lipid matrix structure found within the SC include fatty acids, cholesterol, water, and ceramides. In order to determine the molecular level arrangement of the SC lipids in native human skin, an accurate CG model for all ceramides still needs to be developed.

The main bottleneck to developing the ceramide model is the lack of an accurate atomistic force field. Cholesterol and free fatty acids contained parameters and charge distributions within the CHARMM force field, but the force field is incomplete for a ceramide molecule. Work has been initiated for completing the force field for certain ceramides by performing quantum mechanics calculations. The bonded potential calculated seems to match the potential of similar structures, but the calculated charge distribution appears to provide for inaccurate crystal structures in the atomistic simulations.

Unfortunately, the bilayer studied in this work was in a low-order state as compared to the solid like state found within the SC of the skin. As such, it is still inconclusive if the crystal potential must be optimized for the ceramide molecules or if the crystal potentials for cholesterol and free fatty acids need to be used in place of their amorphous potential. By comparing the target tail radial distribution functions (RDF) from ceramide bilayers, the necessity of a crystalline CG model could be assessed.

Although certain potentials needed to be fitted to a bilayer system for cholesterol and fatty acids, we may be able to avoid this for ceramides. Ceramides naturally form

bilayers in the presence of water unlike simple mixtures of cholesterol and water or fatty acids and water. As such, the structure of ceramides in water is similar to what would be found in the skin, so fitting RDF's from a simple mixture of water and ceramide would provide for the proper potential for later use in mixed lipid systems. As for cross-interactions between ceramides and the other SC lipids, the sampling of the hydrogen-bonding network must be similar in the optimization simulations as what would be seen in a complex bilayer. Simulation of complex atomistic bilayers may not be feasible in every case, because of the number of lipids and water required to duplicate the experimental features.

Another possible pitfall is the compression effect of CG water on the bilayer structure. The lateral packing and arrangement of the molecules are predicted to be accurate in a CG simulation, but the bilayer thickness could be smaller than the experimental findings. However, if robust reverse-mapping techniques are employed on the final CG structure, the atomistic simulation will equilibrate and exhibit an accurate thickness for comparison to experimental results.

Although the bilayer between cholesterol and water was solvated by 50 waters/lipid, this may be too high for SC lipid bilayers with more complex compositions. Experimentally, the self-assembled structures studied contain minimal water amounts. However, the use of water to drive self-assembly may be necessary when considering the preparation schemes involved with experimental self-assembly and the possible involvement of enzymatic processes for assembly *in vivo*. These conditions cannot be reproduced using simulations, but an appropriate amount of hydration may act as an

appropriate substitute. Also, the simulation could be "dehydrated" by removing water from the configuration in the reverse-mapping procedure.

Assuming the full CG skin lipid model can be developed and is shown to be accurate, there are many directions the research can take. By duplicating the simpler skin lipid mixtures discussed in chapter II, we can validate the experimental findings and possibly provide more insight into the role of the different lipids in bilayer structure and self-assembly. In addition, the full model can be used on a mixture of all of the lipids found in the skin to observe self-assembly into the structure observed experimentally in native human skin.

Once the structure has self-assembled and stabilized, the simulation can be reverse-mapped to the atomistic level and equilibrated. With a stable atomistic lipid matrix that has the same structural features as what is seen experimentally, diffusion coefficients of drugs delivered through the skin can be measured easily, and the effectiveness of penetration enhancers can be tested to observe how they affect the structure of the skin lipids. Also, the lipid matrix can be altered to better match diseased skin to provide insight into the mechanisms, which result in depleted barrier function. With structures representing diseased skin, treatment strategies could be proposed and preliminary tests can be performed to assess the effectiveness and feasibility of the new strategies.

REFERENCES

[1] *Gray's Anatomy*, edited by P. Williams (Churchill Livingstone, New York, 1995).

[2] G. Imokawa, A. Abe, K. Jin, Y. Higaki, M. Kawashima, and A. Hidano, Journal of Investigative Dermatology **96** (4), 523 (1991); A. Yamamoto, S. Serizawa, M. Ito, and Y. Sato, Archives of Dermatological Research **283** (4), 219 (1991).

[3] S. Motta, M. Monti, S. Sesana, R. Caputo, S. Carelli, and R. Ghidoni, Biochimica Et Biophysica Acta **1182** (2), 147 (1993).

[4] S. Motta, S. Sesana, R. Ghidoni, and M. Monti, Archives of Dermatological Research **287** (7), 691 (1995).

[5] A. P. M. Lavrijsen, J. A. Bouwstra, G. S. Gooris, A. Weerheim, H. E. Bodde, and M. Ponec, Journal of Investigative Dermatology **105** (4), 619 (1995).

[6] W. M. Holleran, Y. Takagi, and Y. Uchida, Febs Letters **580** (23), 5456 (2006); P. M. Elias, M. L. Williams, W. M. Holleran, Y. J. Jiang, and M. Schmuth, Journal of Lipid Research **49** (4), 697 (2008).

[7] J. A. Segre, Journal of Clinical Investigation **116** (5), 1150 (2006); P. M. Elias, Y. Hatano, and M. L. Williams, Journal of Allergy and Clinical Immunology **121** (6), 1337 (2008).

[8] J. A. Bouwstra and M. Ponec, Biochimica Et Biophysica Acta-Biomembranes **1758** (12), 2080 (2006).

[9] X. Peng, G. Tan, L. B. Lawson, J. He, J. D. Clements, K. D. Papadopoulos, and V. T. John, presented at the 2007 AIChE Annual Meeting, Salt Lake City, Utah, 2007 (unpublished).

[10] R. Notman, M. Noro, B. O'Malley, and J. Anwar, Journal Of The American Chemical Society **128** (43), 13982 (2006); R. Notman, W. K. den Otter, M. G. Noro, W. J. Briels, and J. Anwar, Biophysical Journal **93** (6), 2056 (2007).

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

[12] C. Pare and M. Lafleur, Langmuir **17** (18), 5587 (2001).

[13] R. J. Scheuplein and R. L. Bronaugh, in *Biochemistry and Physiology of the Skin*, edited by L. A. Goldsmith (Oxford University Press, New York, 1983), Vol. II, pp. 1255.

[14] (http://encarta.msn.com/media_461516297/Structure_of_the_Skin.html, 2006).

[15] H. J. Yardley and R. Summerly, Pharmacology & Therapeutics **13** (2), 357 (1981).

[16] R. K. Freinkel and T. N. Traczyk, Journal of Investigative Dermatology **85** (4), 295 (1985); S. Grayson, A. G. Johnsonwinegar, B. U. Wintroub, R. R. Isseroff, E. H. Epstein, and P. M. Elias, Journal of Investigative Dermatology **85** (4), 289 (1985); G. K. Menon, S. Grayson, and P. M. Elias, Journal of Investigative Dermatology **86** (5), 591 (1986).

[17] M. B. Reddy, R. H. Guy, and A. L. Bunge, Pharmaceutical Research **17** (11), 1414 (2000).

[18] O. Simonetti, A. J. Hoogstraate, W. Bialik, J. A. Kempenaar, A. H. G. J. Shrijvers, H. E. Bodde, and M. Ponec, Archives of Dermatological Research **287** (5), 465 (1995).

[19] S. Mitragotri, D. A. Edwards, D. Blankschtein, and R. Langer, Journal of Pharmaceutical Sciences **84** (6), 697 (1995).

[20] J. F. Nagle and S. Tristam-Nagle, Biochimica Et Biophysica Acta-Reviews on Biomembranes **1469** (3), 159 (2000).

[21] P. S. Talreja, N. K. Kleene, W. L. Pickens, T. F. Wang, and G. B. Kasting, Aaps Pharmsci **3** (2), art. no. (2001).

[22] J. A. Bouwstra, G. S. Gooris, J. A. Vanderspek, and W. Bras, Journal of Investigative Dermatology **97** (6), 1005 (1991).

[23] J. A. Bouwstra, G. S. Gooris, W. Bras, and D. T. Downing, Journal of Lipid Research **36** (4), 685 (1995).

[24] M. W. de Jager, G. S. Gooris, I. P. Dolbnya, W. Bras, M. Ponec, and J. A. Bouwstra, Chemistry and Physics of Lipids **124** (2), 123 (2003).

[25] M. W. de Jager, G. S. Gooris, I. P. Dolbnya, W. Bras, M. Ponec, and J. A. Bouwstra, Journal of Lipid Research **45** (5), 923 (2004).

[26] D. C. Swartzendruber, A. Manganaro, K. C. Madison, M. Kremer, P. W. Wertz, and C. A. Squier, Cell and Tissue Research **279** (2), 271 (1995).

[27] P. W. Wertz, D. C. Swartzendruber, K. C. Madison, and D. T. Downing, Journal of Investigative Dermatology **89** (4), 419 (1987); S. Law, P. W. Wertz, D. C. Swartzendruber, and C. A. Squier, Archives of Oral Biology **40** (12), 1085 (1995); K. C. Madison, Journal of Investigative Dermatology **121** (2), 231 (2003).

[28] H. Ohman and A. Vahlquist, Acta Dermato-Venereologica **74** (5), 375 (1994).

[29] L. Norlen, I. Nicander, A. Lundsjo, T. Cronholm, and B. Forslind, Archives of Dermatological Research **290** (9), 508 (1998).

[30] P. Wertz and D. Downing, in *Physiology, Biochemistry, and Molecular Biology of the Skin*, edited by G. LA (Oxford University Press, Oxford, 1991).

[31] J. A. Bouwstra, G. S. Gooris, F. E. R. Dubbelaar, A. M. Weerheim, A. P. Ijzerman, and M. Ponec, Journal of Lipid Research **39** (1), 186 (1998).

[32] J. A. Bouwstra, F. E. R. Dubbelaar, G. S. Gooris, and M. Ponec, Acta Dermato-Venereologica, 23 (2000).

[33] C. Dahl and J. Dahl, in *Biology of Cholesterol*, edited by P. L. Yeagle (CRC Press, Boca Raton, FL, 1988), pp. 147.

[34] P. L. Yeagle, in *Cholesterol in Membrane Models*, edited by L. Finegold (CRC Press, Boca Raton, FL, 1993), pp. 1.

[35] P. L. Yeagle, in *Structure of Biological Membranes*, edited by P. L. Yeagle (CRC Press, Boca Raton, FL, 2005), pp. 243.

[36] K. E. Bloch, CRC Critical Reviews in Biochemistry **14** (1), 47 (1983).

[37] P. L. Yeagle, Biochimica Et Biophysica Acta **822** (3-4), 267 (1985).

[38] P. W. Wertz, Acta Dermato-Venereologica, 7 (2000).

[39] M. Ponec, A. Weerheim, P. Lankhorst, and P. Wertz, Journal of Investigative Dermatology **120** (4), 581 (2003).

[40] K. J. Robson, M. E. Stewart, S. Michelsen, N. D. Lazo, and D. T. Downing, Journal of Lipid Research **35** (11), 2060 (1994); M. E. Stewart and D. T. Downing, Journal of Lipid Research **40** (8), 1434 (1999).

[41] D. Kuempel, D. C. Swartzendruber, C. A. Squier, and P. W. Wertz, Biochimica Et Biophysica Acta-Biomembranes **1372** (1), 135 (1998).

[42] P. Wertz and L. Norlen, in *Skin, Hair, and Nails: Structure and Function*, edited by B. Forslind and M. Lindberg (M Dekker, New York, 2004).

[43] T. J. McIntosh, M. E. Stewart, and D. T. Downing, Biochemistry **35** (12), 3649 (1996).

[44] M. de Jager, G. Gooris, M. Ponec, and J. Bouwstra, Journal of Investigative Dermatology **123** (5), 911 (2004).

[45] S. Wartewig and R. H. H. Neubert, Skin Pharmacology and Physiology **20** (5), 220 (2007).

[46] D. Kessner, M. Kiselev, S. Dante, T. Hauss, P. Lersch, S. Wartewig, and R. H. H. Neubert, European Biophysics Journal with Biophysics Letters **37** (6), 989 (2008).

[47] J. A. Bouwstra, G. S. Gorris, K. Cheng, A. Weerheim, W. Bras, and M. Ponec, Journal of Lipid Research **37** (5), 999 (1996).

[48] J. A. Bouwstra, G. S. Gooris, M. A. Salomonsdevries, J. A. Vanderspek, and W. Bras, International Journal of Pharmaceutics **84** (3), 205 (1992); P. A. Cornwell, B. W. Barry, C. P. Stoddart, and J. A. Bouwstra, Journal of Pharmacy and Pharmacology **46** (12), 938 (1994).

[49] J. A. Bouwstra, J. Thewalt, G. S. Gooris, and N. Kitson, Biochemistry **36** (25), 7717 (1997).

[50] J. Zbytovska, M. A. Kiselev, S. S. Funari, V. M. Garamus, S. Wartewrig, K. Palat, and R. Neubert, Colloids and Surfaces a-Physicochemical and Engineering Aspects **328** (1-3), 90 (2008).

[51] J. A. Bouwstra, G. S. Gooris, F. E. R. Dubbelaar, and M. Ponec, Journal of Investigative Dermatology **118** (4), 606 (2002).

[52] A. Ruettinger, M. A. Kiselev, T. Hauss, S. Dante, A. M. Balagurov, and R. H. H. Neubert, European Biophysics Journal with Biophysics Letters **37** (6), 759 (2008).

[53] J. Ouimet, S. Croft, C. Pare, J. Katsaras, and M. Lafleur, Langmuir **19** (4), 1089 (2003).

[54] J. Ouimet and M. Lafleur, Langmuir **20** (18), 7474 (2004).

[55] M. W. de Jager, G. S. Gooris, I. P. Dolbnya, M. Ponec, and J. A. Bouwstra, Biochimica Et Biophysica Acta-Biomembranes **1664** (2), 132 (2004).

[56] M. W. de Jager, G. S. Gooris, M. Ponec, and J. A. Bouwstra, Journal of Lipid Research **46** (12), 2649 (2005).

[57] D. C. Swartzendruber, P. W. Wertz, D. J. Kitko, K. C. Madison, and D. T. Downing, Journal of Investigative Dermatology **92** (2), 251 (1989).

[58] B. Forslind, Acta Dermato-Venereologica **74** (1), 1 (1994).

[59] L. Norlen, Journal of Investigative Dermatology **117** (4), 830 (2001).

[60] J. R. Hill and P. W. Wertz, Biochimica Et Biophysica Acta-Biomembranes **1616** (2), 121 (2003); P. W. Wertz, Advanced Drug Delivery Reviews **18** (3), 283 (1996).

[61] B. Dahlen and I. Pascher, Chemistry and Physics of Lipids **24** (2), 119 (1979).

[62] M. A. Kiselev, N. Y. Ryabova, A. M. Balagurov, S. Dante, T. Hauss, J. Zbytovska, S. Wartewig, and R. H. H. Neubert, European Biophysics Journal with Biophysics Letters **34** (8), 1030 (2005); M. A. Kiselev, Crystallography Reports **52** (3), 525 (2007).

[63] B. Dahlen and I. Pascher, Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry **B 28** (AUG15), 2396 (1972).

[64] S. A. Pandit and H. L. Scott, Biochimica Et Biophysica Acta-Biomembranes **1788** (1), 136 (2009).

[65] G. H. Peters, F. Y. Hansen, M. S. Moller, and P. Westh, Journal of Physical Chemistry B **113** (1), 92 (2009).

[66] G. A. Khelashvili and H. L. Scott, Journal of Chemical Physics **120** (20), 9841 (2004).

[67] S. A. Pandit, D. Bostick, and M. L. Berkowitz, Biophysical Journal **86** (3), 1345 (2004).

[68] S. A. Pandit, S. W. Chiu, E. Jakobsson, A. Grama, and H. L. Scott, Langmuir **24** (13), 6858 (2008).

[69] T. Rog and M. Pasenkiewicz-Gierula, Biophysical Journal **91** (10), 3756 (2006).

[70] Z. Zhang, S. Y. Bhide, and M. L. Berkowitz, Journal of Physical Chemistry B **111** (44), 12888 (2007).

[71] P. S. Niemela, M. T. Hyvonen, and I. Vattulainen, Biophysical Journal **90** (3), 851 (2006).

[72] P. S. Coppock and J. T. Kindt, Langmuir **25** (1), 352 (2009).

[73] S. Izvekov and G. A. Voth, Journal of Physical Chemistry B **109** (7), 2469 (2005).

[74] G. Toth, Journal of Physics-Condensed Matter **19** (33) (2007).

[75] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, Journal of Chemical Physics **120** (23), 10896 (2004).

[76] A. P. Lyubartsev and A. Laaksonen, Physical Review E **52** (4), 3730 (1995).

[77] D. Reith, M. Putz, and F. Muller-Plathe, Journal of Computational Chemistry **24** (13), 1624 (2003).

[78] G. Milano, S. Goudeau, and F. Muller-Plathe, Journal of Polymer Science Part B-Polymer Physics **43** (8), 871 (2005).

[79] B. Smit, P. Hilbers, K. Esselink, L. Rupert, N. Vanos, and A. Schlijper, Nature **348** (6302), 624 (1990).

[80] S. J. Marrink, A. H. de Vries, and A. E. Mark, Journal of Physical Chemistry B **108** (2), 750 (2004).

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

[82] R. Goetz and R. Lipowsky, Journal of Chemical Physics **108** (17), 7397 (1998).

[83] C. G. Sztrum, O. Hod, and E. Rabani, Journal of Physical Chemistry B **109** (14), 6741 (2005); M. J. Stevens, Journal of Chemical Physics **121** (23), 11942 (2004); P. K. Maiti, Y. Lansac, M. A. Glaser, N. A. Clark, and Y. Rouault, Langmuir **18** (5), 1908 (2002); J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, P. B. Moore, and M. L. Klein, Journal of Physical Chemistry B **105** (40), 9785 (2001); J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, Journal of Physical Chemistry B **105** (19), 4464 (2001); B. J. Palmer and J. Liu, Langmuir **12** (3), 746 (1996).

[84] H. S. Ashbaugh, H. A. Patel, S. K. Kumar, and S. Garde, Journal of Chemical Physics **122** (10), 104908 (2005); Q. Sun and R. Faller, Computers & Chemical Engineering **29** (11-12), 2380 (2005); D. Bedrov, C. Ayyagari, and G. D. Smith, Journal of Chemical Theory and Computation **2** (3), 598 (2006); A. Rakshit and R. C. Picu, Journal of Chemical Physics **125** (16) (2006).

[85] J. Elezgaray and M. Laguerre, Computer Physics Communications **175** (4), 264 (2006).

[86] C. Peter, L. Delle Site, and K. Kremer, Soft Matter **4** (4), 859 (2008).

[87] D. Reith, H. Meyer, and F. Muller-Plathe, Macromolecules **34** (7), 2335 (2001).

[88] M. Schlenkric, J. Brickmann, A. D. MacKerrel Jr., and M. Karplus, in *Biological Membranes: A Molecular Perspective from Computation and Experiment*, edited by K. M. a. B. Roux (Birkhauser, Boston, 1996); S. E. Feller and A. D. MacKerell, Journal Of Physical Chemistry B **104** (31), 7510 (2000); S. E. Feller, K. Gawrisch, and A. D. MacKerell, Journal Of The American Chemical Society **124** (2), 318 (2002).

[89] Z. Cournia, A. C. Vaiana, G. M. Ullmann, and J. C. Smith, Pure and Applied Chemistry **76** (1), 189 (2004).

[90] Z. Cournia, J. C. Smith, and G. M. Ullmann, Journal of Computational Chemistry **26** (13), 1383 (2005).

[91] M. Neumann, Journal of Chemical Physics **82** (12), 5663 (1985).

[92] W. Smith and T. R. Forester, Journal Of Molecular Graphics **14** (3), 136 (1996).

[93] S. J. Plimpton, Journal of Compututational Physics **117** (1), 1 (1995).

[94] M. Allen and D. Tildesley, *Computer Simulation of Liquids*. (Oxford University Press, Oxford, 1989).

[95] A. Einstein, Annalen der Physik **17** (1), 549 (1905).

[96] L. J. Chen, H. J. Qian, Z. Y. Lu, Z. S. Li, and C. C. Sun, Journal of Physical Chemistry B **110** (47), 24093 (2006).

[97] E. R. Chan, A. Striolo, C. McCabe, P. T. Cummings, and S. C. Glotzer, Journal of Chemical Physics **127** (11) (2007).

[98] E. Moreno, R. Cordobilla, T. Calvet, F. J. Lahoz, and A. I. Balana, Acta Crystallographica Section C-Crystal Structure Communications **62**, O129 (2006).

[99] W. G. Hoover, Physical Review A **31** (3), 1695 (1985).

[100] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainen, Journal of Chemical Physics **121** (18), 9156 (2004).

[101] S. J. Marrink, J. Risselada, and A. E. Mark, Chemistry and Physics of Lipids **135** (2), 223 (2005); A. P. Lyubartsev, European Biophysics Journal with Biophysics Letters **35** (1), 53 (2005); M. Kranenburg, M. Venturoli, and B. Smit, Journal of Physical Chemistry B **107** (41), 11491 (2003).

[102] V. Molinero and W. A. Goddard, Journal of Physical Chemistry B **108** (4), 1414 (2004).

[103] L. Y. Hsu and C. E. Nordman, Science **220** (4597), 604 (1983).

[104] L. Y. Hsu, J. W. Kampf, and C. E. Nordman, Acta Crystallographica Section B-Structural Science **58**, 260 (2002).

[105] S. W. Chiu, E. Jakobsson, R. J. Mashl, and H. L. Scott, Biophysical Journal **83** (4), 1842 (2002).

[106] S. A. Pandit, S. Vasudevan, S. W. Chiu, R. J. Mashl, E. Jakobsson, and H. L. Scott, Biophysical Journal **87** (2), 1092 (2004).

[107] R. Goetz, G. Gompper, and R. Lipowsky, Physical Review Letters **82** (1), 221 (1999); M. Venturoli, M. M. Sperotto, M. Kranenburg, and B. Smit, Physics Reports-Review Section of Physics Letters **437** (1-2), 1 (2006).

[108] G. A. Khelashvili, S. A. Pandit, and H. L. Scott, Journal of Chemical Physics **123** (3) (2005).

[109] S. Izvekov and G. A. Voth, Journal of Chemical Theory and Computation **2** (3), 637 (2006).

[110] H. S. Sheih, L. G. Hoard, and C. E. Nordman, Acta Crystallographica Section B-Structural Science **37** (AUG), 1538 (1981).

[111] J. Polson and D. Frenkel, Journal of Chemical Physics **111** (4), 1501 (1999).

[112] K. R. Hadley and C. McCabe, To be Submitted (2009).

[113] B. Guillot, Journal of Molecular Liquids **101** (1-3), 219 (2002); J. K. Holt, Microfluidics and Nanofluidics **5** (4), 425 (2008).

[114] G. Malenkov, Journal of Physics-Condensed Matter **21** (28), 35 (2009).

[115] B. Guillot and Y. Guissani, Journal of Chemical Physics **114** (15), 6720 (2001).

[116] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones, Faraday Discussions **141**, 251 (2009).

[117] P. Mark and L. Nilsson, Journal of Physical Chemistry A **105** (43), 9954 (2001).

[118] C. C. Liew, H. Inomata, and K. Arai, Fluid Phase Equilibria **144** (1-2), 287 (1998).

[119] P. Paricaud, M. Predota, A. A. Chialvo, and P. T. Cummings, Journal of Chemical Physics **122** (24) (2005).

[120] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, Journal of Chemical Physics **79** (2), 926 (1983).

[121] M. W. Mahoney and W. L. Jorgensen, Journal of Chemical Physics **112** (20), 8910 (2000).

[122] B. Hess and N. F. A. van der Vegt, Journal of Physical Chemistry B **110** (35), 17616 (2006).

[123] L. Vlcek and I. Nezbeda, Molecular Physics **102** (5), 485 (2004).

[124] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill, Journal of the American Chemical Society **120** (13), 3166 (1998).

[125] A. Bizjak, T. Urbi, V. Vlachy, and K. A. Dill, Acta Chimica Slovenica **54** (3), 532 (2007).

[126] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331.

[127] G. S. Ayton, H. L. Tepper, D. T. Mirijanian, and G. A. Voth, Journal of Chemical Physics **120** (9), 4074 (2004); S. Izvekov, J. M. J. Swanson, and G. A. Voth, Journal of Physical Chemistry B **112** (15), 4711 (2008); V. Molinero and E. B. Moore, Journal of Physical Chemistry B **113** (13), 4008 (2009).

[128] W. Shinoda, R. DeVane, and M. L. Klein, Soft Matter **4** (12), 2454 (2008).

[129] J. B. MacQueen, presented at the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1967 (unpublished).

[130] D. Steinley, British Journal of Mathematical & Statistical Psychology **59**, 1 (2006).

[131] H. Wang, C. Junghans, and K. Kremer, European Physical Journal E **28** (2), 221 (2009).

[132] A. K. Soper, Journal of Chemical Physics **258** (2-3), 121 (2000).

[133] M. Holtje, T. Forster, B. Brandt, T. Engels, W. von Rybinski, and H. D. Holtje, Biochimica Et Biophysica Acta-Biomembranes **1511** (1), 156 (2001).

[134] J. P. Douliez, A. Leonard, and E. J. Dufourc, Biophysical Journal **68** (5), 1727 (1995).