

# A Coarse-Grained Model of Stratum Corneum Lipids: Free Fatty Acids and Ceramide NS

Timothy C. Moore,<sup>†,‡</sup> Christopher R. Iacovella,<sup>†,‡</sup> Remco Hartkamp,<sup>†,‡</sup> Annette L. Bunge,<sup>§</sup> and Clare M<sup>c</sup>Cabe<sup>\*,†,‡,||</sup>

<sup>†</sup>Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, United States

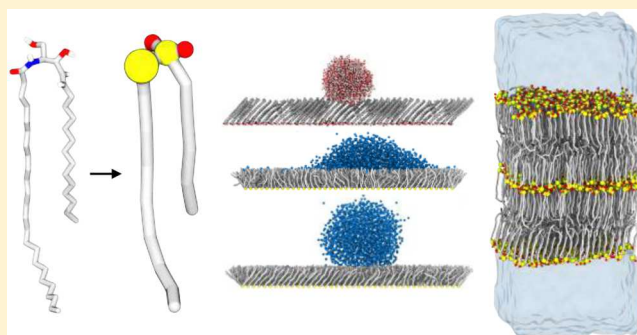
<sup>‡</sup>Vanderbilt University Multiscale Modeling and Simulation (MuMS) Center, Nashville, Tennessee 37235, United States

<sup>§</sup>Department of Chemical and Biological Engineering, Colorado School of Mines, Golden, Colorado 80401, United States

<sup>||</sup>Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States

## S Supporting Information

**ABSTRACT:** Ceramide (CER)-based biological membranes are used both experimentally and in simulations as simplified model systems of the skin barrier. Molecular dynamics studies have generally focused on simulating preassembled structures using atomistically detailed models of CERs, which limit the system sizes and time scales that can practically be probed, rendering them ineffective for studying particular phenomena, including self-assembly into bilayer and lamellar superstructures. Here, we report on the development of a coarse-grained (CG) model for CER NS, the most abundant CER in human stratum corneum. Multistate iterative Boltzmann inversion is used to derive the intermolecular pair potentials, resulting in a force field that is applicable over a range of state points and suitable for studying ceramide self-assembly. The chosen CG mapping, which includes explicit interaction sites for hydroxyl groups, captures the directional nature of hydrogen bonding and allows for accurate predictions of several key structural properties of CER NS bilayers. Simulated wetting experiments allow the hydrophobicity of CG beads to be accurately tuned to match atomistic wetting behavior, which affects the whole system, since inaccurate hydrophobic character is found to unphysically alter the lipid packing in hydrated lamellar states. We find that CER NS can self-assemble into multilamellar structures, enabling the study of lipid systems more representative of the multilamellar lipid structures present in the skin barrier. The coarse-grained force field derived herein represents an important step in using molecular dynamics to study the human skin barrier, which gives a resolution not available through experiment alone.



## I. INTRODUCTION

The stratum corneum (SC) is the outermost layer of the skin, and serves as the barrier between the internal and external environments of the body, preventing, among other things, rapid dehydration and pathogen invasion.<sup>1</sup> In addition to cholesterol and free fatty acids (FFAs), several different ceramides (CERs) comprise the lipid matrix of the SC.<sup>2</sup> To date, 14 subclasses of CER have been identified in human skin that differ in headgroup and acyl tail structure.<sup>3</sup> Of these, non-alpha-hydroxy-sphingosine ceramide (CER NS, Figure 1a) is the most abundant in healthy human skin.<sup>4</sup>

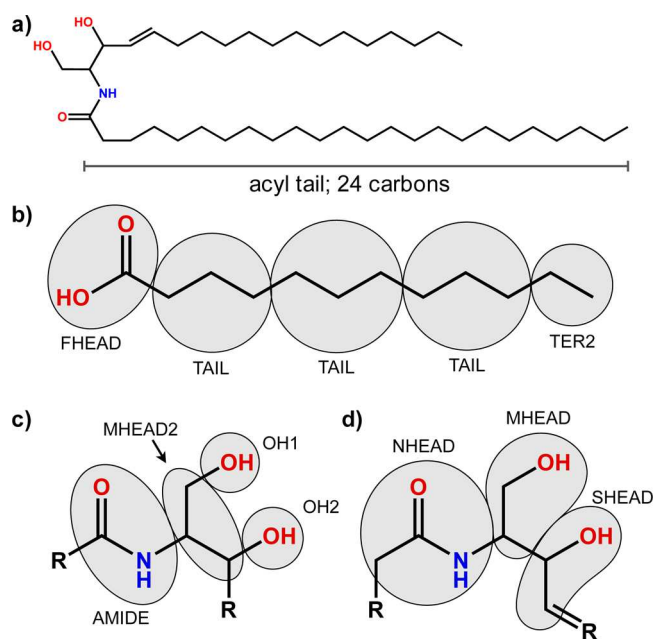
Despite knowledge of the lipid composition in human SC, the connection between lipid composition, structure, and barrier function remains unclear.<sup>5,6</sup> Understanding these connections is essential for developing effective models of skin for use in toxicology assessment and transdermal drug delivery development, as well as for developing effective treatments for skin diseases.<sup>6,7</sup> Molecular simulation studies of SC lipid systems could allow the connections between lipid

composition, structure, and barrier function to be probed, as simulation enables direct visualization of and explicit control over the molecular species in the system. For example, molecular simulations have been used extensively to better understand the structure, phase behavior, thermodynamics, and transmembrane permeability of phospholipid-based bilayer systems.<sup>8–13</sup> However, CER-based bilayers have not been as widely studied as phospholipid-based systems, primarily because of the more complex nature of the lipid organization<sup>14–16</sup> and a lack of accurate atomistic force fields, which was recently addressed.<sup>17</sup> The limited atomistic (i.e., all-atom and united-atom) simulations that have been performed of CER bilayer structures have provided some insight into the lipid interactions,<sup>17–21</sup> e.g., demonstrating strong connections between the phase transition temperature, CER headgroup

Received: August 9, 2016

Revised: August 24, 2016

Published: August 26, 2016



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

structure, and lipid–lipid hydrogen bonding.<sup>17</sup> However, the dense gel and crystalline packings found in CER-rich phases<sup>22–26</sup> significantly reduce lipid mobility as compared to the liquid-crystalline phases typical of biological phospholipid-based bilayers under physiological conditions, resulting in the need for careful equilibration protocols and long equilibration times.<sup>19,27,28</sup> As such, the time and length scales on which certain phenomena occur (e.g., phase-separation, self-assembly) are limiting for all-atom models. Furthermore, bilayer structures consisting of a small number of components are highly simplified models of SC lipid lamellae; to study mixtures relevant to the multicomponent SC that form multilamellar structures, large system sizes must be employed, further increasing the computational cost.

Coarse-grained (CG) molecular models, where groups of atoms are treated as single interaction sites (CG beads), provide an attractive alternative to atomistic models. CG models typically require several orders of magnitude less computational cost to access the same time scale as the corresponding atomistic model, due to the simplified representation of the system. This simplification results in fewer pair interactions, allows implicit treatment of long-ranged electrostatics, and generally softens the energy landscape, allowing a larger time step of integration. Combined, these properties allow CG models to access the long time scales needed to form equilibrium phases via self-assembly and probe the large system sizes needed to represent the SC lipid mixture. CG models have proven effective for the study of various lipid systems,<sup>29–31</sup> including studies that examine the self-assembly of other lipids important to the skin barrier.<sup>32</sup> Recently, Sovova et al.<sup>33</sup> proposed a CG model of CER NS, which has been the

most studied CER via all-atom molecular simulation.<sup>17–20,34</sup>

We note that this model predicts stable lamellar structures as expected of CER conformations in the SC,<sup>33</sup> and is advantageous in that it is compatible with the popular MARTINI force field; however, since it was not derived to match CER structure, the model fails to accurately capture key structural properties of pure CER NS bilayers, overestimating the area per lipid by ~20% and substantially underestimating the tilt angle as compared to experimental and all-atom simulation results (i.e., no tilt is observed with this model).<sup>17,20,33</sup> Furthermore, the CG mapping used in this model (i.e., how the atoms are grouped into CG beads) lacks directional headgroup interactions that have been shown to influence the properties of CER systems,<sup>17,35,36</sup> overall making the CG model more representative of a generic lipid rather than CER NS, and thus of limited utility for accurately probing the behavior of systems representative of the SC.

In this paper, a CG mapping and force field for CER NS are derived to match structural properties calculated from all-atom simulations. The CER NS model proposed herein provides a more explicit treatment of hydroxyl groups than in the MARTINI-compatible CG CER model of Sovova et al.,<sup>33</sup> allowing the model to capture directional headgroup interactions that exist between lipids due to hydrogen bonding.<sup>17,37–39</sup> As CERs and FFAs contain identical tails (i.e., long alkyl chains), a CG FFA model is developed in parallel, using the same tail beads for CER and FFA. The recently developed multistate iterative Boltzmann inversion (MS IBI)<sup>40</sup> scheme is used to optimize these CG force fields. While potentials derived with the original IBI scheme generally reproduce target RDFs with a high degree of accuracy, they often show limited transferability to other states,<sup>41,42</sup> due to the single state nature of the scheme. They may also demonstrate significant artifacts associated with structural correlations between beads (e.g., potential wells and barriers associated with RDF peaks and valleys), and hence can be unrepresentative of the underlying potential landscape.<sup>40</sup> The MS IBI algorithm was developed to address these issues; specifically, the multistate nature of the method provides additional constraints to the optimization, reducing structural artifacts, and yields potentials that are more generally applicable than those derived from the original (i.e., single state) IBI method,<sup>40,43</sup> as is required for simulations that span multiple state points, such as simulations of lipid self-assembly. The structural properties and self-assembly of CER systems are studied using the derived CG model and compared to all-atom and experimental measurements, demonstrating close agreement and providing validation of the model.

The remainder of the paper is organized as follows: first, we provide a description of the CG model and pair potential optimizations before discussing the results obtained for the derived potentials and the validations performed on states not used in the optimization process. We then examine the properties of preassembled CER bilayers and study the self-assembly behavior of both single bilayer (unilamellar) and stacked bilayer (multilamellar) structures. Finally, conclusions and future work are discussed.

## II. MODELS AND METHODS

Deriving CG potentials via MS IBI involves several general steps. First, all-atom simulations are performed at various state points that are chosen to emphasize the different behaviors the CG model should reproduce. For example, for CERs, isotropic

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

**Atomistic Simulations.** All-atom simulations were performed with the fully atomistic CHARMM36 force field,<sup>44</sup> employing the TIP3P water model<sup>45</sup> and CHARMM-compatible headgroup parameters for CERs, which were derived and validated in previous work.<sup>17</sup> All atomistic simulations were performed using the LAMMPS simulation engine,<sup>46,47</sup> with a van der Waals interaction cutoff of 12 Å; the PPPM solver with a real space cutoff of 12 Å was used for the calculation of the long-ranged electrostatic interactions.<sup>48</sup> A time step of 1.0 fs was used with the Nosé–Hoover thermostat in the canonical ensemble (i.e., NVT, constant number of molecules, volume, and temperature) and an additional barostat for simulations in the isothermal isobaric (i.e., NPT) ensemble.<sup>49</sup> For NPT simulations, isotropic pressure control was used for bulk fluid states, while anisotropic pressure control was used for bilayer states, allowing the aspect ratio of the box to change. All-atom simulations used for gathering target data were run for 10 ns after sufficient equilibration with the trajectory written every 100 ps. A list of all the states used is given in Table S1 of the [Supporting Information](#), and the weights given to each state are listed in Table S2 of the [Supporting Information](#); the rationale for the choice of the specific states is discussed in the [Results](#) section. Note that FFA C24:0 corresponds to a fully saturated FFA with a length of 24 carbons. CER NS C16 corresponds to CER NS with an acyl tail containing 16 carbons and is therefore of roughly equal length to the sphingosine tail. Similarly, CER NS C24 corresponds to a CER NS molecule with a longer acyl tail of 24 carbons (as in [Figure 1a](#)).

Bulk fluid simulations (both FFA and CER) were initialized by placing lipids on a  $6 \times 6 \times 6$  cubic lattice in a large simulation box to create a gas of 216 molecules. To randomize the lipids and distribute them evenly throughout the box, NVT simulations were performed for 5 ns at 1300 K. The box volume was then isotropically reduced to achieve the target density over 5 ns, followed by adjusting the temperature to that of the state point of interest. The systems were then equilibrated for 5 ns, which was found to be sufficient for the potential energy to converge, followed by 10 ns of simulation over which target data was collected; this was performed in either the NVT or NPT ensemble, depending upon the state point of interest. The mixed FFA–water systems were initialized in the same manner; these systems contained 6480 water molecules and the appropriate number of lipids for the given concentration. This same approach was also used for the CER headgroups in water, which contained 4000 water molecules and 728 CER headgroups.

Atomistic gel-phase CER NS bilayers were initialized and equilibrated in the same manner as in our previous work.<sup>17</sup> Specifically, initial configurations were generated by placing lipids on a  $10 \times 10$  square lattice to create a leaflet. The leaflet was then replicated and inverted to create a bilayer and the bilayer solvated with 50 water molecules per leaflet. A short energy minimization was performed to remove any high-energy atomic overlaps, followed by 30 ps simulation in the NVT ensemble at 305 K. The system was then simulated for 100 ns at 305 K and 1.0 atm in the anisotropic NPT ensemble and the final 10 ns used to collect target data; we note that, since large oscillations in potential energy and area per lipid were not observed over the 100 ns of simulation, 10 ns is sufficient for collecting data on a property such as the RDF. Furthermore, our previous work<sup>17</sup> has shown that the structural properties of CER NS bilayer systems simulated for 100 ns at 305 K agree well with those from systems slowly cooled from just below the order–disorder transition, providing confidence that our systems are adequately equilibrated with this procedure. Dehydrated bilayer systems were initialized from the final configuration of the hydrated bilayer simulation, and simulated for an additional 40 ns to ensure convergence of the potential energy and area per lipid, with target data collected from the final 10 ns.

**CG Simulations.** All CG simulations utilized the HOOMD-Blue simulation engine.<sup>50,51</sup> Integration of the equations of motion follows Martyna–Tobias–Klein<sup>49</sup> for both NPT and NVT simulations. All CG simulations employed a nonbonded potential cutoff of 12 Å. Since CG beads are treated as electroneutral, no treatment of long-ranged electrostatics was needed. In all cases, a time step of 10 fs was used. Unless otherwise noted, the initial configuration of the CG simulation was the final configuration from the all-atom simulation, mapped to the CG level. For bulk fluid systems, this configuration was replicated twice in each direction so that the CG bulk fluid systems contained 8 times more molecules than the corresponding atomistic systems. For lamellar states, this configuration was replicated twice in each direction in the lamellar plane so that CG lamellar systems contained 4 times more molecules than the corresponding atomistic systems.

**CG Mappings.** The CG mappings used for the lipids studied in this work are shown in [Figure 1](#). The position of each CG bead is taken to be the center of mass of the atoms represented by the bead, with the mass of the CG bead defined to be the mass of the atoms it represents. The mapping for FFA is shown in [Figure 1b](#). The carboxylic acid headgroup atoms are naturally mapped to a single CG bead (FHEAD), with the lipid tail described by “TAIL” beads that represent three methyl groups and a TER2 bead to describe the terminal two methyl groups when necessary, e.g., in FFA C12:0, as shown in [Figure 1b](#). We collectively refer to the TAIL and TER2 beads in the FFA chain as TAILS. We note that this mapping closely resembles that of Hadley and McCabe,<sup>52</sup> with the exception of using a 3:1 rather than a 4:1 mapping for the tail beads.

A four-site mapping for the CER headgroups is used, as shown in [Figure 1c](#). An “AMIDE” bead maps the amide group; an “MHEAD2” bead maps the C1, C2, and C3 carbons (and their associated hydrogens) of the sphingosine tail (note that the name “MHEAD” is reserved for the middle headgroup bead in the three-site model, discussed below); an “OH1” bead maps the hydroxyl group on C1 of the sphingosine chain; and an “OH2” bead maps the hydroxyl group on C3 of the sphingosine chain. This CG mapping treats the hydroxyl groups explicitly to

account for the strong directional interactions between the CER headgroups (i.e., hydrogen bonding). Additionally, this mapping scheme provides a simple route for extension to CER species with different headgroups, since the AMIDE and MHEAD2 beads are common to all CER subclasses. We refer to the beads in the CER headgroup collectively as CERHEADs. This mapping uses the same tail mappings as the FFAs, and as such, TAIL beads in CER and FFA are described by the same interaction potentials. Note, while optimizations use CER NS C24 (see Figure 1a), the validation studies performed also include CER NS C16.

A three-site CER headgroup mapping was also investigated, as shown in Figure 1d to demonstrate the importance of the explicit treatment of hydroxyl groups. Here the effect of the hydroxyl groups is treated implicitly within the backbone structure, rather than being explicitly modeled as separate groups, and as such is similar to the MARTINI CER mapping of Sovova et al.<sup>33</sup> The tails are treated the same as in the four-site CER headgroup model, using the same mappings and force field for the TAILS as in the FFA.

In CG simulations where water is present, a 4:1 *k*-means mapped CG water model was used following the work of Hadley and McCabe in which waters are dynamically assigned to CG beads.<sup>43,53</sup> A dynamic mapping is necessary for water because a 4:1 mapping means atoms mapped to a common bead are necessarily parts of different molecules, and these molecules will not remain associated throughout the course of a simulation due to molecular diffusion. Thus, to map atomistic water to the CG level, a new mapping is effectively defined for each frame of the atomistic trajectory, with clusters of four molecules mapped to a single bead. The location of the center of mass of these beads is then used when calculating RDFs involving CG water. Since the original version of the water model was only optimized for bulk properties, it was recently reparameterized to also capture interfacial behavior.<sup>43</sup> As detailed in Moore et al., the CG water model was optimized against TIP3P all-atom simulations (that is, the same water model used in the all-atom lipid simulations) using both bulk and interfacial properties of water as target states, with the resulting force field accurately capturing structure, density, and droplet stability at 305 K and 1 atm.<sup>43</sup>

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

The bonded interactions were derived from bond/angle distributions calculated from all-atom simulations. CG beads in a given molecule are considered bonded if they occur sequentially (i.e., if they each contain an atom that shares a bond in the all-atom model). Additionally, any three consecutively bonded CG beads interact through an angle-bending potential. No torsional or dihedral interactions were included in the models. Instead, beads separated by three or more bonds interact only through nonbonded pair potentials, whereas the nonbonded pair potential is excluded from beads separated by two or fewer bonds. As in previous work,<sup>52,54</sup> bonded interactions are determined from probability distributions following the approach of Milano et al.<sup>55</sup> Bond and angle interactions were modeled as harmonic springs, with equilibrium values and spring constants derived from the distributions sampled in the fluid-state atomistic simulations mapped to its CG equivalent. After mapping the atomistic trajectories to the CG level, a normalized Gaussian distribution

$p(r)$  was fitted to the sampled distribution for each bond/angle type. A Boltzmann inversion of a Gaussian distribution yields a harmonic potential of the form

$$V(r) = k_B T \ln(p) = K_r (r - r_0)^2$$

where  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature,  $K_r$  is the spring constant determining the stiffness of the harmonic spring,  $r$  is the separation, and  $r_0$  is the equilibrium separation. This procedure was performed for each bond/angle type at each state, and the average  $K_r$  and  $r_0$  over all states for each type was used. Values of  $K_r$  and  $r_0$  for each bond/angle type are given in the Supporting Information.

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

$$V_{i+1}(r) = V_i(r) - 1/N \sum_s \alpha_s(r) k_B T_s \ln \left[ \frac{g_s^{AA}(r)}{g_s^i(r)} \right]$$

where  $\alpha_s(r)$  is a scaling factor to prevent large updates, and to place emphasis on certain parts of the potential;  $k_B$  is the Boltzmann constant; and  $T_s$  is the temperature of state  $s$ ;  $N$  is the number of states used in the optimization. Note that this expression reduces to the original IBI algorithm for a single state point. Following Ref 40,  $\alpha_s(r)$  was chosen to be a linear function of the form  $\alpha_s(r) = \alpha_{s,0} \left( 1 - \frac{r}{r_{\text{cut}}} \right)$ , where  $\alpha_{s,0}$  is the value at  $r = 0$  and determines the maximum weight given to a particular state. This form places more emphasis at smaller separations, and ensures that the potential remains 0 at the cutoff. This allows short-range interactions to be optimized first and to see their effect on long-range behavior, before longer range updates to the potentials are made, which helps to suppress structural artifacts in the force field that may arise due to intermediate and long-range ordering in the system. We also note that, while it could be argued that making smaller updates to the potential at larger separations may lead to thermodynamic inconsistencies in the model, the potentials developed through the MS-BI approach are not only able to match the target RDFs but also satisfy the thermodynamic relationship between density and pressure (including related properties such as area per lipid). Rather than applying a pressure correction as in the original IBI method, CG models derived via MS IBI use a combination of NVT and NPT state points during the optimization. The RDFs are not likely to match if the pair potential predicts the wrong density, since the RDF is normalized by density; thus, NPT states can be used to account for the pressure, as was shown in the derivation of the CG water force field.<sup>43</sup>

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

$$f_{\text{fit}} = 1 - \frac{\int_0^{r_{\text{cut}}} dr |g^i(r) - g^{AA}(r)|}{\int_0^{r_{\text{cut}}} dr [g^i(r) + g^{AA}(r)]}$$

was used to determine stopping criteria. The optimizations were terminated when  $|f_{\text{fit},i} - f_{\text{fit},i-1}| < 0.01$  for each pair being

optimized (i.e., when the fitness function became stable for each pair in the optimization). We note that MS IBI, like the original IBI method, yields numerical force fields that are not constrained to a functional form, thus allowing increased flexibility that is not available when fitting to an analytical form. A more detailed description of the theory behind the MS IBI method can be found in Moore et al.,<sup>40</sup> and an example of using MS IBI to capture several key properties of water, including pressure–density relationships, can be found in a recent publication.<sup>43</sup>

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

Seven distinct targets are used to optimize the FFA and CER nonbonded lipid self-interactions to capture a range of behaviors, including bulk fluid states (both FFA and CER), as well as an FFA monolayer and dehydrated CER bilayers. Target data from three different concentrations of FFA in water were used to optimize the TAILS–WATER interactions, and two distinct states were used to optimize the CERHEADS–WATER interactions, including hydrated CER NS C24 bilayers and CER NS headgroups solvated in water. Optimizations in both the NVT and NPT ensembles are carried out for select targets to ensure that the force field can reproduce the correct density in constant pressure simulations at 1 atm, as previously discussed. Note that not all pair potentials were optimized at all states. Instead, pair potentials were optimized at select states to capture specific behaviors. A complete list of the states used to optimize each pair potential is given in the Supporting Information. Additionally, the weight given to each state is listed in Table S2 of the Supporting Information. The weights were chosen to place more emphasis on the lamellar states; that is, relatively higher weights were given to the monolayer and dehydrated bilayer states, since the ultimate goal of the force field is accurate modeling of lamellar structures present in stratum corneum. The relative values were chosen on the basis of heuristics developed in this work (see Table S2 and Figure S14 in the Supporting Information) and reported in a previous publication.<sup>40</sup> MS IBI calculations were carried out using the open source MSIBI Python library,<sup>56</sup> which also makes use of the MDTraj Python library.<sup>57,58</sup> The optimized nonbonded interaction potentials are freely available via <https://github.com/iModels/sc-ff>.

**Analysis.** Several structural properties were used to characterize the ordered bilayer and multilamellar systems, including the area per lipid (APL), bilayer thickness, nematic order parameter ( $S_2$ ), and tilt angle ( $\theta$ ) of the lipid tails. The hydrophobicity of a surface was characterized by calculating the

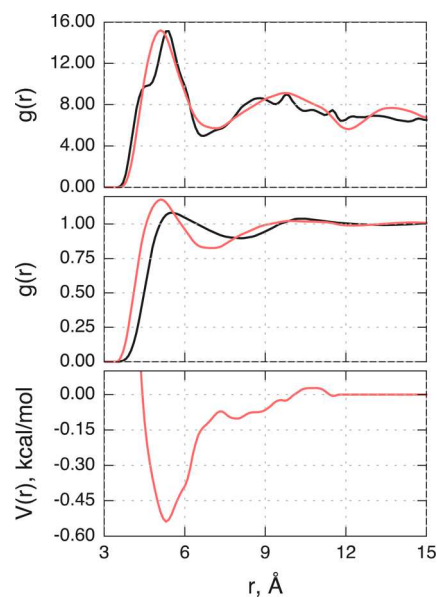
contact angle that a droplet forms with the surface. A more detailed explanation of each of these calculations is provided in the Supporting Information.

### III. RESULTS

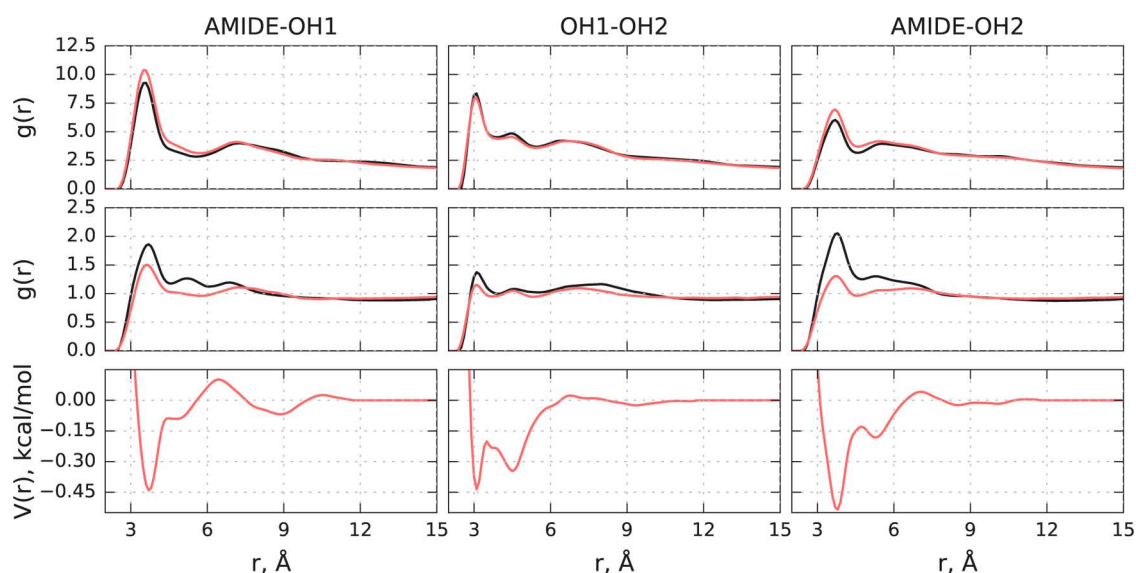
As mentioned above, the lipid–lipid self-interactions were derived first, followed by the lipid–water interactions. Even though our primary goal is to derive a CG force field for CER NS, we note that the force field for FFA, also relevant to the SC, is simultaneously derived. Both CER and FFA contain the same tail beads (i.e., TAIL and TER2); however, the headgroups of FFA molecules interact less strongly than the headgroups of the CER molecules. As such, inclusion of FFA target states alongside CERs helps to lessen the influence of the strong CER headgroup interactions when optimizing potentials involving the TAILS beads.

**III.1. Pure CER. TAIL–TAIL.** Dehydrated CER bilayers (305 K, NPT and NVT) and an FFA monolayer (NVT, 305 K) are used in the TAIL bead optimizations to ensure the force field is able to capture the behavior of ordered lamellar structures. While an ordered structure is clearly an important target for capturing the SC bilayer structure, bulk FFA fluid states in both the NVT (800 K, 0.811 g/mL) and NPT (500 K, 1.0 atm) ensembles are also used to capture the behavior of disordered states expected prior to self-assembly. The inclusion of fluid states is also found to help decrease the level of structural artifacts in the force field that arise from the intermediate and long-ranged structural correlations found in the ordered phases, and thus ensures a better quality force field.<sup>40</sup> Note that elevated temperatures were required to induce fluid phases at the density or pressure of interest. Also, due to periodic boundary conditions, the dehydrated bilayer is effectively an infinitely periodic multilamellar system.

Figure 2 shows the TAIL–TAIL RDFs for the FFA monolayer state and the bulk FFA NPT state, comparing the target and optimized RDFs for each; the optimized pair



**Figure 2.** RDFs and pair potential from the potential optimization for the interaction between TAIL beads. Top, target (black) and CG (red) RDFs from the FFA monolayer state; middle, target (black) and CG (red) RDFs from the FFA bulk fluid NPT state (500 K, 1 atm); bottom, the final potential used to calculate the CG RDFs above.



**Figure 3.** RDFs and the final derived potential between various CERHEAD–CERHEAD beads. Each column corresponds to a particular pair interaction, given at the top of the column. Top row: target (black) and CG (red) RDFs for each pair in the dehydrated NPT bilayer state; middle row, target (black) and CG (red) RDFs for each pair in the bulk NPT fluid pure CER NS state (500 K, 1 atm); bottom row, final derived potential used to calculate the CG RDFs in that column.

potential is also shown. Considering the monolayer state, the CG model captures the main RDF peak locations and heights, although some of the smaller scale features are missed, likely due to the spherical symmetry of the CG potential. The in-plane density of this monolayer system closely compares with the density of the chains commonly found for CERs in a bilayer arrangement. Thus, this system provides a close approximation to the tail ordering found in the CER NS system. Note that this RDF does not decay to unity as the monolayer only fills a small region of the box, resulting in a system density that is lower than the local monolayer density, scaling the value of the RDF; however, this does not impact the optimization as the CG RDFs are normalized by the same factor. Satisfactory agreement is found for the bulk NPT state, where the CG model correctly predicts the presence of a fluid phase and accurately predicts the density at 1 atm, with  $\rho_{CG} = 0.68(5)$  g/mL as compared to  $\rho_{AA} = 0.69(1)$  g/mL. However, we note that the peaks in the CG RDF appear to be shifted to smaller separations by approximately  $0.5 \text{ \AA}$ . As the main focus of the force field development is to accurately represent lamellar structures of the lipids in the SC, we consider such small deviations in the fluid state acceptable. We emphasize that, via MS IBI, a single CG pair potential is optimized that is capable of modeling both ordered and fluid-like states; prior efforts to use the standard IBI methodology to generate CG FFA force fields required different force fields to capture different structural configurations, even though the CG mappings did not change.<sup>52</sup>

As further evaluation of the TAIL–TAIL interaction, structural properties of the CG FFA monolayer were also compared to results from atomistic simulations, since a matching TAIL–TAIL RDF does not guarantee that characteristic monolayer properties will also match. Well-ordered monolayers were observed in simulations of each model, with nematic order parameters  $S_{2,CG} = 0.97$  and  $S_{2,AA} = 0.90$  and tilt angles  $\theta_{CG} = 40(5)^\circ$  and  $\theta_{AA} = 30(10)^\circ$ . We note that the CG force field predicts a slightly increased level of ordering and increased tilt, although the differences in tilt angle are fully

captured by the standard error of the two measurements, thus providing further validation of the pair potentials between TAILS beads.

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

In the atomistic dehydrated CER NS bilayer, in-plane hydrogen bonding is most prevalent between atoms in the groups AMIDE–OH1, OH1–OH2, and AMIDE–OH2 (see the Supporting Information). The RDFs for these pairs are shown in Figure 3. The target AMIDE–OH1 RDF from the dehydrated bilayer state, shown in Figure 3, shows a high, sharp peak at approximately  $3.5 \text{ \AA}$  as a result of the significant hydrogen bonding capacity between atoms in these groups. The CG potential captures this peak and the weaker second peak in height and separation, indicating that the headgroup packing is correctly captured with the CG model. The fluid-phase CG AMIDE–OH1 RDF also agrees well with the target RDF, capturing the location of the first peak and the longer-ranged portion of the RDF ( $r > 7.5 \text{ \AA}$ ) but missing some features in the range  $4.5 \text{ \AA} < r < 7.5 \text{ \AA}$ . Since the primary goal of this CG force field is an accurate structural representation of lamellar structures relevant to the SC lipid matrix, a higher weight was given to the dehydrated bilayer state, and thus again, this small deviation in fluid-phase behavior is considered acceptable.

As shown in Figure 3, the atomistic and CG OH1–OH2 dehydrated bilayer RDFs agree very well, with the three peaks captured in both height and location in the CG simulation. This result indicates that the CG model correctly reproduces the in-

plane ordering of the CER headgroups in the dehydrated bilayer. The sharp first peak in the dehydrated bilayer target RDF is a result of the large number of hydrogen bonds that form between atoms in these groups, which we note is reproduced by the CG model. The CG force field also accurately reproduces the fluid RDF, which is expected given the similarity between the fluid and dehydrated bilayer OH1–OH2 RDFs. Similarly, the target and CG dehydrated bilayer AMIDE–OH2 RDFs agree with a high level of accuracy, as shown in Figure 3. Several hydrogen bonds are also formed between atoms in these groups, which is manifested in the high peak in the dehydrated bilayer RDF at 3.5 Å. This peak is accurately captured in the CG simulation, again further indicating the in-plane packing of the CER headgroups in the dehydrated bilayer RDF is captured in the CG model. The CG model captures the general features of the bulk fluid phase, although quantitative agreement is not seen for the peak height; again, since our primary focus is the ordered phase, this deviation is considered to be acceptable.

In general, a good fit is found for all remaining RDFs at all states (see Figures S1–S7 of the Supporting Information), although we note that interactions involving the AMIDE bead tend not to reproduce the first peak in the fluid phase RDFs as accurately as seen in the other interactions. Whereas the other CERHEADs beads contain one or fewer polar groups, the AMIDE bead contains two polar groups, resulting in a high degree of shape anisotropy in the underlying interactions that is not easily accounted for by a spherically symmetric potential. Although this problem is not unique to MS IBI,<sup>42,59</sup> it becomes more evident when using a multistate coarse-graining methodology, in which the derived potential must find the best compromise between the fluid structure at the different target states, as dictated by the weight given to each target state. Although the weights clearly have an influence on the fit, when an additional optimization was performed in which higher weights were given to the bulk fluid states (listed as  $\alpha_{s,0}$  (high bulk) in Table S2), as expected the fluid phase CG RDFs containing an AMIDE bead showed good agreement with the atomistic RDFs, at the expense of accurate fits in the bilayer states (see Figure S14). Again, since our primary focus is quantitative agreement in the ordered phase, higher weight was given to those state points as compared to the bulk fluid, and thus, the small quantitative deviations in the bulk fluid properties are considered acceptable. Additionally, we note that the density of the pure CER NS bulk fluid NPT state calculated with the CG model, 0.69(3) g/mL, only deviates from the atomistic value of 0.71(1) g/mL by 4%.

Compared to the four-site model, relatively poor fits are observed for the ordered lamellar phases, as the CG models do not appear to capture the long-range in-plane structure, although we note bulk fluid properties are captured with similar accuracy to the four-site model. RDFs and potentials for the three-site CER NS model are provided in Figures S12 and S13 of the Supporting Information.

**Pure CER Validation.** As validation of the CG CER interactions beyond RDFs, the structural properties of dehydrated CER NS C24 bilayers simulated using the CG model are compared to those calculated from simulations of the all-atom (AA) model. Table 1 summarizes the properties of the dry bilayer calculated with each model in the anisotropic NPT ensemble. The use of this ensemble allows the lengths of the orthogonal simulation box to change independently, allowing

**Table 1. Structural Properties of Dehydrated CER NS C24 Bilayers Simulated via CHARMM36-CER All-Atom (AA) and the CG Models Derived Here<sup>a</sup>**

model	APL (Å <sup>2</sup> )	S <sub>2</sub>	thickness (Å)	θ (deg)
AA	40.0(1)	0.983	57.1(1)	12(5)
four-site CG	41.90(7)	0.976	55.44(7)	7(1)
three-site CG	48.99(7)	0.956	58.31(8)	9.5(5)

<sup>a</sup>The four-site model corresponds to Figure 1c, while the three-site model corresponds to Figure 1d. The numbers in parentheses indicate the variation in the last digit based on the standard deviation (e.g., 40.0(1) is 40.0 with a standard deviation of 0.1).

for a less biased comparison of the different models, as systems start from preassembled bilayer configurations.

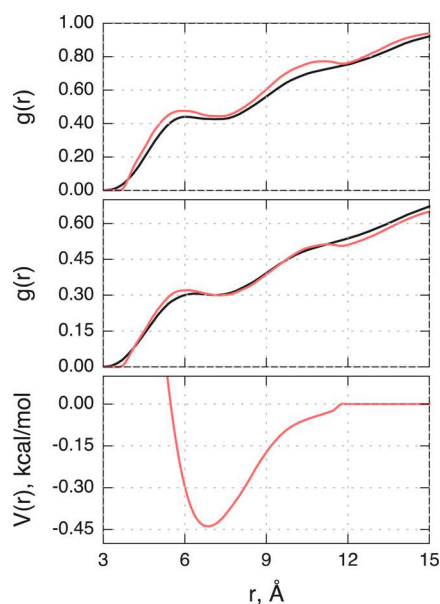
Overall, the CG force field provides accurate predictions of the key structural features of the dehydrated CER bilayer. Only minor variations in APL and S<sub>2</sub> are found between the CG and atomistic simulations, with the CG simulations producing a slightly more compact bilayer (i.e., smaller thickness) than the atomistic simulations. The tilt angle is slightly underpredicted in the CG simulations but is within the error of the atomistic system measurement. When comparing the volume per lipid (VPL) of the bilayers, only minor variations of the ratio of the CG to atomistic results is observed, with  $VPL_{CG}/VPL_{AA} = 1.02$ . As such, we can conclude that the CG CER force field accurately captures the structure and density of the dehydrated bilayer phase of CER NS. Table 1 also shows the properties of the dehydrated bilayer simulations of the three-site headgroup model depicted in Figure 1d. Note that the TAIL and TER2 beads are the same for both mappings and identical targets were used for optimizing both the three- and four-site mappings. Thus, differences in properties are a direct result of the chosen mapping. The simulation of the three-site model shows a significant overprediction of the APL as compared to the atomistic simulation. This inaccuracy is likely related to the lack of in-plane ordering in the three-site model, which is apparent in the dry bilayer RDFs, as shown in the Supporting Information (see Figures S12 and S13). The VPL of this model is also overpredicted compared to the atomistic simulations, with  $VPL_{CG}/VPL_{AA} = 1.25$ . We note that the APL for the three-site CG CER model is comparable to that obtained from the three-site MARTINI-compatible model of Sovova et al.,<sup>33</sup> although the lipids in our three-site CG model show appreciable tilt, which is not captured by the MARTINI model. As such, it is clear that the three-site CER headgroup model is an improvement over the MARTINI model, but that the four-site CER headgroup mapping better captures the in-plane packing of the CER headgroups, enabled by the more explicit description of the hydroxyl groups; thus, this treatment is essential to accurately capture the properties of CER NS. This is also observed for the hydrated bilayer systems, as discussed below. Furthermore, when a higher weight is given to the bulk fluid states in the potential optimization, we find that the APL of the dehydrated bilayer increases significantly to a value of 49.8(2) Å<sup>2</sup>. This is a result of capturing more of the bulk fluid behavior than the lamellar behavior in the potentials during the optimization. Thus, our chosen weights (i.e., higher weights given to the lamellar states) appear to provide a good balance between accurately capturing the structure of the ordered lamellar phases and accurate prediction of the bulk density, even though some of the fluid phase RDFs are slightly shifted compared to the atomistic target data.

**III.2. Optimization of Lipid–Water Interactions.** Here, the results of the lipid–water potential derivations are presented, focusing on beads within the CER molecule. Interactions are first derived using MS IBI. However, we note that system density alone can drive lipid–water structural correlations, even in systems that are hydrophobic; the MS IBI routine (and IBI in general) cannot necessarily differentiate between a correlation that is induced by density and one induced by an underlying attraction. Thus, even though MS IBI produces a potential that captures the structural features of the target RDFs, the resulting potential may not provide adequate hydrophobic behavior. To address this issue, MS IBI is coupled with surface wetting simulations, which are used to validate and, if necessary, tune specific interactions such that the atomistic wetting behavior is observed in the CG model. In this approach, targeted modifications are made to the CG interactions that were derived via MS IBI and that reproduce the correct excluded volume and RDFs, such that the appropriate hydrophobic/hydrophilic character is captured in the surface wetting simulations.

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

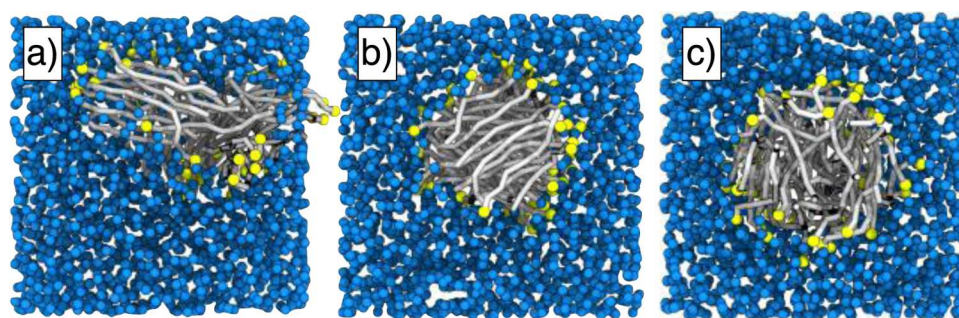
Figure 4 shows the RDFs for the most dilute and concentrated FFA–water targets, along with the optimized CG potential. Note that the highest concentration state corresponds to the system that starts from a dispersed state at each stage in the CG optimization. From the figure, we can see that both target RDFs are accurately reproduced by the CG pair potential. It is also apparent that the hydrophobic interactions are sufficient to drive phase-separation, given that the system starts dispersed and visually the FFA phase-separates from water, as shown in Figure 5a.

To ensure the interactions are sufficiently hydrophobic, a wetting simulation was performed in which a FFA monolayer was initialized with a water droplet on the tail side of the leaflet (as opposed to the headgroup side). The droplet forms a spherical cap on the surface with a characteristic contact angle. Snapshots of the droplet from atomistic and CG simulations are shown in Figure 6a and b, respectively. From the figure, we can see that the atomistic model shows formation of a droplet with little surface wetting and thus a large contact angle  $\varphi_{AA} =$

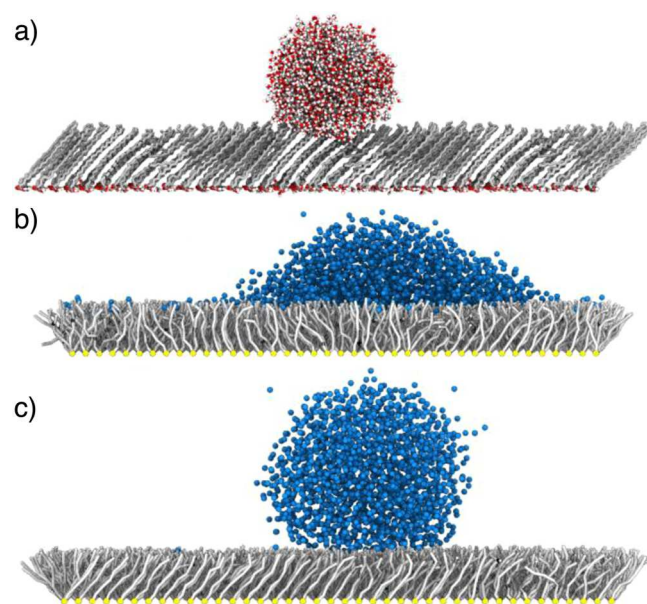


**Figure 4.** TAIL–WATER RDFs, target (black) and CG (red), from the most dilute (top) and most concentrated (middle) states. Bottom: final derived potential, before tuning to match the wetting behavior.

$133^\circ$ , while the contact angle calculated using the MS IBI-derived TAILS–WATER potentials is  $\varphi_{CG} = 40^\circ$ . We note that the WATER–WATER CG potential was derived to capture the stability of a water droplet and provides a close approximation of the atomistic surface tension of a water droplet in a (near) vacuum.<sup>43</sup> As such, the deviation from atomistic behavior lies in the TAILS–WATER interactions, indicating the derived CG force field contains insufficient hydrophobic character, despite the ability to phase-separate, as discussed above. To adjust the hydrophobic character of the TAILS beads, the MS IBI derived potentials were systematically scaled, with a factor of 0.25 found to provide good agreement with a contact angle of  $\varphi_{CG'} = 135^\circ$  (where the prime indicates use of the scaled potentials), as shown in Figure 6c. Examining the behavior of the hydrated monolayer, the atomistic simulations demonstrate a high level of nematic ordering with  $S_{2,AA} = 0.950$  at the droplet contact (where a lipid is considered in contact with the droplet if the terminal bead is within 7 Å of a water bead in the droplet). The unscaled CG potential optimized via MS IBI, which does not capture the contact angle, appears to show a disruption in ordering near the droplet with  $S_{2,CG} = 0.622$ , as compared to  $S_{2,CG'} = 0.921$  with the scaled potential. This result suggests that the lipid hydrophobicity not only affects the lipid–water interface but also plays a role in the structure of the lipid phase, which is relevant to hydrated lipid systems. Simulating the bulk FFA–WATER system with the scaled potentials results in behavior that also better matches the atomistic model, as shown in Figure 5b and c, where the simulation using the scaled CG potentials produces a cylindrical FFA aggregate in better agreement with the atomistic model, rather than an elliptical aggregate produced with the unscaled CG potential. By scaling the TAILS–WATER potentials, the CG force field is able to phase-separate, reproduce the atomistic wetting behavior, and maintain the lipid order upon wetting. Therefore, the derived TAIL–WATER and TER2–WATER potentials scaled by a factor of 0.25 are transferred to the CER–water optimizations and used henceforth.



**Figure 5.** Snapshots of FFA–water simulations using different force fields. (a) CG simulation using the derived force field before tuning the interactions. (b) CG system with potentials after scaling the FHEAD–WATER and TAILS–WATER interactions. (c) Atomistic system mapped to the CG level. CG beads colored as follows: FFA chains, silver with yellow FHEAD beads; water, blue spheres.



**Figure 6.** Snapshots from the droplet simulations of (a) the atomistic model, (b) the CG model before tuning TAILS–WATER potentials, and (c) the CG model with the TAIL–WATER potentials scaled by 0.25. Atoms are colored as follows: C, silver; O, red; H, white; methyl hydrogens are not shown for clarity. CG beads are colored as follows: TAILS, silver; FHEAD, yellow; WATER, blue.

**CERHEADS–WATER.** While hydrophobicity-driven phase-separation is the first step in lipid self-assembly, hydrophilic interactions help drive the formation of organized structures.<sup>32,60</sup> Therefore, it is important that the lipid headgroups possess the correct hydrophilic character to drive self-assembly. The CER–WATER MS IBI optimization used three states to optimize the interactions between water and the CER

headgroup beads: hydrated CER NS C24 bilayers (305 K, NVT and NPT) to capture the interfacial behavior and CER NS headgroups dispersed in water (1.0 atm, 400 K) to capture the solvation of the lipid headgroups necessary for bilayer self-assembly. We note that the headgroups in water were simulated at 400 K to ensure that the headgroups were well dispersed throughout the water. Close matches to the target RDFs are found, as shown in Figures S8–S11 of the [Supporting Information](#) along with the derived potentials. To determine if the CERHEADs beads have sufficient hydrophilic character, we examine the wetting properties of a CER–WATER interface. Atomistic wetting simulations of CER headgroups show complete wetting of the headgroups by water, viz.,  $\theta_{AA} \rightarrow 0^\circ$ , indicating a hydrophilic surface. This is expected, given the available hydrogen bonding sites in the CER headgroups. Wetting simulations using the derived CG force field show the same behavior, complete wetting of the surface, indicating the MS IBI optimizations captured the hydrophilic nature of the CER headgroups. As such, the CERHEADs–WATER potentials were not modified from their derived form. We note that the same procedure was performed for the three-site CER headgroup mapping model; sufficient hydrophilicity of these headgroup beads was observed and hence no scaling was performed.

**III.3. Validation.** As validation of the fully derived CG model (i.e., where all interactions are represented), we simulate hydrated preassembled single-component bilayers composed of CER NS C16 and CER NS C24 with the derived CG model and compare to results obtained from simulations of atomistic models and to experimental data, where available. [Table 2](#) summarizes the comparison between the CG and atomistic CHARMM–CER simulation results, where overall good agreement is seen, with the APL, nematic order, and bilayer thickness accurately reproduced. While neither model shows a strong correlation between CER length and tilt angle,

**Table 2. Structural Properties of CER NS C16/CER NS C24 Bilayers at 305 K and 1.0 atm Simulated Using the CG Models Derived Here and the CHARMM36–CER All-Atom Model (AA)<sup>a</sup>**

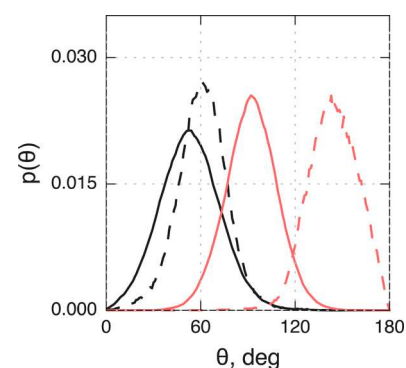
system	model	APL ( $\text{\AA}^2$ )	$S_2$	$d$ ( $\text{\AA}$ )	$\theta$ (deg)
CER NS C16	AA	42.4(2)	0.998	42.5(5)	24.3(9)
	CG	42.0(1)	0.982	44.51(9)	6.3(7)
CER NS C24	AA	39.9(2)	0.979	56.2(4)	22(4)
	CG	41.99(9)	0.981	56.6(1)	7(1)
	three-site CG	46.2(1)	0.914	61.5(2)	9(2)

<sup>a</sup>For CER NS C24, the terminal eight carbons of the acyl tail are excluded from the moment of inertia of the tail, so as to exclude the unequal portion of the tails from the calculation. The numbers in parentheses are the variation in the final reported digit, based on the standard deviation.

appreciably more tilt is observed in the atomistic systems than the CG systems. While this and other differences between the atomistic and CG simulations in Table 2 can be attributed, at least in part, to the loss of details in the CG model, the softened free energy landscape associated with the CG models should allow the systems to relax more efficiently than atomistic models, even over the same time scale. Furthermore, as we discuss in the context of self-assembly below, small deviations in lipid packing in the CG simulations result in measurable differences in tilt angle, which may also account for the perceived underprediction of tilt angles in the CG simulations. Note that only CER NS C24 was used in the force field derivation; therefore, these results also illustrate the predictive power of the CG model in the ability to accurately capture CER NS C16 behavior. Additionally, comparisons are made for CER NS C24 simulated with the three-site CG model derived here (see Table 2), where, again, the APL is overpredicted as compared to the atomistic simulation and is similar to the APL calculated with the MARTINI CER model ( $46 \text{ \AA}^2$ ).<sup>33</sup>

As an additional validation, the properties calculated from simulations of the CG and atomistic models for pure systems in Table 2 compare favorably with the limited experimental data on CER-based systems. APL values calculated from experimental studies of CER monolayers range from  $37.8$  to  $45 \text{ \AA}^2$  for CER NS C16<sup>61,62</sup> and  $39.0 \text{ \AA}^2$  for CER NS C18.<sup>38</sup> Monolayers displaying tilt angles ranging from  $0$  to  $14^\circ$  for pure CER NS C16 and CER NS C18 have been reported from experimental measurements,<sup>38,62</sup> which we note are closer to the values predicted by the CG models than the atomistic model. Other simulation studies using different atomistic force fields also fall within similar ranges; e.g., using the united atom GROMOS-based force field of Berger,<sup>63</sup> Notman et al. calculated an APL of  $37.4 \text{ \AA}^2$  for CER NS C24 at  $323 \text{ K}$ ,<sup>20</sup> while a value of  $39.8 \text{ \AA}^2$  was reported for pure CER NS C16 at  $305 \text{ K}$  with the same model in other work.<sup>17</sup> Additionally, Metcalf and Pandit calculated the APL of a pure CER NS C16 bilayer to be  $\sim 44 \text{ \AA}^2$  at  $323 \text{ K}$  using a different united atom model.<sup>64</sup> Tilt angles ranging from  $17^\circ$  for pure CER NS C24 at  $323 \text{ K}$ <sup>20</sup> to  $26^\circ$  for pure CER NS C16 at  $323 \text{ K}$ <sup>64</sup> have been reported from simulation studies. We note that the tilt angle in CG simulations significantly deviates from the atomistic simulations; this may be related to the loss of detail in the CG model, in which the local packing of the tails is altered compared to the atomistic model, resulting in different tilt behavior. However, as noted above, the tilt angles calculated in the CG simulations are actually closer to experimental values than the tilt angles from the atomistic simulations. Thus, overall, the four-site CER headgroup CG model developed here produces results for preassembled bilayers consistent with both experiment and various atomistic-level models.

In addition to the aforementioned structural metrics, we also compare the orientation of the OH groups between the CG and atomistic models for the CER NS C24 systems because of the importance of hydrogen bonding network formation in CER-based systems.<sup>37–39</sup> The orientation of the MHEAD2–OH1 vector relative to the outward bilayer normal (i.e., pointing toward the water) in the pure CER NS C24 bilayer, shown in Figure 7, is preferentially aligned approximately  $60^\circ$  off the bilayer normal in the atomistic simulation, suggesting the OH1 group interacts with the water phase. This behavior is captured with the CG model, where the MHEAD2–OH1 vector prefers an approximately  $50^\circ$  angle with respect to the bilayer normal. In contrast, the MHEAD2–OH2 vector prefers



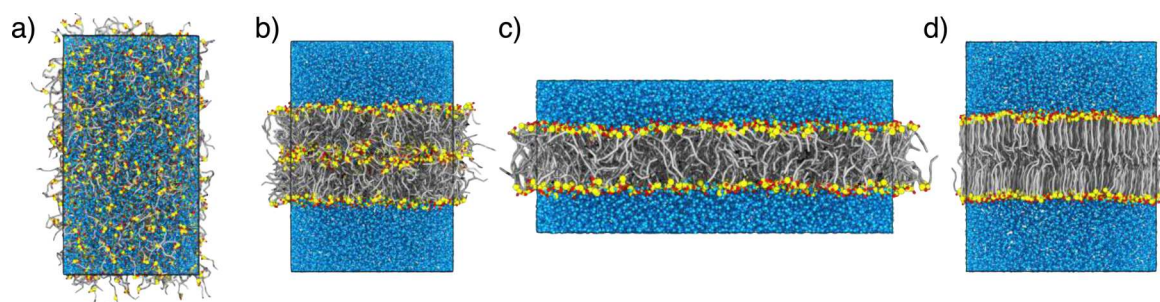
**Figure 7.** Orientation of the MHEAD2–OH vectors with respect to the bilayer normal in the CER NS C24 bilayer. A value of  $0^\circ$  indicates the OH bead lies directly between the MHEAD2 bead and the lipid–water interface; a value of  $90^\circ$  means the OH group is lying flat in the bilayer plane. Black lines, OH1; red lines, OH2; solid lines, CG; dashed lines, atomistic mapped to the CG level.

an orientation with the OH2 bead below the MHEAD2 bead (i.e., not between the MHEAD2 and WATER beads), indicating it is more involved in the hydrogen bonding network formed between the CER headgroups compared to the OH1 bead. This behavior is qualitatively captured by the CG model, where the MHEAD–OH2 vector tends to point away from the lipid–water interface, indicated by the MHEAD2–OH2 angle that is  $>90^\circ$  in Figure 7. However, there is a quantitative discrepancy between this property in the CG and atomistic simulations, where the CG system prefers an angle of approximately  $95^\circ$ , compared to  $135^\circ$  in the atomistic system. This discrepancy does not appear to have any significant impact on the APL or in-plane structure of the lipids, and is likely a consequence of the loss of detail inherent to CG models. This comparison further suggests that the CG model, at least on a qualitative level, captures the hydrogen bond network formation between the CER headgroups.

#### IV. CER NS C24 SELF-ASSEMBLY

It is unlikely that preassembled atomistic systems would be able to transition from metastable states to lower energy states because of the time scales on which major morphological transitions occur compared to the time scales that are accessible with atomistic models. Thus, pre- and self-assembled structures studied via atomistic simulations are likely to be “stuck” in metastable configurations. For example, Das et al. studied SC lipid self-assembly using a united-atom force field, finding that the lipids self-assemble into a lamellar phase under confinement but form micellar structures in the bulk,<sup>28</sup> in contrast to experimental studies that suggest the formation of lamellar phases in bulk systems;<sup>65,66</sup> this discrepancy is likely due to the need for much longer simulation times to form equilibrium phases, currently inaccessible to atomistic models. Given the reduced computational cost of CG models, they can be used to access the time scales required to examine the self-assembly of SC lipids.

Here, we examine the self-assembly of CER NS C24 in water, motivated by its abundance in the SC. Systems composed of 800 lipids and 10,000 water beads are studied, which represents a water to lipid ratio well above that of a fully hydrated bilayer. To remove dependence on the initial configuration, the CG-mapped atomistic bilayer (replicated twice in the  $x$  and  $y$  directions to 4 times the original size) system was heated to



**Figure 8.** Self-assembly of CER NS C24 into an ordered bilayer: (a) initial heating above mixing temperature; (b) cooling from high temperature; (c) full compression; (d) final configuration after compression and expansion. CG beads are colored as follows: TAILS, silver cylinders; AMIDE and MHEAD2, yellow spheres; OH1 and OH2, red spheres; WATER, blue spheres.

**Table 3. Properties of Self-Assembled Bilayers as a Function of Expansion and Compression Time (Relaxation Time)<sup>a</sup>**

final structure	relaxation time (ns)	APL (Å <sup>2</sup> )	$f_{\text{top}}/f_{\text{bottom}}$	$\theta_{\text{top}}/\theta_{\text{bottom}}$ (deg)
bilayer	20	40.00(2)	0.513/0.487	5.2/9.2
	40	40.10(2)	0.511/0.489	5.0/7.2
	60	39.69(3)	0.509/0.491	5.2/9.0
stacked bilayer	200	42.76(2)	0.243/0.250/0.244/0.264	9.6/6.6/8.1/4.9
	400	43.09(3)	0.240/0.255/0.253/0.253	9.5/5.5/6.3/6.1
	600	42.83(2)	0.258/0.246/0.253/0.244	5.6/7.1/5.9/11.1

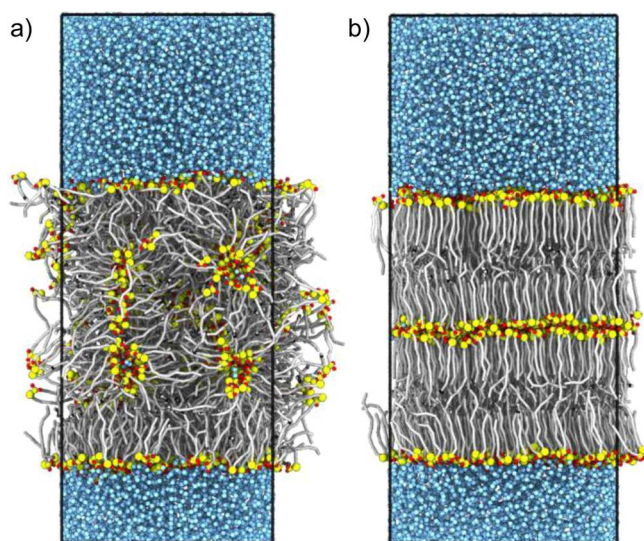
<sup>a</sup> $f_{\text{top}}/f_{\text{bottom}}$  represents the fraction of lipids in the top and bottom leaflets of the self-assembled bilayers;  $\theta_{\text{top}}/\theta_{\text{bottom}}$  represents the average tilt angle of the lipids in the top and bottom leaflets. For the stacked systems,  $f_{\text{top}}/f_{\text{bottom}}$  represents the fraction of lipids in each of the four layers, and  $\theta_{\text{top}}/\theta_{\text{bottom}}$  represents the average tilt angle of the lipids in each of the four layers. The APL of the stacked systems was calculated using the average number of lipids in each layer. The standard error in the tilt angle was less than 3° for each measurement.

1500 K for 20 ns, during which time the lipids and water became completely mixed and disordered, as shown in Figure 8a. Upon cooling to 305 K, a bilayer containing structural defects formed, as shown in Figure 8b. To increase the relaxation rate of the defects, the system was then expanded in the bilayer plane to double the original area over 10 ns while preserving the box volume, as shown in Figure 8c. Upon deforming the box to the original aspect ratio over 10 ns, a defect-free bilayer formed, as shown in Figure 8d. After switching to the anisotropic NPT ensemble, the self-assembled bilayer was found to be stable for 200+ ns. This process was repeated three times with different expansion and compression rates to examine the influence of the relaxation rate on the final structure. The APLs of the bilayers over the final 50 ns of the NPT simulations are compared in Table 3, with little variation between the different systems observed. Interestingly, the self-assembled CG systems are found to pack more densely than the comparable preassembled bilayers, resulting in APLs that are closer to the value from the atomistic simulation (see Table 2, CER NS C24). The time scale associated with this relaxation influences the distribution of lipids between the leaflets in the bilayer structure, with shorter time scales associated with larger deviations from an ideal (i.e., equal) distribution, although all deviations are within 2.6% of ideal. These small variations, however, do create noticeable changes to the average tilt angle of the leaflets (although the average APL is relatively unaffected), with a larger tilt observed in the leaflets with fewer lipids, as shown in Table 3. This trend is expected—with a fewer number of lipids in a leaflet, the lipid tails tilt relative to the bilayer normal to decrease their spacing to an optimal level.

Since the simulated bilayer is a periodic structure (as periodic boundary conditions are used), it is reasonable to assume that the self-assembled structure is dependent on the size and shape of the box. For example, one would expect two bilayers to form in the previously described CER NS C24 system if the box area

parallel to the bilayer plane were reduced by a factor of 2. To explore this concept, the previous self-assembly protocol was repeated, with the exception that the aspect ratio of the simulation box at the end of the initial compression phase was set to half that of the initial self-assembly simulation (i.e., half the cross-sectional area of the bilayer and twice the thickness). As with the bilayer self-assembly, a defected lamellar structure initially formed, where the lipids were well-ordered at the lipid–water interface and less ordered in the interior of the membrane, as shown in Figure 9a. After subsequent expansion and compression cycles, well-ordered, multilamellar structures (i.e., stacked bilayers) formed, with a small amount of water between the middle leaflets (0.35–0.5, water molecules per interior lipid, compared to 50 in the bulk), as shown in Figure 9b. The relaxation rates of the stacked bilayer defects were much slower than that of the single bilayer systems, with 200 ns of both expansion and compression required to remove the defects, compared to 20 ns for the single bilayer assembly. Subsequent replication of this self-assembly protocol yielded similar structures, illustrating reproducibility. We note that, without the expansion and compression to increase the relaxation rate of the defects, a defect-free stacked bilayer structure formed over the course of 1.3 μs at 305 K. Thus, the expansion and compression protocol is used as a means to increase the rate of exploration of the phase space of the CER–water system, rather than to influence the final state of the system.

Examining the multilamellar systems, we note that similar structural trends are observed in comparison to the self-assembled single bilayer systems. For example, the self-assembled multilamellar systems also show deviations from an ideal distribution of lipids per leaflet, although the deviations are larger in the multilamellar systems (see Table 3), which may be expected given the differences between a lipid–water and lipid–lipid interface. A similar relationship between the tilt



**Figure 9.** Self-assembly of CER NS C24 into a stacked bilayer: (a) after changing the aspect ratio of the simulation cell; (b) after subsequent area expansion and compression. Beads colored as follows: TAILS, silver cylinders; AMIDE and MHEAD2, yellow spheres; OH1 and OH2, red spheres; WATER, blue spheres.

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

The lipid–lipid interface in multilamellar structures makes it possible for lipids to adopt different configurations, which may alter the properties of the system. For example, several models of SC lipid organization have been proposed, with a key difference between the models being whether the CER tails adopt an extended conformation (with the two tails on opposite sides of the headgroup), a hairpin conformation (with both tails on the same side of the headgroup), or some combination of the two. For example, the stacked monolayer model assumes an extended conformation for the CERs,<sup>57</sup> while the sandwich model assumes a hairpin conformation of the CERs.<sup>5,68</sup> In contrast, the single gel phase model proposes a mixture of extended and hairpin CER conformations,<sup>69</sup> and the armature reinforcement model proposes a transition from extended to hairpin CER conformations upon hydration.<sup>66,70</sup>

To explore the CER conformations in the stacked bilayer configurations, we examine the angle  $\theta_{\text{tails}}$  between the two tails of each CER molecule.  $\theta_{\text{tails}}$  was defined as the angle between the two vectors from the MHEAD2 bead to the terminal bead of each tail and thus distinguishes between the lipid being in an extended ( $\theta_{\text{tails}} > 90^\circ$ ) or hairpin ( $\theta_{\text{tails}} < 90^\circ$ ) conformation. From the simulations, we find that the vast majority of the lipids in the multilamellar system exist in a hairpin conformation, with no correlation between conformation and expansion/compression time (for all systems, between 91 and 94% of lipids exist in a hairpin conformation). As might be expected, the hairpin configuration is strongly preferred at the lipid–water interface due to the hydrophobic nature of the lipid tails; however, when considering the lipids in the inner leaflets, the number of lipids adopting a hairpin conformation decreases to between 82 and 87%. Thus, while the hairpin conformation

appears to be the preferred CER conformation in the pure, multilamellar CER NS C24 system, roughly 15% of the lipids in the inner leaflets adopt an extended conformation. Thus, the pure CER NS system is most consistent with the hairpin conformation assumed in the sandwich model. However, we note that the CER conformation in the complex lipid mixtures of human SC, which include both cholesterol and FFAs, could be different than those observed for pure CER NS. It is reasonable to assume that the intramolecular bonded forces dictate the CER conformation to an extent. Therefore, to validate the intramolecular potentials and ensure they are not biasing the extended vs hairpin conformations in the stacked bilayer system, a single CER molecule in a vacuum was simulated with the CG and atomistic force fields, and the distributions of the angle between the two tails ( $\theta_{\text{tails}}$ ) and end-to-end distance of the tails calculated with each model. Good agreement between the distributions calculated with the two models was found, with each predicting, on average, a hairpin conformation of the two tails (see Figure S15 of the Supporting Information). Therefore, we conclude the observed hairpin vs extended distribution is physically relevant and not an artifact of the intramolecular bonded forces in the CG force field.

## V. CONCLUSION

We have shown that MS IBI is an appropriate method for deriving CG models of SC lipids, specifically CER NS, which is the most abundant CER in human SC. The transferable nature of the MS IBI-derived force field is illustrated by the fact that the TAIL beads of CER NS and FFA are treated as the same type, and accurately model the tail behavior of the two different types of molecule. This transferability is promising for the development of models for other SC lipids, as the TAILS beads can confidently be transferred to describe the acyl tails, leaving only the headgroup interactions to be derived. Further, the inclusion of explicit hydroxyl groups on the CER headgroups allows the model to accurately capture the in-plane packing of the molecules. This is an important property of any CER model, especially, for example, when studying permeability, where the primary permeability barrier is at the lipid–water interface.<sup>71</sup> In contrast, an implicit treatment of the hydroxyl groups does not allow the model to capture this behavior, which results in less accurate structural properties of lamellar phases with the CG model. This CG mapping also allows for easy extension to other CER species, since the AMIDE and MHEAD2 beads are common to all CERs.

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

While this paper derives a model specific to CER NS, it also provides the necessary framework for developing force fields to accurately model a realistic SC lipid mixture. Recent

experimental work has focused on developing lipid mixtures that adequately reproduce the barrier properties of human skin.<sup>36,66,72–74</sup> CG SC lipid models can offer molecular-level insight into phenomena that cannot be observed through experiment and atomistic molecular simulation, e.g., the self-assembly into and transitions between well-ordered lamellar phases. CG models are thus a practical tool for complementing experimental work in this field, e.g., aiding in the discovery of promising lipid mixtures that reproduce the properties of human SC. CG models for the other lipid classes of SC (FFA and cholesterol) will be published in a future publication, allowing simulation studies of representative SC lipid mixtures.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.6b08046.

Tables containing the states used to gather target data and which states were used to optimize each pair interaction; a table containing the parameters for the CG bonded interactions; a description of the various analyses performed on the simulation trajectories; a summary of the hydrogen bonding calculated from atomistic bilayer simulations; and RDFs and potentials for interaction pairs that are not presented in the text (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: c.mccabe@vanderbilt.edu. Phone: 615 322-6853. Fax: 615 343-7951.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by grant number R01 AR057886-01 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Science Foundation grant number CBET-1028374. This work was conducted in part using computational resources provided by the National Energy Research Scientific Computing Center, supported by the Office of Science of the Department of Energy under Contract No. DE-AC02-05CH11231 and the Advanced Computing Center for Research and Education at Vanderbilt University.

## ■ REFERENCES

- (1) Yardley, H. J.; Summerly, R. Lipid Composition and Metabolism in Normal and Diseased Epidermis. *Pharmacol. Ther.* **1981**, *13* (1970), 357–383.
- (2) Gray, G. M.; Yardley, H. J. Different Populations of Pig Epidermal Cells: Isolation and Lipid Composition. *J. Lipid Res.* **1975**, *16* (6), 441–447.
- (3) Rabionet, M.; Gorgas, K.; Sandhoff, R. Ceramide Synthesis in the Epidermis. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2014**, *1841* (3), 422–434.
- (4) Ponec, M.; Weerheim, A.; Lankhorst, P.; Wertz, P. W. New Acylceramide in Native and Reconstructed Epidermis. *J. Invest. Dermatol.* **2003**, *120* (4), 581–588.
- (5) Bouwstra, J. A.; Dubbelaar, F. E. R.; Gooris, G. S.; Ponec, M. The Lipid Organisation in the Skin Barrier. *Acta Derm. Venereol. Suppl. (Stockh)*. **2000**, *80*, 23–30.
- (6) van Smeden, J.; Janssens, M.; Gooris, G. S.; Bouwstra, J. A. The Important Role of Stratum Corneum Lipids for the Cutaneous Barrier

Function. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2014**, *1841* (3), 295–313.

- (7) van Smeden, J.; Janssens, M.; Kaye, E. C. J.; Caspers, P. J.; Lavrijsen, A. P. M.; Vreeken, R. J.; Bouwstra, J. A. The Importance of Free Fatty Acid Chain Length for the Skin Barrier Function in Atopic Eczema Patients. *Exp. Dermatol.* **2014**, *23* (1), 45–52.
- (8) Feller, S. E. Molecular Dynamics Simulations of Lipid Bilayers. *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 217–223.
- (9) Lindahl, E.; Sansom, M. S. Membrane Proteins: Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18* (4), 425–431.
- (10) Lyubartsev, A. P.; Rabinovich, A. L. Recent Development in Computer Simulations of Lipid Bilayers. *Soft Matter* **2011**, *7* (1), 25.
- (11) Bennett, W. F. D.; Tieleman, D. P. Computer Simulations of Lipid Membrane Domains. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828* (8), 1765–1776.
- (12) Rabinovich, A. L.; Lyubartsev, A. P. Computer Simulation of Lipid Membranes: Methodology and Achievements. *Polym. Sci., Ser. C* **2013**, *55* (1), 162–180.
- (13) Bemporad, D.; Luttmann, C.; Essex, J. W. Computer Simulation of Small Molecule Permeation across a Lipid Bilayer: Dependence on Bilayer Properties and Solute Volume, Size, and Cross-Sectional Area. *Biophys. J.* **2004**, *87* (1), 1–13.
- (14) Groen, D.; Gooris, G. S.; Barlow, D. J.; Lawrence, M. J.; van Mechelen, J. B.; Demé, B.; Bouwstra, J. A. Disposition of Ceramide in Model Lipid Membranes Determined by Neutron Diffraction. *Biophys. J.* **2011**, *100* (6), 1481–1489.
- (15) Školová, B.; Hudská, K.; Pullmannova, P.; Kovacik, A.; Palat, K.; Roh, J.; Palat, K.; Fleddermann, J.; Estrela-Lopis, I.; Vávrová, K. Different Phase Behavior and Packing of Ceramides with Long (C16) and Very Long (C24) Acyls in Model Membranes: Infrared Spectroscopy Using Deuterated Lipids. *J. Phys. Chem. B* **2014**, *118*, 10460–10470.
- (16) Mojumdar, E. H.; Gooris, G. S.; Barlow, D. J.; Lawrence, M. J.; Deme, B.; Bouwstra, J. A. Skin Lipids: Localization of Ceramide and Fatty Acid in the Unit Cell of the Long Periodicity Phase. *Biophys. J.* **2015**, *108* (11), 2670–2679.
- (17) Guo, S.; Moore, T. C.; Iacovella, C. R.; Strickland, L. A.; McCabe, C. Simulation Study of the Structure and Phase Behavior of Ceramide Bilayers and the Role of Lipid Head Group Chemistry. *J. Chem. Theory Comput.* **2013**, *9* (11), 5116–5126.
- (18) Pandit, S. A.; Scott, H. L. Molecular-Dynamics Simulation of a Ceramide Bilayer. *J. Chem. Phys.* **2006**, *124* (1), 014708.
- (19) Das, C.; Noro, M. G.; Olmsted, P. D. Simulation Studies of Stratum Corneum Lipid Mixtures. *Biophys. J.* **2009**, *97* (7), 1941–1951.
- (20) Notman, R.; den Otter, W. K.; Noro, M. G.; Briels, W. J.; Anwar, J. The Permeability Enhancing Mechanism of DMSO in Ceramide Bilayers Simulated by Molecular Dynamics. *Biophys. J.* **2007**, *93* (6), 2056–2068.
- (21) Gupta, R.; Rai, B. Molecular Dynamics Simulation Study of Skin Lipids: Effects of the Molar Ratio of Individual Components over a Wide Temperature Range. *J. Phys. Chem. B* **2015**, *119* (35), 11643–11655.
- (22) Shah, J.; Atienza, J. M.; Duclos, R. I.; Rawlings, A. V.; Dong, Z.; Shipley, G. G. Structural and Thermotropic Properties of Synthetic C16:0 (Palmitoyl) Ceramide: Effect of Hydration. *J. Lipid Res.* **1995**, *36* (9), 1936–1944.
- (23) Bouwstra, J. A.; Gooris, G. S.; Cheng, K.; Weerheim, A.; Bras, W.; Ponec, M. Phase Behavior of Isolated Skin Lipids. *J. Lipid Res.* **1996**, *37* (5), 999–1011.
- (24) Moore, D. J.; Rerek, M. E.; Mendelsohn, R. FTIR Spectroscopy Studies of the Conformational Order and Phase Behavior of Ceramides. *J. Phys. Chem. B* **1997**, *101* (44), 8933–8940.
- (25) Brief, E.; Kwak, S.; Cheng, J. T. J.; Kitson, N.; Thewalt, J.; Lafleur, M. Phase Behavior of an Equimolar Mixture of N-Palmitoyl-D-Erythro-Sphingosine, Cholesterol, and Palmitic Acid, a Mixture with Optimized Hydrophobic Matching. *Langmuir* **2009**, *25* (13), 7523–7532.

- (26) Paz Ramos, A.; Lafleur, M. The Chain Length of Free Fatty Acids Influences the Phase Behaviour of Stratum Corneum Model Membranes. *Langmuir* **2015**, *31* (42), 11621–11629.
- (27) Hoopes, M. I.; Noro, M. G.; Longo, M. L.; Faller, R. Bilayer Structure and Lipid Dynamics in a Model Stratum Corneum with Oleic Acid. *J. Phys. Chem. B* **2011**, *115* (12), 3164–3171.
- (28) Das, C.; Noro, M. G.; Olmsted, P. D. Lamellar and Inverse Micellar Structures of Skin Lipids: Effect of Templating. *Phys. Rev. Lett.* **2013**, *111* (14), 148101.
- (29) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812–7824.
- (30) Shinoda, W.; DeVane, R. H.; Klein, M. L. Multi-Property Fitting and Parameterization of a Coarse Grained Model for Aqueous Surfactants. *Mol. Simul.* **2007**, *33* (1–2), 27–36.
- (31) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109* (7), 2469–2473.
- (32) Hadley, K. R.; M<sup>c</sup>Cabe, C. A Simulation Study of the Self-Assembly of Coarse-Grained Skin Lipids. *Soft Matter* **2012**, *8* (17), 4802.
- (33) Sovova, Z.; Berka, K.; Otyepka, M.; Jurecka, P. Coarse-Grain Simulations of Skin Ceramide NS with Newly Derived Parameters Clarify Structure of Melted Phase. *J. Phys. Chem. B* **2015**, *119* (10), 3988–3998.
- (34) Notman, R.; Anwar, J.; Briels, W. J.; Noro, M. G.; den Otter, W. K. Simulations of Skin Barrier Function: Free Energies of Hydrophobic and Hydrophilic Transmembrane Pores in Ceramide Bilayers. *Biophys. J.* **2008**, *95* (10), 4763–4771.
- (35) Moore, D. J.; Rerek, M. E.; Mendelsohn, R. Role of Ceramides 2 and 5 in the Structure of the Stratum Corneum Lipid Barrier. *Int. J. Cosmet. Sci.* **1999**, *21* (5), 353–368.
- (36) Garidel, P.; Fölting, B.; Schaller, I.; Kerth, A. The Microstructure of the Stratum Corneum Lipid Barrier: Mid-Infrared Spectroscopic Studies of Hydrated Ceramide:palmitic Acid:cholesterol Model Systems. *Biophys. Chem.* **2010**, *150* (1–3), 144–156.
- (37) Pascher, I. Molecular Arrangements in Sphingolipids. Conformation and Hydrogen Bonding of Ceramide and Their Implication on Membrane Stability and Permeability. *Biochim. Biophys. Acta, Biomembr.* **1976**, *455* (2), 433–451.
- (38) Löfgren, H.; Pascher, I. Molecular Arrangements of Sphingolipids. The Monolayer Behaviour of Ceramides. *Chem. Phys. Lipids* **1977**, *20* (4), 273–284.
- (39) Rerek, M. E.; Chen, H.; Markovic, B.; Van Wyck, D.; Garidel, P.; Mendelsohn, R.; Moore, D. J. Phytosphingosine and Sphingosine Ceramide Headgroup Hydrogen Bonding: Structural Insights through Thermotropic Hydrogen/Deuterium Exchange. *J. Phys. Chem. B* **2001**, *105* (38), 9355–9362.
- (40) Moore, T. C.; Iacovella, C. R.; M<sup>c</sup>Cabe, C. Derivation of Coarse-Grained Potentials via Multistate Iterative Boltzmann Inversion. *J. Chem. Phys.* **2014**, *140* (22), 224104.
- (41) Bayramoglu, B.; Faller, R. Coarse-Grained Modeling of Polystyrene in Various Environments by Iterative Boltzmann Inversion. *Macromolecules* **2012**, *45* (22), 9205–9219.
- (42) Bayramoglu, B.; Faller, R. Modeling of Polystyrene under Confinement: Exploring the Limits of Iterative Boltzmann Inversion. *Macromolecules* **2013**, *46* (19), 7957–7976.
- (43) Moore, T. C.; Iacovella, C. R.; M<sup>c</sup>Cabe, C. Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion. In *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*; Snurr, R. Q., Adjiman, C. S., Kofke, D. A., Eds.; Springer: Singapore, 2016; pp 37–52.
- (44) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114* (23), 7830–7843.
- (45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.
- (46) Plimpton, S. J. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (47) LAMMPS web page, <http://lammps.sandia.gov/> (accessed Jan 1, 2016).
- (48) Plimpton, S. J.; Pollock, R.; Stevens, M. J. Particle Mesh Ewald and rRESPA for Parallel Molecular Dynamics Simulations. *Proc. Eighth Siam Conf. Parallel Process. Sci. Comput.* **1997**, 1–13.
- (49) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101* (5), 4177.
- (50) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General Purpose Molecular Dynamics Simulations Fully Implemented on Graphics Processing Units. *J. Comput. Phys.* **2008**, *227* (10), 5342–5359.
- (51) HOOMD-blue web page, <http://glotzerlab.engin.umich.edu/hoomd-blue/> (accessed Jan 1, 2016).
- (52) Hadley, K. R.; M<sup>c</sup>Cabe, C. A Coarse-Grained Model for Amorphous and Crystalline Fatty Acids. *J. Chem. Phys.* **2010**, *132* (13), 134505.
- (53) Hadley, K. R.; M<sup>c</sup>Cabe, C. On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties. *J. Phys. Chem. B* **2010**, *114* (13), 4590–4599.
- (54) Hadley, K. R.; M<sup>c</sup>Cabe, C. A Structurally Relevant Coarse-Grained Model for Cholesterol. *Biophys. J.* **2010**, *99* (9), 2896–2905.
- (55) Milano, G.; Goudeau, S.; Müller-Plathe, F. Multicentered Gaussian-Based Potentials for Coarse-Grained Polymer Simulations: Linking Atomistic and Mesoscopic Scales. *J. Polym. Sci., Part B: Polym. Phys.* **2005**, *43* (8), 871–885.
- (56) MSIBI, <https://github.com/ctk3b/msibi> (accessed Jan 1, 2016).
- (57) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109* (8), 1528–1532.
- (58) MDTraj, <http://mdtraj.org/latest/> (accessed Jan 1, 2016).
- (59) Faller, R.; Reith, D. Properties of Poly(isoprene): Model Building in the Melt and in Solution. *Macromolecules* **2003**, *36* (14), 5406–5414.
- (60) Marrink, S. J.; Lindahl, E.; Edholm, O.; Mark, A. E. Simulation of the Spontaneous Aggregation of Phospholipids into Bilayers. *J. Am. Chem. Soc.* **2001**, *123* (35), 8638–8639.
- (61) Brockman, H. L.; Momsen, M. M.; Brown, R. E.; He, L.; Chun, J.; Byun, H.-S.; Bittman, R. The 4,5-Double Bond of Ceramide Regulates Its Dipole Potential, Elastic Properties, and Packing Behavior. *Biophys. J.* **2004**, *87* (3), 1722–1731.
- (62) Scheffer, L.; Solomonov, I.; Weygand, M. J.; Kjaer, K.; Leiserowitz, L.; Addadi, L. Structure of Cholesterol/ceramide Monolayer Mixtures: Implications to the Molecular Organization of Lipid Rafts. *Biophys. J.* **2005**, *88* (5), 3381–3391.
- (63) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72* (5), 2002–2013.
- (64) Metcalf, R.; Pandit, S. A. Mixing Properties of Sphingomyelin Ceramide Bilayers: A Simulation Study. *J. Phys. Chem. B* **2012**, *116*, 4500–4509.
- (65) Wartewig, S.; Neubert, R. H. H. Properties of Ceramides and Their Impact on the Stratum Corneum Structure: A Review. Part 1: Ceramides. *Skin Pharmacol. Physiol.* **2007**, *20* (5), 220–229.
- (66) Kessner, D.; Ruettinger, A.; Kiselev, M. A.; Wartewig, S.; Neubert, R. H. H. Properties of Ceramides and Their Impact on the Stratum Corneum Structure. Part 2: Stratum Corneum Lipid Model Systems. *Skin Pharmacol. Physiol.* **2008**, *21* (2), 58–74.
- (67) Swartzendruber, D. C.; Wertz, P. W.; Kitko, D. J.; Madison, K. C.; Downing, D. T. Molecular Models of the Intercellular Lipid Lamellae in Mammalian Stratum Corneum. *J. Invest. Dermatol.* **1989**, *92*, 251–257.

(68) Kuempel, D.; Swartzendruber, D. C.; Squier, C. A.; Wertz, P. W. In Vitro Reconstitution of Stratum Corneum Lipid Lamellae. *Biochim. Biophys. Acta, Biomembr.* **1998**, *1372* (1), 135–140.

(69) Norlén, L. Skin Barrier Structure and Function: The Single Gel Phase Model. *J. Invest. Dermatol.* **2001**, *117*, 830–836.

(70) Kiselev, M. A. Conformation of Ceramide 6 Molecules and Chain-Flip Transitions in the Lipid Matrix of the Outermost Layer of Mammalian Skin, the Stratum Corneum. *Crystallogr. Rep.* **2007**, *52* (3), 525–528.

(71) Das, C.; Olmsted, P. D.; Noro, M. G. Water Permeation through Stratum Corneum Lipid Bilayers from Atomistic Simulations. *Soft Matter* **2009**, *5* (22), 4549.

(72) de Jager, M. W.; Gooris, G. S.; Dolbnya, I. P.; Bras, W.; Ponec, M.; Bouwstra, J. A. Novel Lipid Mixtures Based on Synthetic Ceramides Reproduce the Unique Stratum Corneum Lipid Organization. *J. Lipid Res.* **2004**, *45* (5), 923–932.

(73) Rerek, M. E.; Van Wyck, D.; Mendelsohn, R.; Moore, D. J. FTIR Spectroscopic Studies of Lipid Dynamics in Phytosphingosine Ceramide Models of the Stratum Corneum Lipid Matrix. *Chem. Phys. Lipids* **2005**, *134* (1), 51–58.

(74) Gooris, G. S.; Bouwstra, J. A. Infrared Spectroscopic Study of Stratum Corneum Model Membranes Prepared from Human Ceramides, Cholesterol, and Fatty Acids. *Biophys. J.* **2007**, *92* (8), 2785–2795.

SI for Pub.

A Coarse-grained Model of Stratum Corneum Lipids: Free Fatty Acids and  
Ceramide NS

Timothy C. Moore,<sup>1,2</sup> Christopher R. Iacovella,<sup>1,2</sup> Remco Hartkamp,<sup>1,2</sup> Annette L. Bunge,<sup>3</sup>  
and Clare M<sup>c</sup>Cabe<sup>1,2,4</sup>

<sup>1</sup>*Department of Chemical and Biomolecular Engineering, Vanderbilt University,  
Nashville, TN 37235*

<sup>2</sup>*Vanderbilt University Multiscale Modeling and Simulation (MuMS) Facility, Nashville,  
TN 37235*

<sup>3</sup>*Department of Chemical and Biological Engineering, Colorado School of Mines,  
Golden, CO 80401*

<sup>4</sup>*Department of Chemistry, Vanderbilt University, Nashville, TN 37235*

**Table S1.** Target states used for MS IBI optimizations. The states used to optimize each interaction are described in the text and given in Table S2. State condition specifies the ensemble – states with a pressure listed were simulated in the NPT ensemble; states with a density listed were simulated in the NVT ensemble. The dehydrated bilayer values are the area per lipid and the thickness (i.e., the length of the simulation box normal to the bilayer plane since water is absent from this state).

<b>Name</b>	<b>State Condition</b>	<b>Temperature, K</b>	<b>Label for Table S2</b>
<b>Pure FFA C24:0</b>			
bulk fluid	1.0 atm	500	FFA1
bulk fluid	0.811 g/mL	800	FFA2
monolayer	30.0 Å <sup>2</sup> /chain	305	MONO
<b>FFA-water</b>			
0.4 mol% FFA	1.0 atm	305	W-FFA1
0.5 mol% FFA	1.0 atm	305	W-FFA2
0.85 mol% FFA	1.0 atm	305	W-FFA3
<b>Pure CER NS</b>			
bulk fluid	1.0 atm	500	CER1
bulk fluid	0.721 g/mL	500	CER2
dehydrated bilayer	1.0 atm	305	DEH-BIL-NPT
dehydrated bilayer	39.96 Å <sup>2</sup> /lipid; 57.18 Å	305	DEH-BIL-NVT
<b>CER NS-water</b>			
bilayer	1.0 atm	305	BIL-NPT
bilayer	40.0 Å <sup>2</sup> /lipid	305	BIL-NVT
CER headgroups	1.0 atm	400	W-CERHEAD

**Table S2.** States used to optimize each pair interaction and the weight  $\alpha_{s,0}$  given to each state (as described in the main text). Pairs are denoted by bold text, with the states used to optimize a given interaction listed below the interaction name. The state condition names correspond to the right-most column of Table S1. To illustrate the dependence of the chosen weights on the force field, an additional optimization of the lipid-lipid self-interaction potentials was performed, where relatively higher weights are given to the bulk fluid states, given as  $\alpha_{s,0}$  (high bulk). Note that lipid-water interactions were not optimized with  $\alpha_{s,0}$  (high bulk).

<b>States used to optimize each interaction</b>	$\alpha_{s,0}$	$\alpha_{s,0}$ (high bulk)
<b>TER2-TER2</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
<b>TER2-TAIL</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
<b>TER2-FHEAD</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
<b>TAIL-TAIL</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
MONO	0.4	0.1
DEH-BIL-NPT	0.8	0.1
DEH-BIL-NVT	0.8	0.1
<b>TAIL-FHEAD</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
<b>FHEAD-FHEAD</b>		
FFA1	0.1	0.8
FFA2	0.1	0.8
<b>CERHEADs-CERHEADs</b>		
CER1	0.1	0.8
CER2	0.1	0.8
DEH-BIL-NPT	0.8	0.1
DEH-BIL-NVT	0.8	0.1
<b>CERHEADs-TAILs</b>		
CER1	0.1	0.8
CER2	0.1	0.8
<b>FHEAD-WATER</b>		
W-FFA1	0.3	-
W-FFA2	0.3	-
W-FFA3	0.3	-
<b>TAILs-WATER</b>		
W-FFA1	0.3	-
W-FFA2	0.3	-
W-FFA3	0.3	-
<b>CERHEADs-WATER</b>		
BIL-NPT	0.3	-
BIL-NVT	0.3	-
W-CERHEAD	0.3	-

**Table S3.** Bond and angle parameters for the 4-site CG model. For bonds,  $x_0$  represents the equilibrium bond length in Å and  $k_x$  the harmonic bond stretching constant in kcal/mol-Å<sup>2</sup>. For angles,  $x_0$  represents the equilibrium angle in degrees and  $k_x$  the harmonic angle bending constant in kcal/mol-rad<sup>2</sup>.

<b>Name</b>	$x_0$	$k_x$
<b>Bonds</b>		
TER2-TAIL	3.41	35.576
TAIL-TAIL	3.92	20.928
TAIL-FHEAD	3.28	24.296
TAIL-AMIDE	3.01	103.857
AMIDE-MHEAD2	2.92	190.045
MHEAD2-TAIL	3.62	27.575
MHEAD2-OH1	2.326	45.307
MHEAD2-OH2	2.067	151.987
<b>Angles</b>		
TER2-TAIL-TAIL	155.0	7.93
TAIL-TAIL-TAIL	158.0	6.33
TAIL-TAIL-FHEAD	148.0	11.63
TAIL-TAIL-AMIDE*	158.0	6.33
MHEAD2-AMIDE-TAIL	106.0	29.0
AMIDE-MHEAD2-TAIL	107.0	5.285
TAIL-TAIL-MHEAD2*	158.0	6.33
AMIDE-MHEAD2-OH1	109.0	8.388
AMIDE-MHEAD2-OH2	105.0	10.759
TAIL-MHEAD2-OH1	128.0	5.671
TAIL-MHEAD2-OH2	69.0	361.765
OH1-MHEAD2-OH2	115.0	12.22

\*Angles treated the same as TAIL-TAIL-TAIL because the sampled distributions from the AA simulations were very similar to the TAIL-TAIL-TAIL distribution.

All constants are given for a harmonic potential of the form:

$$V(x) = \frac{1}{2} k_x (x - x_0)^2,$$

where  $x_0$  is the equilibrium value and  $k_x$  is the spring constant. Note the factor of  $\frac{1}{2}$  is not included in the spring constant.

## Analysis

*Area Per Lipid:* For bilayer systems, the area per lipid (*APL*) was taken as the cross-sectional area of the simulation box divided by the number of lipids per leaflet.

*Bilayer Thickness:* For hydrated bilayers, the bilayer thickness was calculated as the distance between the lipid-water dividing surfaces on either side of the bilayer. The dividing surface was defined as the location where the water density drops to  $e^{-1}$  of the bulk value in the direction normal to the lipid-water interface. For dehydrated systems the bilayer thickness was taken as the magnitude of the box vector normal to the bilayer plane.

*Nematic Order Parameter:* The nematic order parameter,  $S_2$ , measures the global orientational order of the lipid tails in each leaflet. A value of 1 indicates perfect orientational order, where all lipid tails are oriented in the same direction. A value of 0 indicates a completely isotropic system. To calculate  $S_2$ , the director of each tail in a leaflet is taken as the eigenvector associated with the smallest eigenvalue of the moment of inertia tensor of the tail. For CERs, a separate director is calculated for each tail, whereas an FFA molecule is described by a single director. The director for each tail is used to construct the nematic tensor of the system; the nematic order parameter is defined as the largest eigenvalue of the nematic tensor. For bilayer systems, we report  $S_2$  as the average  $S_2$  over each leaflet in the system.

*Tilt Angle:* The tilt angle of the lipids in a bilayer leaflet are determined by calculating the angle between the tail director vector and the bilayer normal for each tail in the system, and taking the average over all tails in the system.

*Contact Angle:* To tune hydrophobic/hydrophilic interactions, simulations were performed of water droplets on lipid surfaces. The hydrophobicity/hydrophilicity of particular beads was tuned by scaling the CG pair interaction with water until the CG contact angle agreed with the AA contact angle. To calculate the contact angle, the density profile of an ideal spherical cap was fit to the density profile of water on the

surface, with the assumption that the water forms a spherical cap on the surface. The density profile fit gives the height  $h$  of the spherical cap, along with the radius of the sphere  $R$ , from which the contact angle can be calculated as:

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

*Hydrogen Bonding:* An analysis of the hydrogen bonding in the atomistic simulations was performed to identify which CER headgroup CG beads interact most strongly. Hydrogen bond donors were defined as nitrogen or oxygen atoms bonded to a hydrogen, and hydrogen bond acceptors were defined as any nitrogen or oxygen atom. Thus, a hydrogen bond is represented as D-H...A, where D denotes the donor atom, H is the hydrogen involved in the hydrogen bond (and is covalently bonded to the donor atom), and A denotes the acceptor atom. A geometric criterion was used to determine if a hydrogen bond exists between a donor/acceptor pair, following the definition used by Ferrario et al.:<sup>1</sup> a hydrogen bond is considered to exist if 1) the donor-acceptor distance  $r_{AD} \leq 3.5 \text{ \AA}$ , 2) the hydrogen-acceptor distance  $r_{AH} \leq 2.6 \text{ \AA}$ , and 3) the hydrogen-donor-acceptor angle  $\theta_{DHA} \leq 30^\circ$ .

### **In-Plane Hydrogen Bonding in Dehydrated CER NS Bilayer**

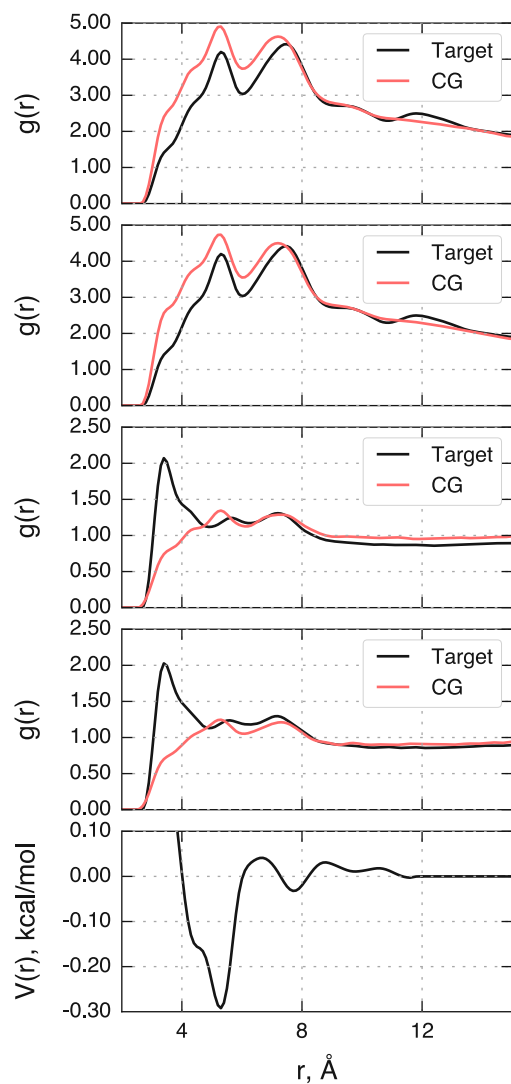
The hydrogen bonding behavior of the lipid headgroups in the dehydrated CER NS bilayer is considered here. Note that although the dehydrated bilayer is effectively an infinitely repeating lamellar structure through the periodic boundaries, only in-plane hydrogen bonding is considered, that is, hydrogen bonding between molecules in the same leaflet; trans-leaflet hydrogen bonds are not considered here.

**Table S4.** Summary of in-plane hydrogen bonding in the atomistic dehydrated CER NS C24 bilayer mapped to the CG level. Note that the acceptor-donor pairs are the CG beads and that a H-bond is considered to exist between CG beads if there is a hydrogen bond between atoms represented by the CG beads. Numbers in parenthesis represent the variation in the final reported digit based on the standard deviation.

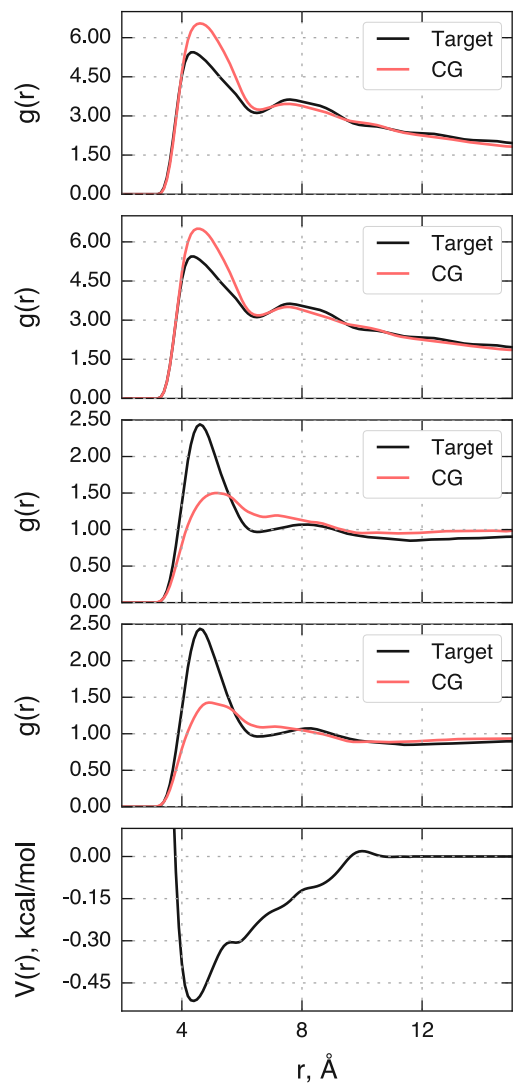
<b>Pair</b>	<b>Hydrogen bonds per lipid</b>
OH1-AMIDE	0.20(4)
OH1-OH2	0.15(5)
OH2-AMIDE	0.10(2)
OH1-OH1	0.03(1)
OH2-OH2	0.020(1)
AMIDE-AMIDE	0.017(6)

## RDFs and potentials for pair interactions not reported in the text.

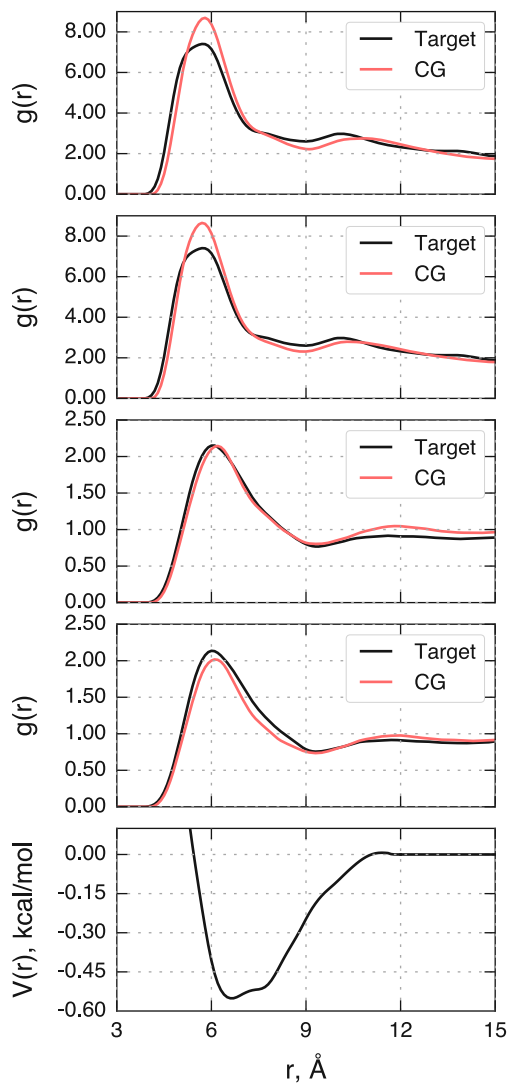
### 4-Site CER Headgroup Mapping



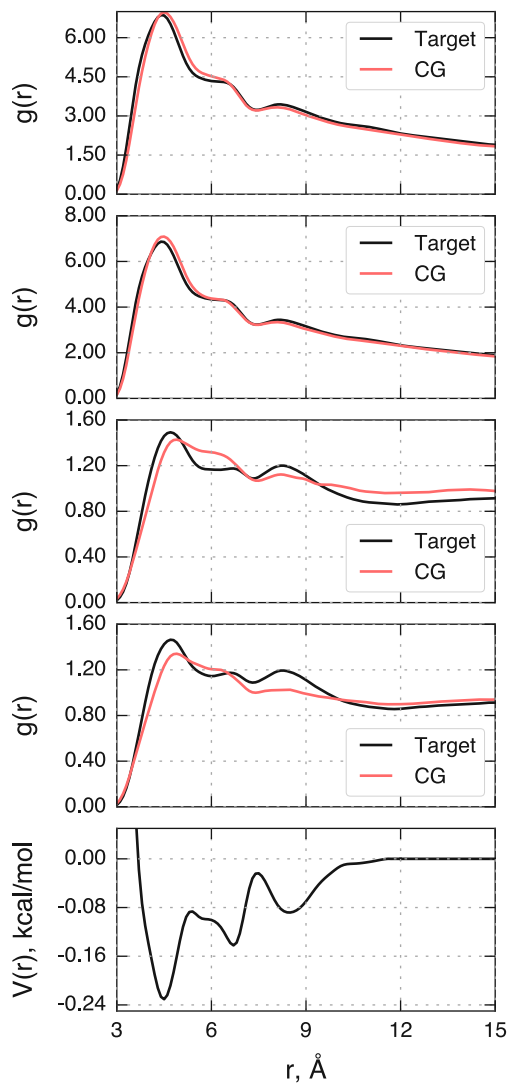
**Figure S1.** AMIDE-AMIDE RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential. We note that the first peak in the fluid state RDF is not captured with the CG model. As emphasized in the main text, the goal of this force field is to reproduce fluid structure in lamellar states; thus, a larger weight was given to the dehydrated bilayer and a better fit is found for the RDF at that state.



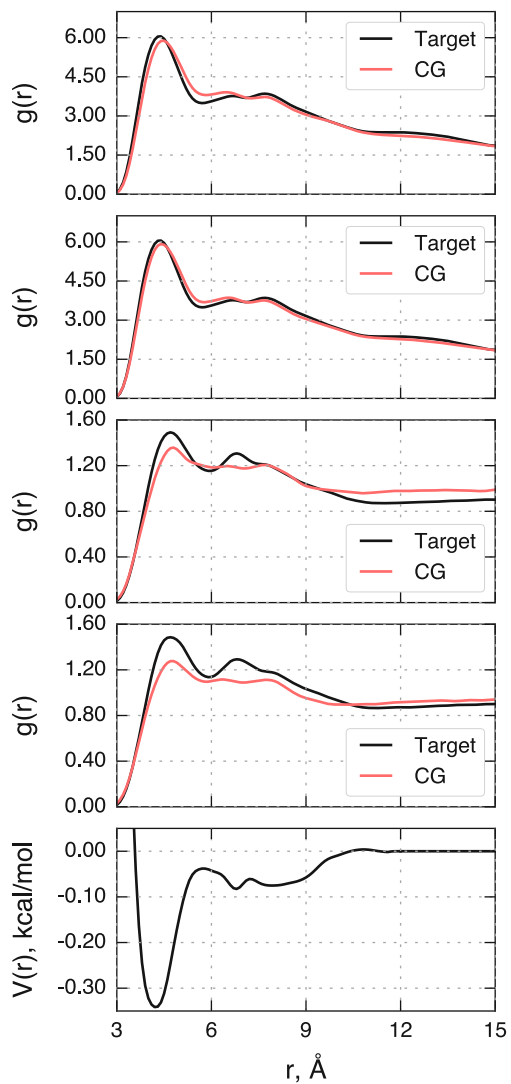
**Figure S2.** AMIDE-MHEAD2 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



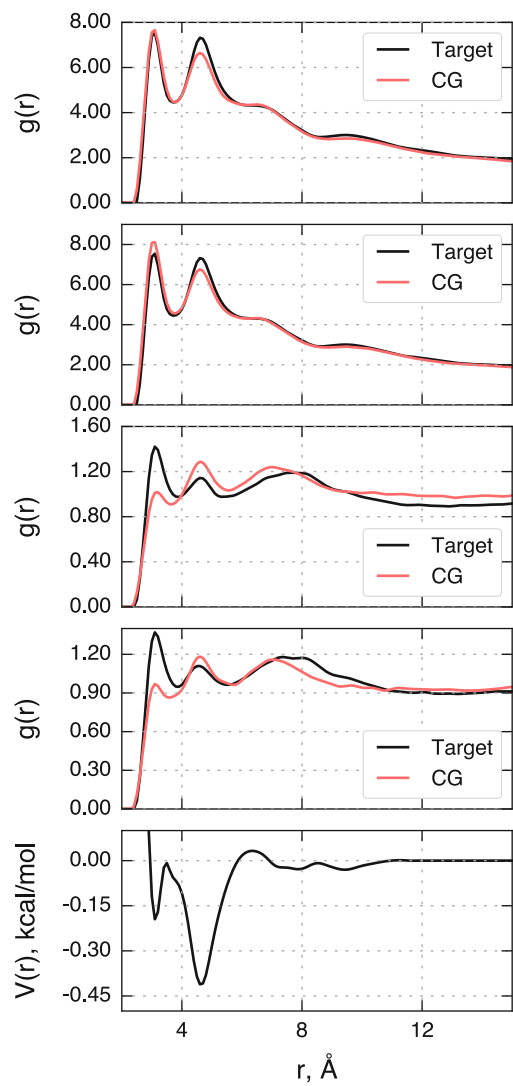
**Figure S3.** MHEAD2-MHEAD2 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



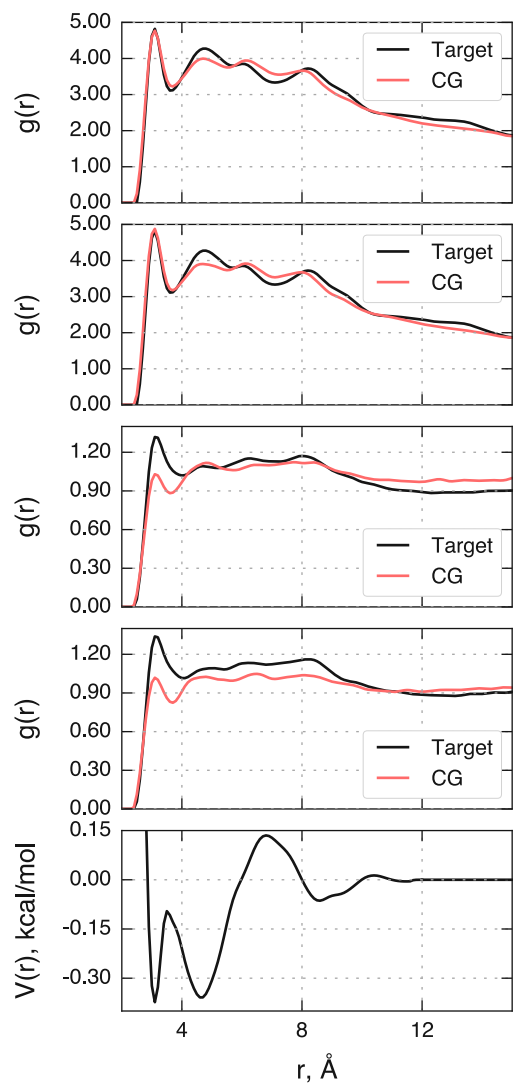
**Figure S4.** MHEAD2-OH1 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



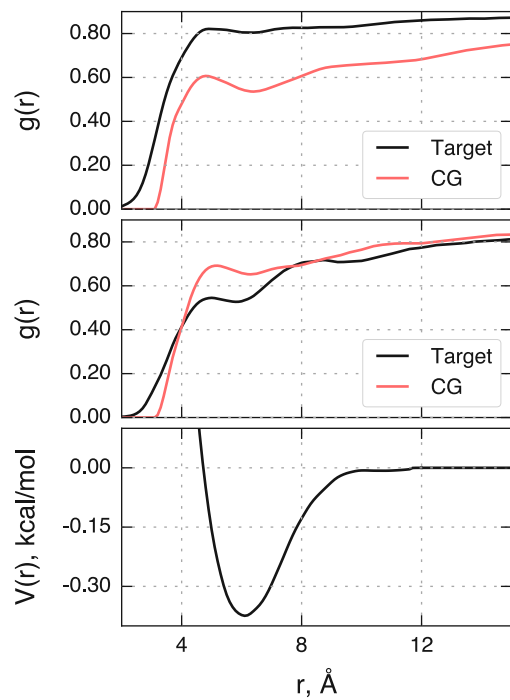
**Figure S5.** MHEAD2-OH2 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



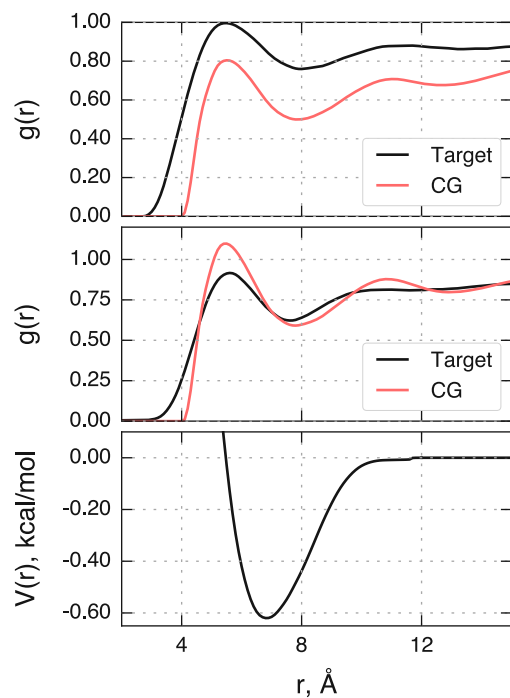
**Figure S6.** OH1-OH1 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



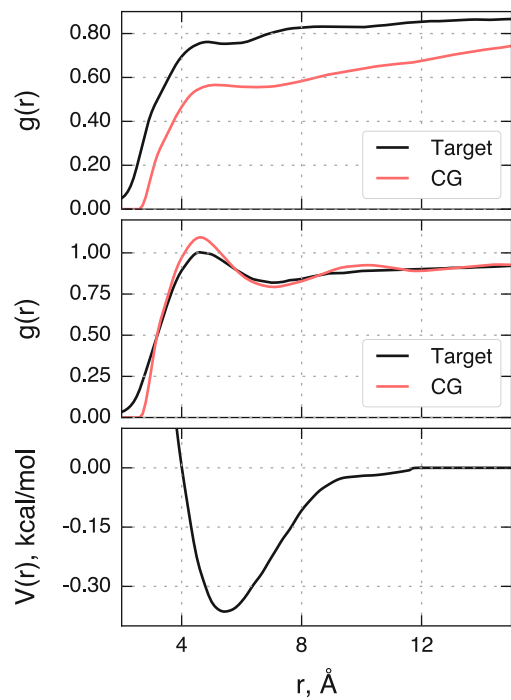
**Figure S7.** OH2-OH2 RDFs and potential. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential.



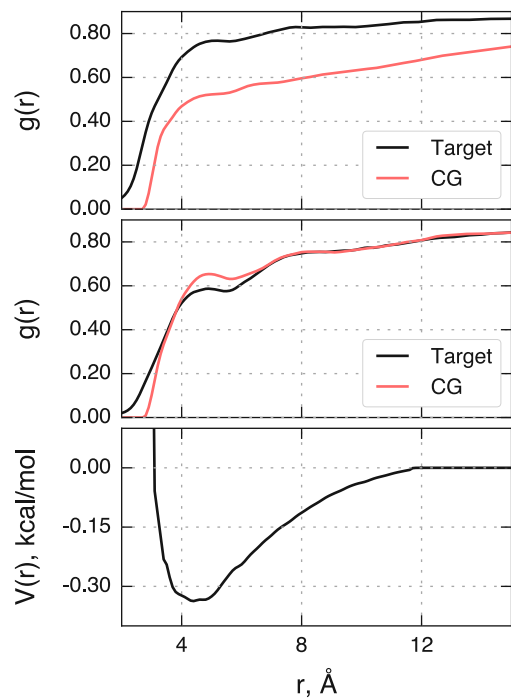
**Figure S8.** AMIDE-WATER RDFs and potential. Top: CER headgroups in water RDFs; middle: hydrated NPT bilayer RDFs; bottom: final derived potential.



**Figure S9.** MHEAD2-WATER RDFs and potential. Top: CER headgroups in water RDFs; middle: hydrated NPT bilayer RDFs; bottom: final derived potential.



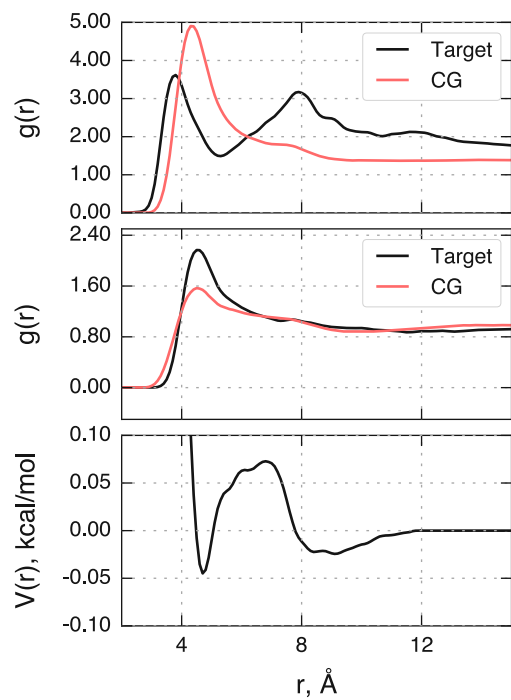
**Figure S10.** OH1-WATER RDFs and potential. Top: CER headgroups in water RDFs; middle: hydrated NPT bilayer RDFs; bottom: final derived potential.



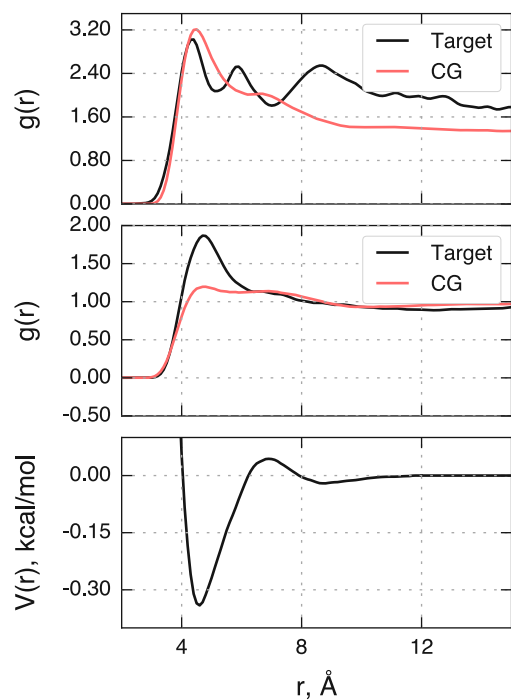
**Figure S11.** OH2-WATER RDFs and potential. Top: CER headgroups in water RDFs; middle: hydrated NPT bilayer RDFs; bottom: final derived potential.

### *3-Site CER Headgroup Mapping*

Here, we present the RDFs and potentials between the headgroup beads in the 3-site CER headgroup mapping. These RDFs illustrate the main shortcoming of using a 3-site mapping for the CER headgroups: the ordering in the lamellar phases is not captured when using a 3-site mapping. This was the motivating force behind developing the 4-site CER headgroup CG model. This lack of CER headgroup ordering is apparent in the NHEAD-MHEAD RDF, shown in Figure S12. Atoms in these groups form more hydrogen bonds with each other than between any other pair in the atomistic simulations. Thus, this pair represents the hydrogen bonding network formed between the lipid headgroups. The target RDF for this pair at the dehydrated bilayer state shows two distinct peaks, indicating there is significant ordering between these headgroup beads, likely a result of the hydrogen bonding present between atoms in these beads. The CG RDF inaccurately predicts the first peak, and only predicts a minor shoulder for the second peak, clearly showing a lack of headgroup ordering and hydrogen bond formation. The atoms in the NHEAD and SHEAD beads also form a relatively high number of hydrogen bonds, so this pair can also serve as a measuring stick for the headgroup packing. The NHEAD-SHEAD RDFs further reveal that the headgroup packing is not captured with the 3-site CER headgroup mapping, as shown in Figure S13. The dehydrated bilayer target RDF clearly shows three distinct peaks, however, only the first is reproduced with the CG model. The lack of order in the CERHEADs-CERHEADs RDFs combined with inaccurate structural properties of the lamellar phases of CER NS C24 predicted with the 3-site CER headgroup mapping model clearly show that a CG model must account for the directional headgroup interactions in CER.



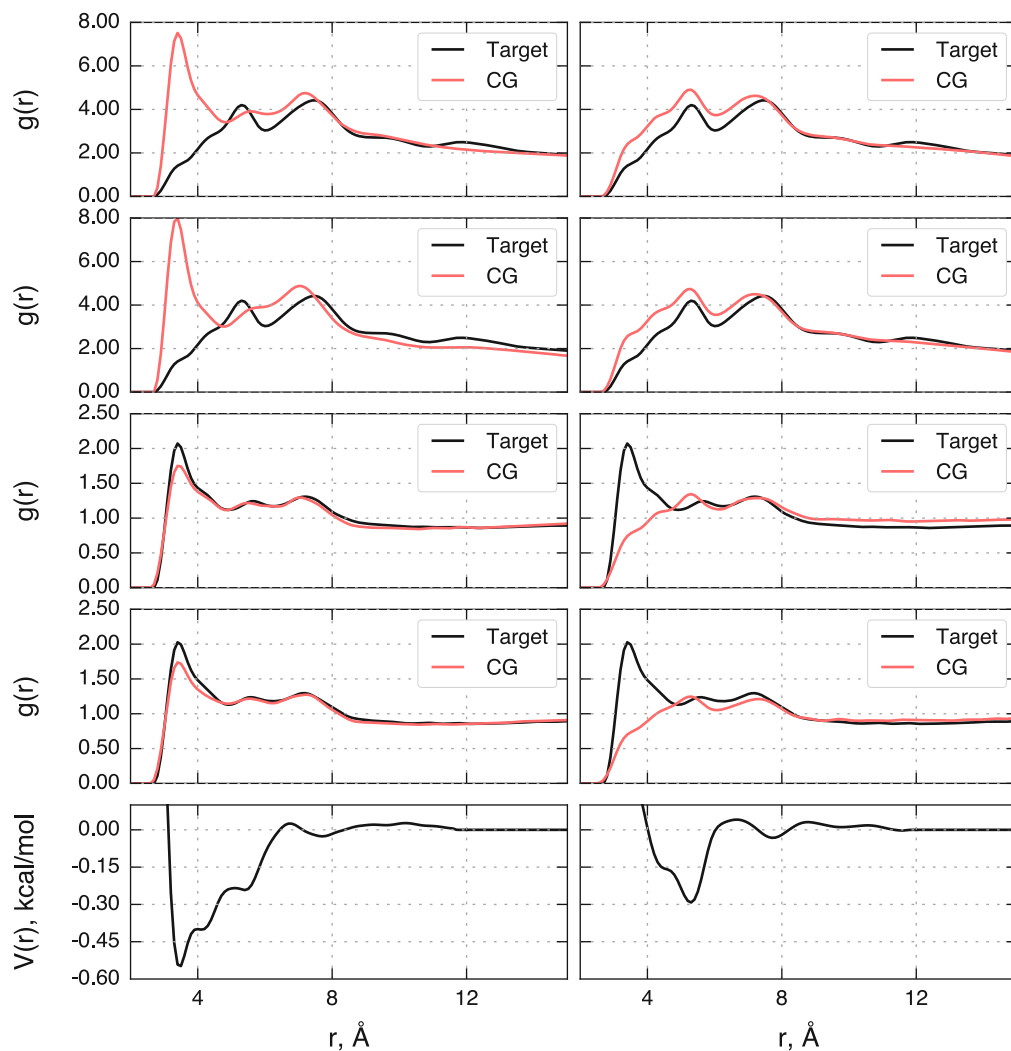
**Figure S12.** NHEAD-MHEAD RDFs and final derived potential. Top: dehydrated bilayer RDFs; middle: isotropic fluid RDFs; bottom: final derived NHEAD-MHEAD potential.



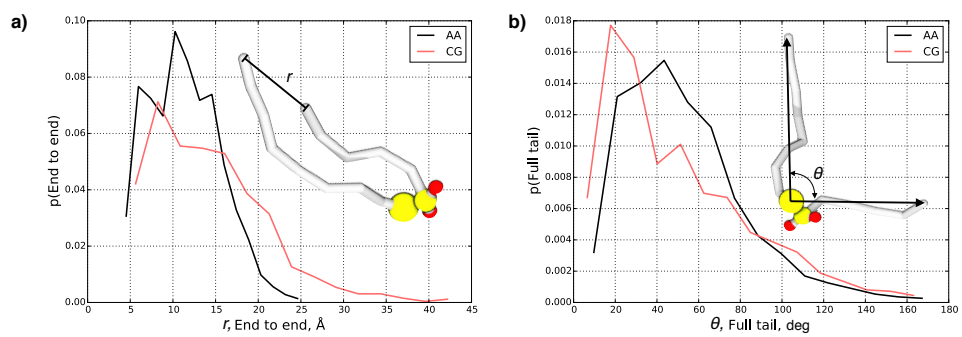
**Figure S13.** NHEAD-SHEAD RDFs and final derived potential. Top: dehydrated bilayer RDFs; middle: isotropic fluid RDFs; bottom: final derived NHEAD-SHEAD potential.

### Effect of Relative Weights on Optimized Force Field

In addition to the potential optimization described at length in the main text, an additional optimization was performed to investigate the effect of the weight given to each state on the optimized force field. In this case, a higher weight was given to the bulk fluid states, listed as  $\alpha_{s,0}(\text{high bulk})$  in Table S2. The effect of this selection of weights is that the fit of the AMIDE-AMIDE target RDF at the bulk states improves, albeit at the expense of the fit at the bilayer states where a strong first neighbor correlation appears that is not present in the target RDF, as shown in Figure S14. Additionally, the *APL* of the dehydrated bilayer increases to  $49.8(2) \text{ \AA}^2$  when using the force field derived with  $\alpha_{s,0}(\text{high bulk})$ ; this value is significantly higher than the atomistic value of  $40.0(1) \text{ \AA}^2$ , indicating that more of the bulk fluid behavior is captured as expected. Given that the ultimate goal of this force field is accurate representation of the structural properties of stratum corneum lipid lamellae, the higher weights reported in Table S2 were assigned to the lamellar states in the force field optimization. The results presented here serve to further illustrate the effect of varying the weights given to the different states in the optimization and are consistent with observations in earlier work.<sup>2</sup>



**Figure S14.** AMIDE-AMIDE RDFs and potentials from different optimizations to illustrate the effect of the relative weighting factors given to each state. Left: weights listed as  $\alpha_{s,0}$ (high bulk) in Table S2; right: weights listed as  $\alpha_{s,0}$  in Table S2. From top to bottom: dehydrated NPT bilayer RDFs; dehydrated NVT bilayer RDFs; isotropic NVT fluid RDFs; isotropic NPT fluid RDFs; final derived potential. We note that the fluid phase RDFs match the target more accurately when a higher weight was given to the lamellar states; however, the fits to the dehydrated bilayer RDFs are significantly less accurate than when a higher weight was given to the lamellar states. Given that the ultimate goal of this force field is accurate representation of the structural properties of stratum corneum lipid lamellae, the higher weights reported in Table S2 were assigned to the lamellar states in the force field optimization. The results presented here serve to further illustrate the effect of varying the weights given to the different states in the optimization and are consistent with observations in earlier work.



**Figure S15.** CER NS conformations from simulations of a single molecule in a vacuum. a) end-to-end distance between the two tails; b) angle between each of the 2 ceramide tails.

## References

1. Ferrario, M., Haughney, M., McDonald, I. R. & Klein, M. L. Molecular-dynamics simulation of aqueous mixtures: Methanol, acetone, and ammonia. *J. Chem. Phys.* **93**, 5156 (1990).
2. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **140**, (2014).