JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

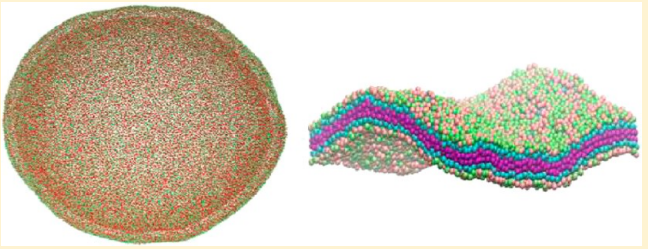
Hybrid Approach for Highly Coarse-Grained Lipid Bilayer Models

Anand Srivastava and Gregory A. Voth*

Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics and Computation Institute, University of Chicago, 5735 S. Ellis Ave., Chicago, Illinois 60637, United States

Supporting Information

ABSTRACT: We present a systematic methodology to develop highly coarse-grained (CG) lipid models for large scale biomembrane simulations, in which we derive CG interactions using a powerful combination of the multiscale coarse-graining (MS-CG) method, and an analytical form of the CG potential to model interactions at short-range. The resulting hybrid coarse-graining (HCG) methodology is used to develop a three-site solvent-free model for 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), and a 1:1 mixture of 1,2-dioleoyl-*sn*-glycero-3-phospho-L-serine (DOPS) and DOPC. In addition, we developed a four-site model of DOPC, demonstrating the capability of the HCG methodology in designing model lipid systems of a desired resolution. We carried out microsecond-scale molecular dynamics (MD) simulations of large vesicles, highlighting the ability of the model to study systems at mesoscopic length and time scales. The models of DLPC, DOPC, and DOPC/DOPS have elastic properties consistent with experiment and structural properties such as the radial distribution functions (RDF), bond and angle distributions, and the *z*-density distributions that compare well with reference all-atom systems.



1. INTRODUCTION

Cells are the building blocks of multicellular organisms, and lipid bilayers serve as the membrane boundary that surrounds the cell and its organelles.[1−3] The membrane selectively controls the diffusion of ions, proteins, and other molecules into and out of the cell based on the cell's biological requirements.[4] Besides having a critical role in mechanical protection and compartmentalization, lipids also play a major role in cellular signaling used to trigger a host of activities in the living organism. An example of a biological signaling process is a synaptic transmission where neurotransmitters in synaptic vesicles mediate the communication across neurons.[5,6] Recently, a deeper understanding of the sensitivity of lipids to their surroundings and the signaling mechanisms of systems such as the G-protein coupled receptors[7] has led to important breakthroughs in drug design. Advanced experimental techniques such as neutron diffraction, electron microscopy, and atomic force microscopy[8−11] are frequently used to study lipid systems. Experiments have revealed a range of biological functions that lipids support and regulate and have also given insights into the mechanism involved during these processes.[12−14,11] Nonetheless, numerous membrane behaviors and the mechanisms underlying the biological functions are currently beyond experimental reach, especially in studies involving in vivo processes where it becomes increasingly difficult to control experimental conditions.[15,16]

Computational modeling and simulation techniques[17−20] are increasingly being used as an additional research tool to study biological membrane systems and enhance our understanding of the mechanisms involved at the atomic and cellular level. Modeling real biological membrane systems, made up of numerous types of lipids (including cholesterol) with variable compositions, is still challenging.[21,22] The time and length scales of membrane processes may range from nanoseconds[23,24] to seconds for vesicle formation and fusion. Some processes, such as pore formation[25,26] and ion permeation[27,28] through membranes, can be studied by simulating small bilayer patches, while others, such as cellular vesiculation, endocytosis, and similar cellular processes, require modeling of very large systems, including whole liposomes.[29,30] Mechanisms that involve computationally tractable length and time scales and that might require interpretation for results of atomistic resolution are usually modeled using the all-atom representation. However, it is computationally infeasible to simulate membrane processes, such as synaptic transmission and formation and fusion of cells,[31,1,32,7,33] with an all-atom representation, since the time and length scales extend to seconds and micrometers, respectively. For example, modeling a liposome 200 nm in diameter with an all-atom representation may entail propagating equations of motion for many millions of atoms. It would take months for such simulations to obtain a microsecond of sampling required to harvest statistically significant data with currently available computing resources. As such, coarse-grained (CG) simulations are increasingly being explored as viable alternatives for computational studies of membrane systems at experimentally relevant length and time scales.[34−36]

Received: August 30, 2012

   dx.doi.org/10.1021/ct300751h | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

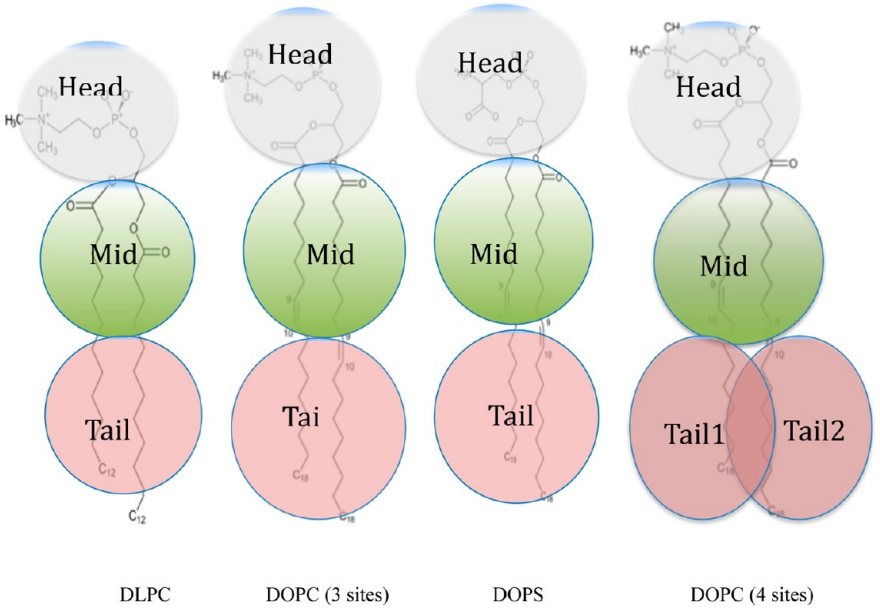


Figure 1. Coarse-grained representation of the HCG models of DLPC, DOPC, and DOPS lipid molecules described in this study.

Coarse-graining of molecular systems is mostly done by defining CG sites that are made up of chemically connected atoms.[35] The objective is to have as many atoms as is feasible in a given reduced unit to obtain a coarser representation of the system, thus minimizing the degrees of freedom considered. High frequency internal degrees of freedom, which are assumed not to greatly affect the physical properties of interest, are averaged out in the process. Several CG lipid models of lipids have been successfully developed in the recent years. The united atom models were the earliest versions of CG lipids where hydrogen atoms were merged with their backbone atoms.[37] Less resolved CG models, such as those involving four to six carbon atoms per CG site,[38,39] are known to give satisfactory global properties while maintaining local chemical specificity to a certain extent. However, the application of such models to study large time and length scale membrane phenomena is hindered due to an explicit CG representation of solvent molecules. Solvent-free models of lipid systems provide a significant leap in accessing large length and time scales. Based on the extent of the reduced representation, the solvent-free models may be categorized into "regular" CG models[38,40,39,41−44] and very "aggressive" CG models.[45,46,34,47,41,48,49] The earliest version of solvent-free lipid models[50,51] were represented with empirical pair potentials designed to mimic the hydrophobic interactions and maintain cohesion for a consistent bilayer morphology.

Due to the limitations of the traditional pair-potentials, a broad fluid bilayer phase is not always found in a solvent-free model.[46] Very aggressive solvent-free CG lipid models with three to ten sites per lipid have been recently developed, allowing the control of the properties of the fluid phase via tunable parameters.[45,46,48,49] In such models, membrane properties such as bending rigidity, area compression modulus, lateral diffusion coefficient, and flip-flop rate can be varied over relevant ranges by a careful selection of parameters. However, since the modeling schemes for these models are based on designing interaction parameters that can reproduce experimental observations, a systematic protocol to develop aggressive CG models for individual lipids from the underlying atomistic forces, with limited fitting of parameters to match experimental results, was not the focus of that work.

In this paper, we develop a hybrid approach to aggressively coarse-grain lipid systems. The method is based on the MS-CG methodology,[52−54] a variational method in which the data from all-atom simulations is used to determine the optimal CG force field for a chosen CG representation. The CG force field parametrization depends on a set of MD configurations and a predefined basis set for the CG force field. Lipid models with various degrees of coarse-graining have been successfully developed in the past with MS-CG using a cubic spline basis set[55,42] or a more accurate and flexible B-spline basis set.[43] The MS-CG method was also used to develop a single-site solvent-free hybrid analytical systematic (HAS) ellipsoidal lipid model[41] where anisotropic interactions were modeled with an enhanced Gay−Berne potential.[56] In this later work, the MS-CG approach was used to model the in-plane lipid center of mass interactions to maintain bilayer integrity.[41]

The hybrid coarse-grained (HCG) models presented here are instead full CG "bead" based models. They can reproduce the atomic-level RDFs and the *z*-density distributions and have elastic properties consistent with experimental data. Due to the aggressive level of coarse-graining inherent in the present work, however, limited atomistic-level sampling is obtained for the short-range (SR) interactions, a feature that gives rise to poor statistics for the MS-CG force-matching process in those regions. To surmount this challenge, a hybrid scheme is proposed in this work where the MS-CG method is supplemented with an analytical force field in the sparsely sampled SR configurational space in order to give the highly CG sites a proper repulsion and excluded volume. This methodology can then be used to systematically develop CG models with user-defined resolution, for individual and mixed lipid systems. We apply here the method to develop three-site solvent free models for lipid bilayers of 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC), 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), and a 1:1 mixture of 1,2-dioleoyl-*sn*-glycero-3-phospho-L-serine (DOPS) and DOPC. In addition, we

dx.doi.org/10.1021/ct300751h | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX

Table 1. Atoms from All-Atom Representation Associated with Each CG Site[a]

CG site	DOPC atoms	DLPC atoms	DOPS atoms	DOPC atoms (2 tails)
head	1−44 (44)	1−27 (27)	1−37 (37)	1−37 (37)
mid	45−66, 92−113 (44)	28−53, 76−84 (35)	38−59, 85−106 (44)	38−57, 85−104 (40)
tail	67−91, 114−138 (50)	54−75, 85−106 (44)	60−84, 107−131 (50)	58−84 (27), 105−131 (27)

[a]The number representing individual atom in all-atom representation can be obtained from ref 64.

develop a four-site DOPC model that illustrates the adjustable resolution capability of the model.

The remainder of this article is organized as follows: Section 2 briefly introduces the HCG scheme and explains the CG representations for DLPC, DOPC, and DOPS used in this work, while also reviewing briefly the theory behind the MS-CG method. The details about the SR analytical functional form and bonded interactions are also discussed. The features of the all-atom reference systems used in this work to obtain the MS-CG interactions parameters are covered in section 3, while results including the validation and application of the model for large flat bilayer and liposomes are discussed in section 4. Section 5 presents conclusions and future directions.

2. HCG METHOD

2.1. CG Representation. The DLPC, DOPC, and DOPS lipid molecules are reduced to three-site highly CG models, as shown in Figure 1. The solvent effects are effectively embedded in the CG force field interactions. We name the CG sites of the lipid molecule, head, mid, and tail, based on their topological location. The atoms from all-atom representation that constitute each CG site for the different lipids are listed in Table 1. The individual mass of each CG site is equal to the sum of masses of all the atoms constituting that CG site, and they are listed in Table 2.

Table 2. Mass of the CG Sites Used in the HCG Models[a]

CG site	mass DOPC sites (amu)	mass DLPC sites (amu)	mass DOPS sites (amu)
head	196.160	182.137	185.134
mid	311.440	157.147	311.440
tail	278.524	140.270	278.524

[a]Mass of each CG site is equal to the sum of masses of atoms that constitute the CG site, as given in Table 1.

While using the MS-CG procedure to develop highly CG models for lipid systems, a minimum number of CG sites were sought that could faithfully reproduce the structural and elastic properties. A cutoff of 4−5 Å was chosen for analytical form beyond which the CG interaction parameters were obtained completely from the MS-CG force-matching method. The lipids molecules could be successfully reduced to three-site models using this cutoff. The HCG model, while significantly reducing the number of degrees of freedom, thus retains the effective chain disorder. This is because the MS-CG force field represents the many-body potential of mean force between CG sites and hence its entropic components.[53,54,57]

Nevertheless, it is noteworthy that, in the three-site CG model, the two tails are combined as a single CG site. Lipid tails have a large amount of disorder and representing the two tails with a single CG site can lead to loss of accuracy in structural features like tail−tail RDFs function and may compromise the bilayer stability. To explore this, we also developed a four-site CG model of DOPC where each tail is represented with a CG site. A comparison between tail−tail RDF between three-site and four-site models highlights this issue and is described in the Results section. It is also shown later that careful selection of the SR analytical interaction parameters can lead to a stable bilayer with acceptable elastic properties even with a mismatched tail−tail RDF. The current scheme can be easily extended to four or more CG site models and for other lipids.

The three-site mapping of pure DLPC and DOPC bilayers resulted in six pairwise nonbonded, two bonded, and one angular interactions. For the mixed DOPC/DOPS bilayer system, the DOPC and DOPS are represented as separate head groups whereas the mid and tail groups are modeled with the same CG parameters. This results in four additional nonbonded interactions with a total of ten nonbonded pairs of interactions. For the four-site model of DOPC, there are two tails, modeled with the same interactions, leaving six pairs of nonbonded, two bonded interactions, and one angular interaction.

2.2. MS-CG Interaction Parameters. The details of the MS-CG method are provided elsewhere.[52−54,42,43] In short, the MS-CG method is based on a variational principle where the CG force field is represented as a linear combination of arbitrary functions of unknown variables. The forces obtained from this CG force field are matched in a least-squared sense to the all-atom MD simulation (or other sources) to arrive at a final set of MS-CG force field parameters. The matching between CG and atomistic forces minimizes the residual[52,47,53,54,42,43]

$$\chi^2 = \min\langle |F(\mathbf{R}^N; \phi) - \mathbf{f}(\mathbf{r}^n)|^2 \rangle \quad (1)$$

where $\mathbf{f}(\mathbf{r}^n)$ represents the reference "atomistic forces" projected on the CG sites calculated from the atomistic MD simulations and $\mathbf{F}(\mathbf{R}^N;\phi)$ represents the CG forces calculated based on the assumed arbitrary function. Here, $\mathbf{r}^n = \{\mathbf{r}^1,\mathbf{r}^2, ..., \mathbf{r}^n\}$ and $\mathbf{R}^N = \{\mathbf{R}^1, \mathbf{R}^2, ..., \mathbf{R}^n\}$ denote Cartesian spatial coordinates for the all-atom and CG representations, respectively. The predefined CG force field (called a basis set[53]) contains N_d parameters, $\phi = \{\phi_1, \phi_2, ..., \phi_{N_d}\}$, which are determined using the variational minimization of the residual shown.

The CG potential $U(\mathbf{R}^N;\phi)$ is obtained by integrating the CG force results from the MS-CG calculations, which is relatively straightforward if the nonbonded CG interactions are defined to be pairwise decomposable. It has been shown theoretically that the minimum value of the residual is obtained when $\mathbf{F}(\mathbf{R}^N; \phi)$ is the gradient of the exact many-body potential of mean force, $U(\mathbf{R}^N)$, governing the CG variables $\mathbf{R}^N$.[58,59,53,54,60,61] However, in practice, the optimal $U(\mathbf{R}^N; \phi)$ function is only an approximation to the exact $U(\mathbf{R}^N)$ because the basis set for the assumed force-field is incomplete. The degree of sampling in the underlying atomistic MD simulation can also affect the quality of the result, especially in an aggressive CG model such as the one developed in this work. As a result, as is discussed in the next section, an analytical component in the interaction potential can be incorporated to overcome the incomplete MS-CG sampling in the short-range.

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

2.3. Analytical Form for Less Sampled Short-Range Interactions. While using the MS-CG methodology, we found that statistically insufficient CG force field parameters were obtained in the short-range, usually for R_{IJ} < 5 Å. Figure 2

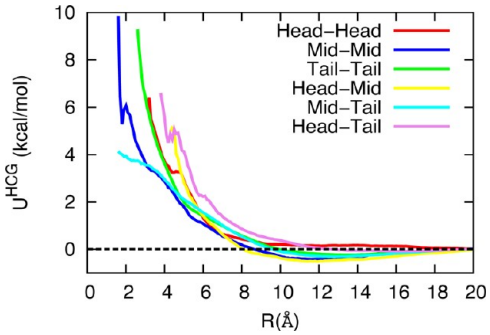


Figure 2. MS-CG-derived pairwise interaction between CG sites in the DOPC 3-site model. MS-CG provides an accurate set of CG interactions for well-sampled regions of phase space. Due to poor sampling in the short-range, the MS-CG method does not guarantee accurate interactions at close range. This is particularly true for highly CG models with a large number of atoms coarse-grained into a single site as shown in Figure 1. An analytical interaction potential with user-selected parameters is used to represent the short-ranged interaction.

illustrates this point, as the MS-CG derived nonbonded interactions for all the pairs in the three-sites DOPC lipid model are shown. As is evident in Figure 2, the MS-CG interactions have rather poor sampling in the SR region.

An analytical form was thus used to obtain the interactions in the sparsely sampled SR domain. The analytical function selected for the short-range interaction was of the form,

$$U_{SR}(R_{IJ}) = 4\varepsilon\left(\left(\frac{\sigma}{R_{IJ}}\right)^{12} - \left(\frac{\sigma}{R_{IJ}}\right)^{6}\right) \text{ for all } R_{IJ} \leq R_{SR} \tag{2}$$

where ε, σ, and R_{SR} are the three parameters that define the short-range interaction analytically. The constants ε and σ are the standard Lennard-Jones parameters; R_{SR} is the higher cutoff for the analytical part, and it is kept to the minimum permissible value (generally <5 Å) such that the bulk of the interaction space is defined by the MS-CG interactions. An informed set of values are chosen as initial guess for the fitting parameters for eq 2 to make the process of obtaining the analytical form more systematic. The ε and σ parameters are chosen such that the interactions obtained from analytical form and the MS-CG are C1 continuous around the lowest available range of MS-CG interaction data. Once all the interactions are defined (including the bonded interactions explained in next section), the three parameters are adjusted by running CG simulations on a small bilayer patch (~50 Å × 50 Å). A zero-surface tension boundary condition is maintained by coupling the CG bilayer with an anisotropic barostat in the x and y dimensions and keeping the z-dimension of the simulation box fixed. The area per lipid (A_L), fluctuations in the area per lipid, and the thickness of the CG bilayer patch are monitored for 20 million time steps with an integration time step of 50 fs. No external constraint with respect to A_L or K_A is applied to the CG system. However, if a stable bilayer is not formed or physical properties such as A_L and K_A of the CG system are not comparable to AA data, the user defined parameter R_{SR} is varied and the other two parameters (σ and ε) are recalculated by fitting the MS-CG data to eq 2, starting from the new R_{SR}. The final R_{SR} parameters are listed in Table S4 in the Supporting Information. The preliminary estimates on the elastic properties such as the area compressibility modulus and bending modulus can be made from the knowledge of the above observables. If the A_L, fluctuations in A_L, bilayer thickness, and the RDFs are within the range of values observed in the all-atom simulations, the new HCG parameters are tested on large bilayer patch (40 nm × 40 nm) systems. The parameters are finalized once the large flat bilayer is found stable for very long simulations (~40 million CG time steps). The steps involved in the obtaining the full HCG parameters are elucidated using a flowchart diagram in Figure 3.

2.4. Bonded Interactions. The Inverse Boltzmann (IB) method[62,63] was used to obtain the CG force field for the bonded interactions. The bond and angle probability distributions were sampled from the atomistic MD trajectory. Given the probability distributions, the CG bond potentials $U^{CG}_{Bond}(R)$ and the angular potential $U^{CG}_{Angle}(\theta)$ are derived from the IB relations as

$$U^{CG}_{Bond}(R) = -k_B T \ln P^{CG}_{Bond}(R)$$
$$U^{CG}_{Angle}(\theta) = -k_B T \ln P^{CG}_{Bond}(\theta) \tag{3}$$

where R and θ are bond-length and bong angle respectively, k_B is the Boltzmann's constant, and T is the absolute temperature. The bond and angle probability distributions are represented as $P^{CG}_{Bond}(R)$ and $P^{CG}_{Angle}(\theta)$, respectively, and are normalized by R^2 for the bond length and $\sin(\theta)$ for the angles. Since the reference probability distributions were not Gaussian, the bonded and angular interactions were represented as tabulated potentials instead of the harmonic approximation.

The force-field parameters (both bonded and nonbonded) for the HCG models of all the lipids are provided as tables in the Supporting Information.

3. BUILDING THE SOLVENT-FREE HCG MODEL

Broadly speaking, the steps involved in the construction of a HCG model for a given lipid system are enumerated below:

1. Generate an all-atom reference trajectory for the given lipid system. The details of the reference all-atom MD simulations used in this work are discussed in section 3.1.
2. Choose the CG representation of the lipids (three sites, four sites, etc.) and obtain the effective reference CG coordinates and forces for each configuration from the reference all-atom trajectory.
3. Find the probability distribution for the CG bonds and angles from the reference CG coordinates and obtain the bond and angle-bending interaction parameters using the IB method, as discussed in section 2.4.
4. Apply the MS-CG based force-matching algorithm on the reference CG coordinates and forces to obtain the MS-CG nonbonded parameters for all possible pair interactions. The details of the MS-CG parametrization are discussed in section 3.2.
5. The portion of each SR nonbonded interaction that is not available from MS-CG is next supplemented with an analytical function using the adjustable parameters ε, σ, and R_{SR}, as discussed in section 2.3. These parameters are then tested for large bilayer patches (~40 nm × 40 nm) to ensure the stability of large bilayer systems over long time scales (~100 ns).

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

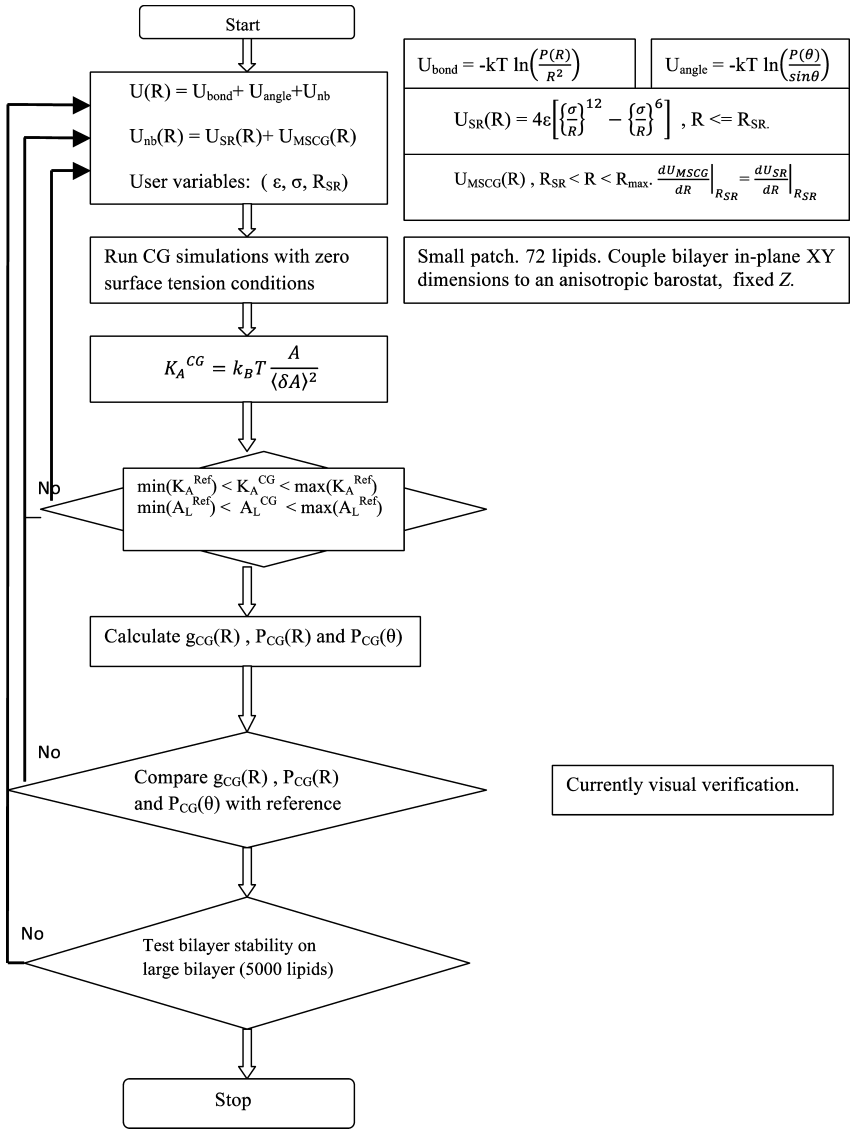


Figure 3. Flowchart showing the steps involved in the HCG modeling protocol.

It should be noted that fitting the MS-CG data to simple potential functions like Lennard-Jones or Morse potential without going through the above-mentioned protocol does not guarantee a stable bilayer. Such a simpler protocol would be desirable but was seen to fail. The MS-CG derived data for the DOPC lipid system was fitted to a Morse potential and CG simulations were carried out for a 5000 lipid bilayer system. MS-CG nonbonded pair interaction data for 3-sites DOPC lipid fitted to Morse potential is shown in Figure S1, Supporting Information. The fitted Morse potential parameter values are provided in Table S5, Supporting Information. The bilayer structure is lost within few nanoseconds of simulation time. This shows that a simple parametrization of a global nonbonded function such as Morse potential to the forces in the force matching may not always work well. Carefully designed HCG parameters are needed that uniquely define the interactions that keep the bilayer morphology intact and reproduce the structural and physical properties well.

3.1. Reference All-Atom MD Simulations. The reference all-atom MD system for DLPC consisted of a bilayer with 50 DLPC lipid molecules and 1500 water molecules. A total of 72 DOPC lipid molecules with 2100 water molecules were used to generate the all-atom trajectory for the HCG modeling of DOPC. The mixed DOPC/DOPS system consisted of 36 DOPC and 36 DOPS lipid with 2100 water molecules, and the system was neutralized using Na^+ and Cl^-counterions. The physiological salt concentration of 0.3 M was used for all systems. All three systems had equal composition in each layer. It should be noted that the parametrization process is composition dependent and has to be repeated if the type of lipid or ratio of lipids in case of a mixed lipid system is changed. The goal of the present work is to develop highly coarse-grained models that can be greatly scaled up in system size.

The initial coordinates for the lipid bilayer for DLPC and DOPC were obtained from an existing all-atom simulation.[64] For the mixed DOPC/DOPS system, half of the DOPC molecules in each layer were replaced with DOPS molecules. The bilayer systems were then solvated before placing them in a periodic cell. The ratio of number of water molecules to number of lipid molecules was chosen to be around 30:1.[65] CHARMM36 force field parameters[64] were used for lipid molecules while water molecules and counterions were simulated using the TIP3P force-field.[66] The cutoff for nonbonded van der Waals interactions was 12 Å, and no

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

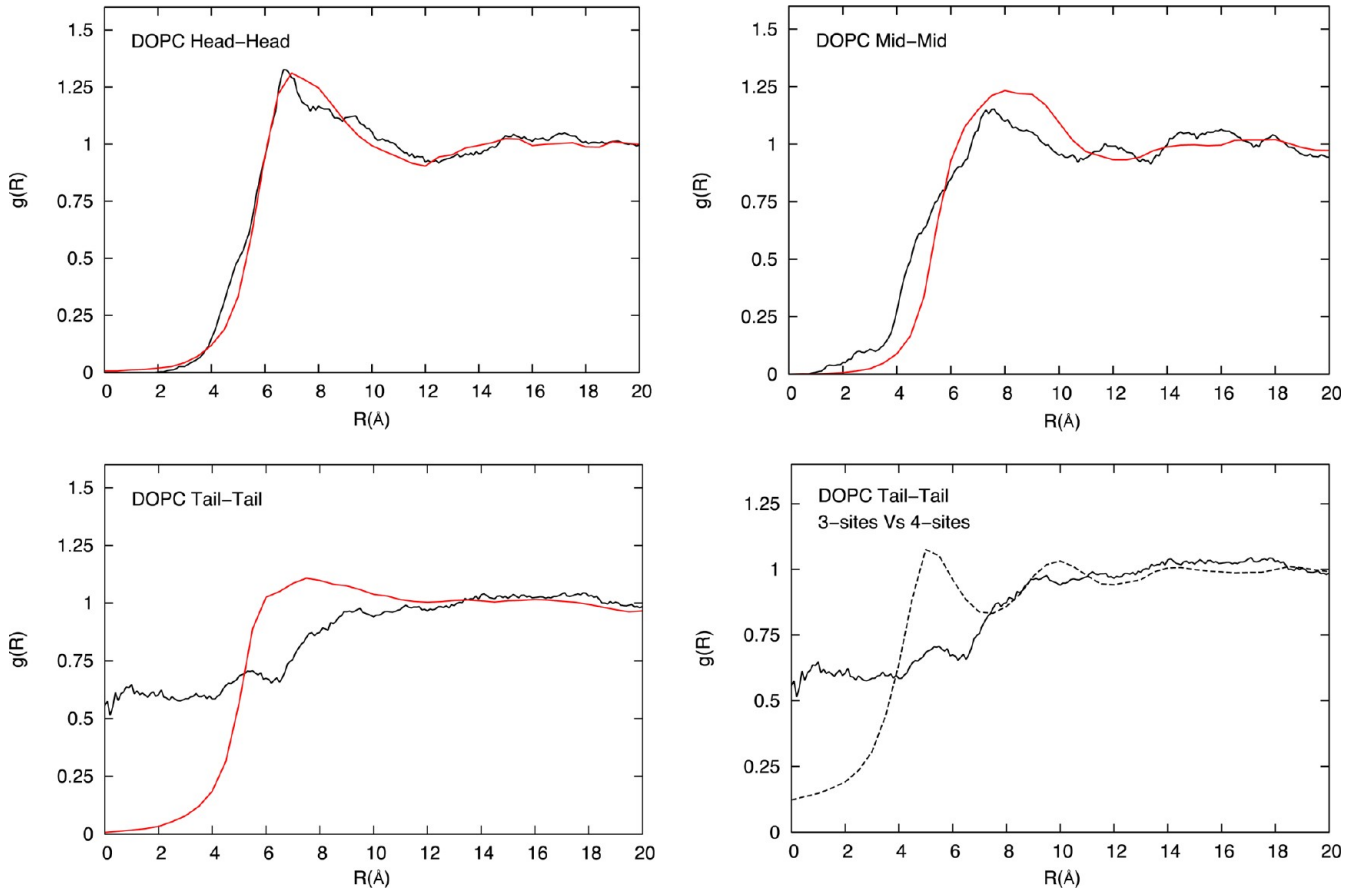


Figure 4. (a) Comparison of head−head RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC lipid system. (b) Mid−mid RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC lipid system. (c) Tail−tail RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC lipid system. (d) Tail−tail RDFs from atomistic simulation with one tail CG representation (solid line) and two tails CG representation (dashed line) for the DOPC lipid system.

long-range correction was imposed. A LJ switching function over 8 to 12 Å was used in the MD simulations and the long-range electrostatic interactions were calculated via the particle mesh Ewald summation.[67] An integration time step of 2 fs was used, and the SHAKE algorithm[68] was applied to constrain all C−H bonds. After an initial energy minimization, each bilayer system was run for about 40 ns in the constant NPT ensemble using the Nose-Hoover thermostat and Langevin piston at 1.01 bar and 310 K.

The average cell dimension in the last 20 ns of NPT run for DLPC was 39.72 Å × 39.72 Å × 58.84 Å. The average value of A_L was 63.12 Å^2 for the DLPC lipid bilayer system. The average A_L for DOPC was found to be 68.20 Å^2. These data are consistent with previous results from the CHARMM36 force field and experimental data.[69] The average A_L for mixed DOPC/DOPS lipid from all atom system was found to be 65.8 Å^2. The A_L for the mixed lipid system is less than the pure DOPC lipid system. This may appear counterintuitive because of the high probability of electrostatic repulsion taking place between the anionic phosphoserine (PS) head groups present in the system. However, recent all-atom simulations and experiments[70,71] have provided conflicting evidence with respect to A_L for mixed anionic lipids. The presence of anionic head groups in mixed lipid systems can lead to increased intra- and intermolecular hydrogen bonding and strong ion-lipid interactions, which may reduce the effective A_L of the system.[70] The A_L may also depend on the salt concentration as shown in ref 72. In the absence of any definitive reference data for area per lipid of DOPC/DOPS system, the converged data from the all-atom simulations were used for all future studies. Once A_L for an individual system converged in the NPT simulations, the different systems were run under the constant NVT ensemble for the next 60 ns. The last 40 ns of the constant NVT ensemble trajectory was used to sample reference forces and coordinates for the MS-CG calculations.

All the simulations were performed with the parallel MD program LAMMPS.[73] Visualizations of the trajectory were performed with VMD,[74] and analysis was performed using LAMMPS and GROMACS.[75]

3.2. Details of the Force-Matching. The effective reference CG coordinates and forces for each configuration were calculated from the all-atom trajectory. The forces due to water molecules were incorporated into the effective force calculation of the CG sites. Atomic coordinates and forces were sampled every 1 ps for a total of 40 000 MD frames collected as input for the MS-CG procedure. B-Spline basis functions were used to interpolate the MS-CG forces. For B-Spline-based parametrization details, the reader is referred to the Methods section in ref 43. The final CG force-table results were then interpolated and integrated to obtain tabular pair potential interactions for each different lipid model. The SR nonbonded interaction parameters were obtained, as discussed in section 2.3. The full interaction parameters (including the SR

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

interactions) for all the three lipid models are given in the Supporting Information.

4. SIMULATION RESULTS AND DISCUSSION

Three sets of CG simulations results are presented for each lipid system in this section. In the first set of CG simulations results, individual models for DLPC, DOPC, and mixed DOPC/DOPS are validated against the corresponding reference systems by comparing RDFs, bond and angle distribution, and the *z*-density profile. Mean squared displacement (MSD) for DLPC is calculated and compared with the all-atom system to show the fluid nature of the HCG model. The second set of results is obtained by simulating a large flat membrane of 5000 lipids (~40 nm × 40 nm in area) and used primarily to measure important elastic properties such as the area compressibility modulus and the bending modulus. Undulation spectra and the lateral pressure distribution are also calculated and shown for the DLPC HCG model. Lastly, the HCG models are tested with large liposome simulations (~120 nm). The liposome's resilience to deformation is also tested to confirm model's viability for large-scale simulations.

4.1. Model Validation. The CG systems were created on the basis of the reference systems discussed in section 3.1. The CG systems for DLPC and DOPC consisted of 50 and 72 lipids, respectively. The mixed DOPC/DOPS system had 72 lipids with 1:1 composition. All CG simulations were carried out for up to 200 ns CG simulation time in the constant NVT ensemble for direct comparison with the reference system. All simulation conditions were the same as those described in section 3.1 for the all-atom reference system. The structural properties were calculated for the CG simulation and validated with corresponding data from all-atom simulation.

4.1.1. RDFs. The two-dimensional RDF plots for the selected sites for DOPC, DLPC, and mixed DOPC/DOPS are shown in Figures 4, 5, and 6, respectively. Considering the ultra low-

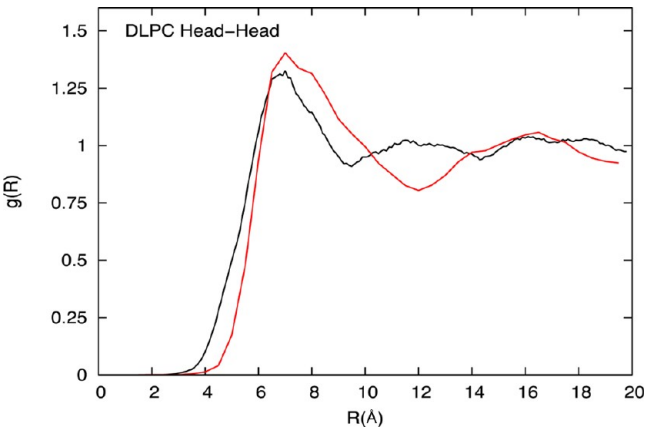


Figure 5. Comparison of head–head RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DLPC lipid system.

resolution of the CG models in this work, the RDF plots compare reasonably well with the all-atom MD results for the chosen site–site interactions. The HCG model is based on the MS-CG method that can, in principle, capture the many-body free-energy landscape accurately,[53,54,57] and the agreements in RDFs indicate the underlying atomistic origin of the HCG model. The RDFs, particularly for the head–head pairs, have good agreement with the all-atom RDFs, as shown in Figures 4a, 5, and 6 for DOPC, DLPC, and mixed DOPC/DOPS lipid bilayers, respectively. Unlike the DOPC and DLPC HCG model, there are two types of head groups in mixed DOPC/DOPS HCG model, and the RDF plots between the different types of the headgroup (PC–PC, PC–PS, and PS–PS) are shown in Figure 6b–d. RDF plots for all the head pairs compare well with the all-atom MD reference results. It is worth noting that the RDF for the PS–PS head groups is more structured (and has a higher peak value) than that for all other head–head distributions in this study. This suggests that the DOPC and DOPS lipids are not homogeneously distributed, and the DOPS lipids tend to aggregate more than the DOPC although there is no evidence of isolated clusters being formed. This may also explain the slightly reduced A_L observed in mixed lipid system, which may be caused due to the presence of anionic head groups giving rise to increased hydrogen bonding, strong ion-lipid interactions, and higher possibility of salt bridges.[72,70]

The HCG model predictions for the mid and tail sites, which represent the more disordered region of the bilayer, are found to differ more greatly from the corresponding all-atom pair correlation function. This is due to the high lipid chain-packing irregularities and increased chain flexibility away from the head groups. While a limited deviation is found in the RDF plots for mid–mid pair in DOPC, as shown in Figure 4b, there are particularly large discrepancies for the tail–tail pair-correlation, as shown in Figure 3c. This discrepancy is especially pronounced in the SR region.

It is to be expected that the native structural correlation will be lost to a certain extent in an aggressively CG model especially for the sites that represent sections of two different tails with a single CG site. As shown in Figure 4c, the tail–tail all-atom 2-dimensional RDF function has nonzero values (~0.5–0.6) in the SR of 0–6 Å. This is clearly an artifact of representing two distinct tails, which are highly flexible and sometimes intertwining, with a single CG site. This causes the centroid positions of two CG tail sites from different lipids in the same layer to overlap with each other, especially when projected on the two-dimensional *xy* plane, as was done to calculate the two-dimensional RDF. Figure 4d highlights this effect where the all-atom RDF plots for tail–tail pair in three-site HCG model for DOPC are compared with the tail–tail pair of four-site DOPC model with two CG tail sites. The overlap reduces with four-site models, and adding more sites could improve this further, although at the cost of overall computational efficiency.

In the three-site HCG model, accuracy in the tail–tail RDF is traded for bilayer stability. We found that when the interaction for the SR was kept very soft, to better reproduce the reference all-atom RDF data, the stability of the bilayer is compromised, and the bilayer tends to aggregate in a micelle-like structure due to strong effective interaction between tail groups. To have a stable bilayer, the range of repulsive interaction for tail–tail interaction must be expanded to avoid strong artificial cohesion at short ranges, which leads to micelle-like structure.

The aggressive nature of the HCG model is a primary reason for the mismatch in pair-correlation between the all-atom MD simulation and the CG simulation data. Additionally, certain features of the MS-CG method may also influence the RDFs. Since the MS-CG uses the pair wise interaction approximation, many body interactions are usually omitted in the parametrization process, although this is not always the case.[61] This assumption especially affects solvent-free CG models. More-

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

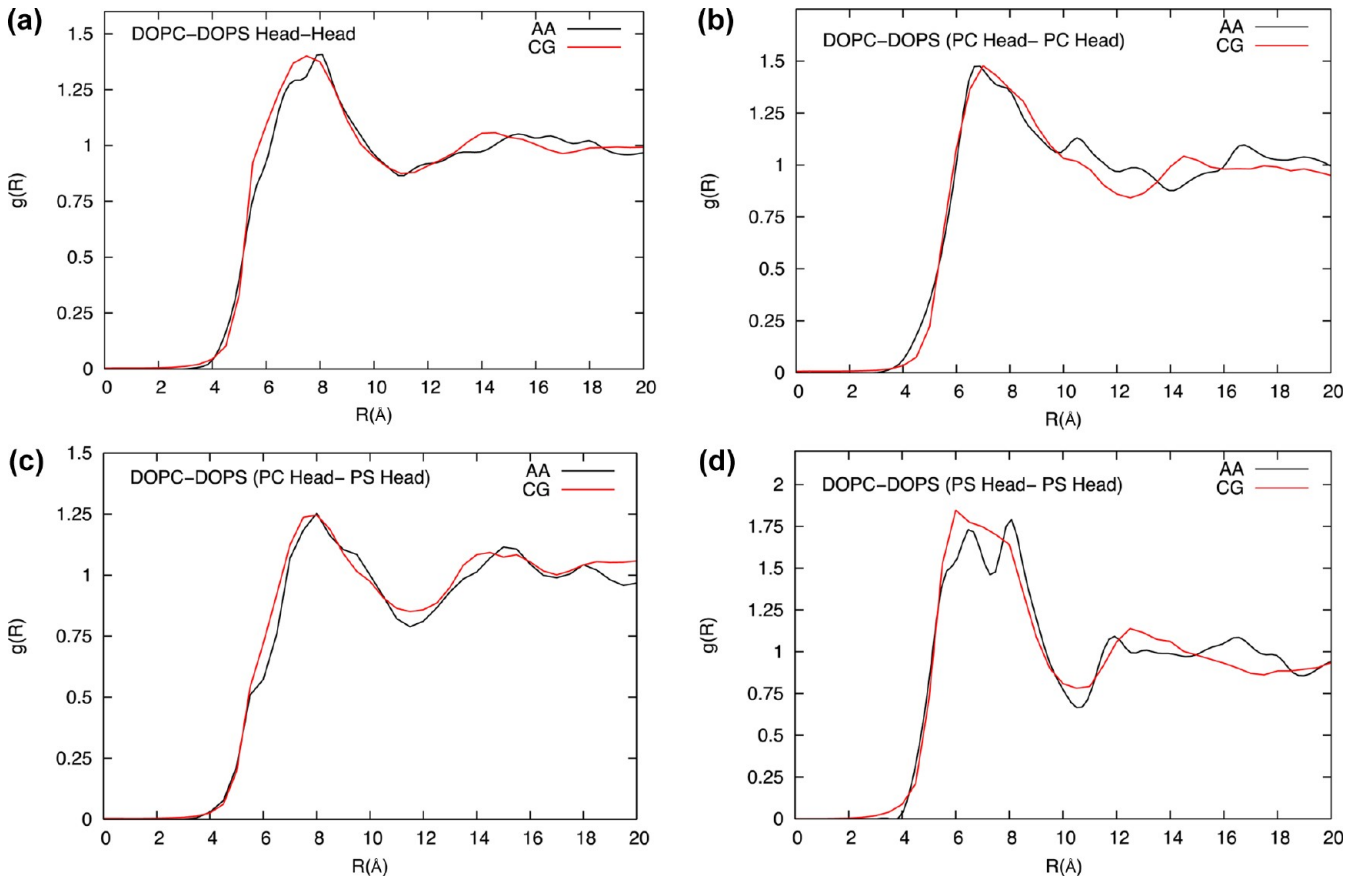


Figure 6. (a) Comparison of head–head RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC/DOPS lipid system. (b) Head (PC)–head (PC) RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC/DOPS lipid system. (c) Head (PC)–head (PS) RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC/DOPS lipid system. (d) Head (PS)–head (PS) RDFs from atomistic simulation (black) and solvent-free HCG simulation (red) for DOPC/DOPS lipid system.

over, the minimization process in the MS-CG procedure is based on a least-squares method that is dependent on the available sampling.[43]

4.1.2. z-Density Distribution. The number density distribution of individual sites in DOPC, DLPC, and DOPC/DOPS are shown in Figure 7 as a function of the coordinates normal to the bilayer. This *z*-density profile shows good agreement with the all-atom trajectory. The *z*-density profile can also be used to estimate the bilayer thickness. It is seen that DOPC, which has longer tails (18-carbons per tail), is slightly thicker than DLPC (12-carbons per tail). The agreement with all-atom data implies that the bilayer thickness for the CG system is in the same range as that observed in the all-atom simulations.

4.1.3. Bond Probability Distribution. The CG bond length probability distributions for HCG DOPC, DLPC, and mixed DOPC/DOPS lipid models are compared with the corresponding all-atom MD system in Figure 8 and show good agreement with the reference all-atom distributions. It is observed that the longer tailed DOPC have a higher mean bond length than the DLPC lipid model. In addition, the DOPC lipids have broader bond distribution due to the presence of unsaturated bonds. In general, the mid–tail bond length distribution for the all-atom system is broader than the distribution obtained from the HCG model, which may be attributed to the strong cohesion between tail groups from the two leaflets. This cohesive force is critical in the formation of the condensed fluid phase for the HCG lipid model, especially in the absence of water that is a known source of cohesion for lipid systems. In this work, an imperfect bond length and angle distribution is acceptable in exchange for a stable fluid bilayer. To form a fluid phase bilayer system, the interaction potential must allow sufficient cohesion while retaining enough disorder to prevent gelation or solidification. The advantage of the HCG model is that such interactions arise naturally out of the MS-CG methodology.

4.1.4. Two-Dimensional MSD. The two-dimensional MSD vs time plot for the center of mass for the DLPC HCG model is shown in Figure 9. Although this CG model is not designed to reproduce the exact dynamics, MSD calculations are carried out to show the fluid nature of the HCG model. The MSD plot represents the mobility of the lipid molecules in the plane of the lipid bilayer sheet. The lateral diffusion constant of the lipid is measured to be 1.05×10^{-7} cm^2/s for DLPC from the MSD data. The linearity of MSD is tested up to 200 ns to verify that the CG system acts as a fluid over long time scales. The experimental value of the two-dimensional diffusion constant for the DLPC is $(3.0 \pm 0.6) \times 10^{-8}$ cm^2/s.[69,76] Clearly, the dynamics in the CG system is much faster than experimental value and the all-atom MD results, as expected.[55] This is consistent behavior in a CG system where the increased diffusion is attributed to the loss of degree of freedom leading to a smoother free energy landscape. Accurate diffusion in CG models is not expected without either rescaling time or using a more general dynamics with proper frictional terms.[77−80]

4.2. Large Flat Bilayer (5000 lipids, 38 nm × 38 nm system). The second set of CG simulations was carried out with 5000 lipids, consisting of 15 000 total CG sites for DLPC,

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

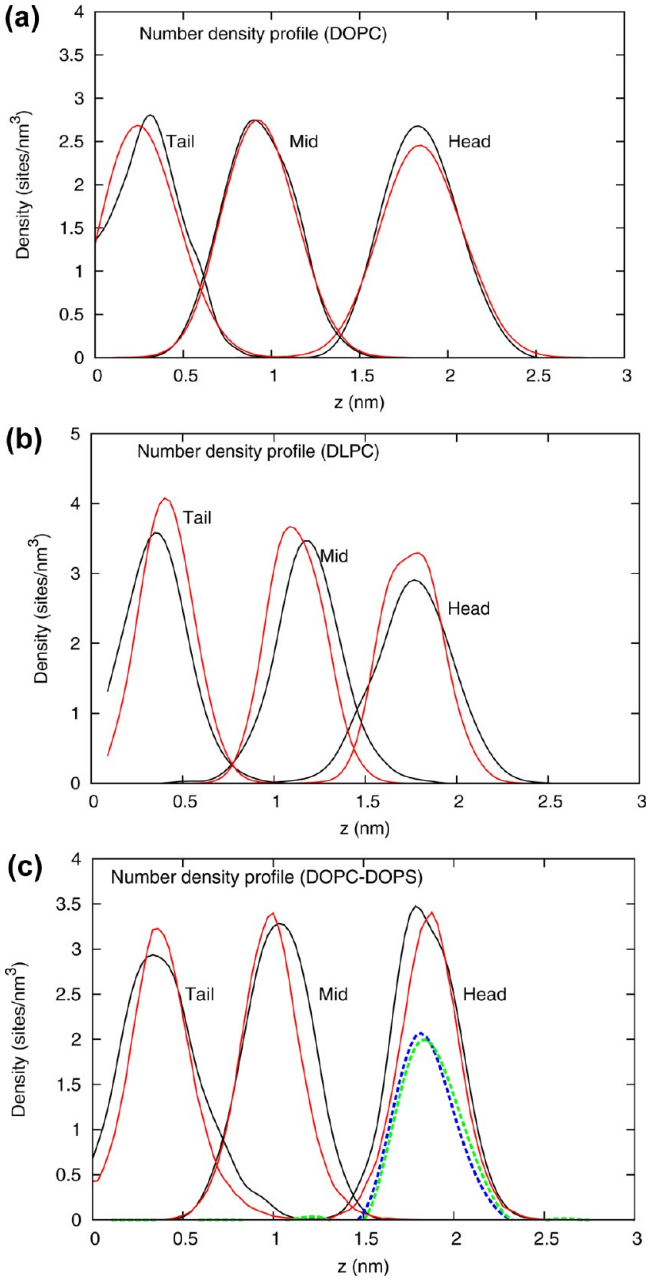


Figure 7. (a) z-Density profile for three CG sites from the atomistic simulation (black) and the CG simulation (red) of the DOPC HCG model. (b) z-Density profile for three CG sites from the atomistic simulation (black) and the CG simulation (red) of the DLPC HCG model. (c) z-Density profile for three CG sites from the atomistic simulation (black) and the CG simulation (red) of the DOPC/DOPS HCG model. The blue and green dashed lines represent the z-density profiles of the DOPC and DOPS headgroups, respectively. No heterogeneity with respect to distribution in z-dimension is observed.

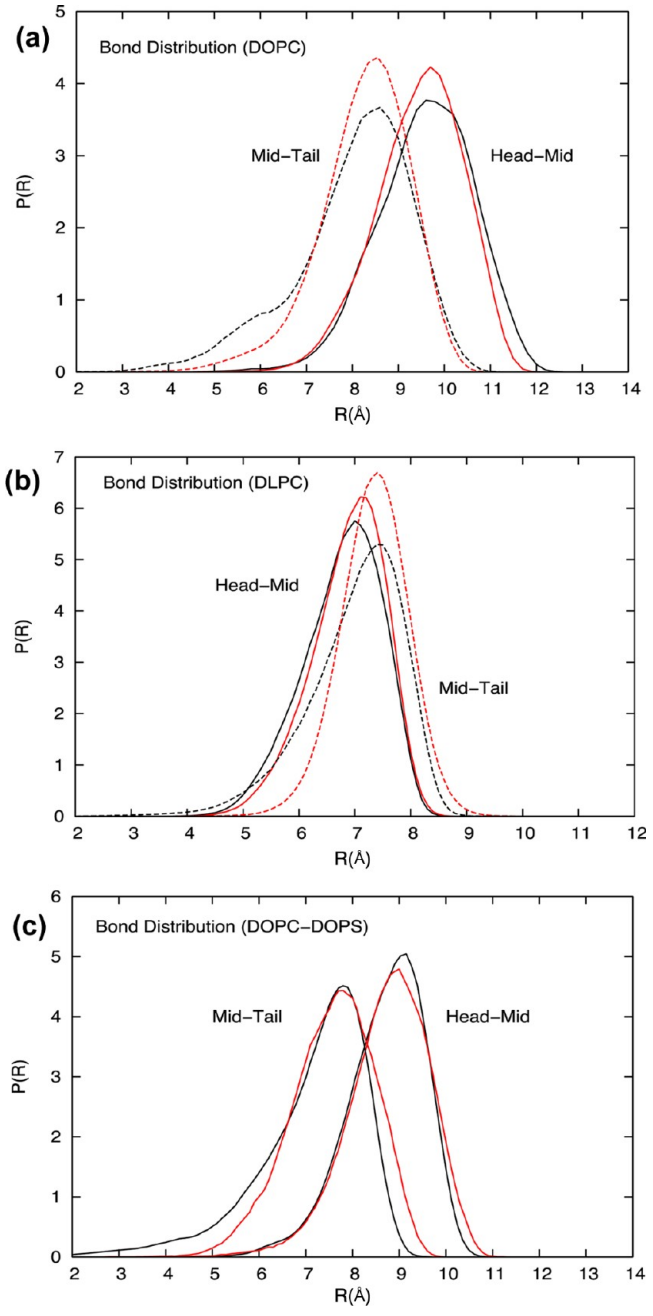


Figure 8. (a) Comparison of bond probability distributions from the atomistic simulation and the CG simulation of DOPC HCG model. (b) Bond probability distributions from the atomistic simulation and the CG simulation of DLPC HCG model. (c) Bond probability distributions from the atomistic simulation and the CG simulation of DOPC/DOPS HCG model.

DOPC, and DOPC/DOPS membranes. The scale of the coarse-graining can be gauged by comparing the system size with respect to an equivalent all-atom system. For example, an equivalent membrane system with DLPC lipids would consist of 530 000 DLPC atoms and a total of approximately one million atoms including the water molecules. Figure 10a shows the top and side view of the DLPC flat bilayer after 2 μs of CG simulation. The CG simulations were performed for 40 million 50 fs time steps at 310 K and zero surface tension.

4.2.1. Area Compressibility Modulus. The value of A_L for DLPC, DOPC, and mixed DOPC/DOPS HCG models was found to be 57.63 Å^2, 66.2 Å^2, and 61.4 Å^2, respectively. The values are slightly less than that observed in the all-atom systems, which are 58.20 Å^2 and 67.4 Å^2 for DLPC and DOPC, respectively. CG systems tend toward a smaller A_L than the corresponding all-atom system because of the softer effective potential energy landscape. Moreover, the A_L in CG simulation is calculated by taking the average area of the simulation box with respect to lipids in one leaflet rather than the actual surface

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

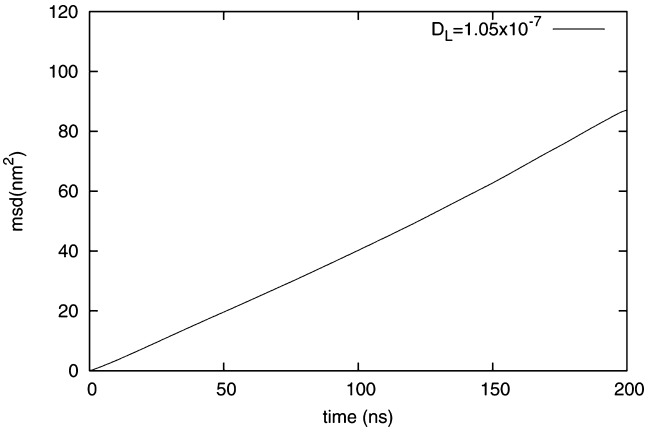


Figure 9. Mean-squared displacement of DLPC molecules calculated using center of mass diffusion. The experimental value of lateral diffusion constant D_L for DLPC is 3×10^{-8} ($\pm$ 20%) cm^2/s.

area of the lipid. Some decrease in A_L can also be attributed to the undulation modes observed in large systems.[81]

The area compressibility K_A for the CG lipid system was calculated using the fluctuation formula:

$$K_A = k_B T \frac{\langle A_L \rangle_{N\gamma T}}{\frac{n_L}{2}(\langle A_L^2 \rangle_{N\gamma T} - \langle \bar{A}_L \rangle_{N\gamma T}^2)} \quad (4)$$

where $\bar{A}_L$ is the average A_L obtained under zero surface tension conditions and n_L is the number of lipids in the system. Figure 11 shows the fluctuations in A_L in the three-site DOPC HCG model for 400 ns. The area compressibility moduli for the three lipid systems studied in this paper are reported in Table 3. A comparison with reference experimental and all-atom simu-

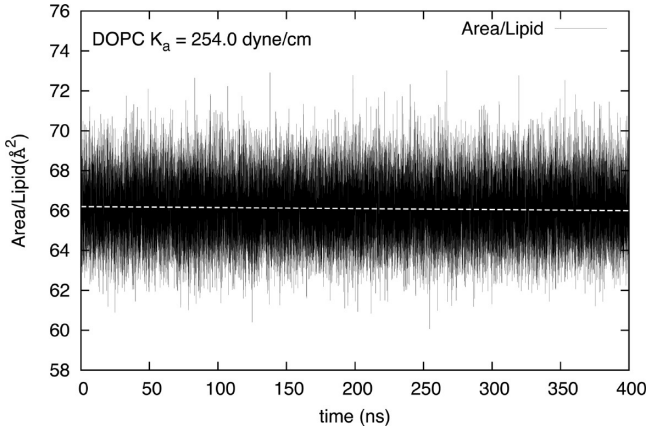


Figure 11. Area per lipid variation with time in CG simulations with DOPC HCG model. The compressibility modulus is calculated using the average area per lipid and its fluctuations.

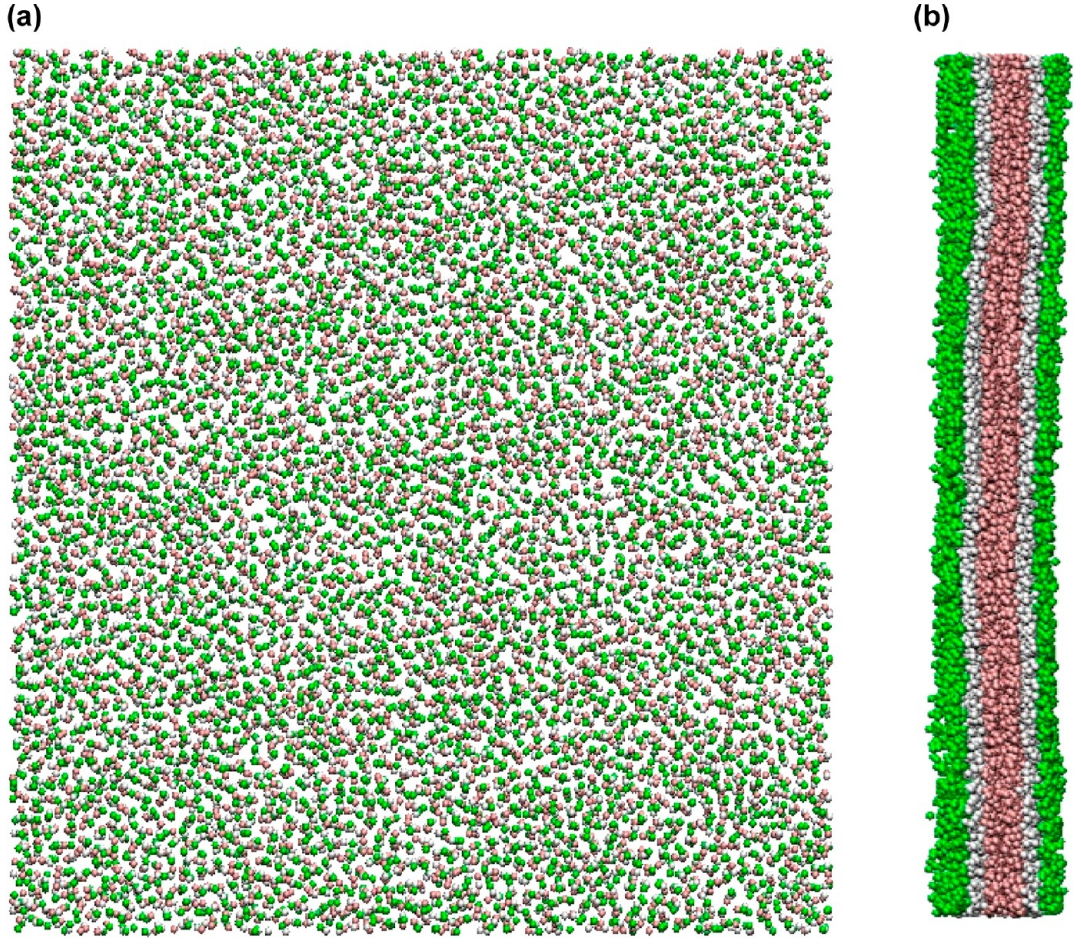


Figure 10. Snapshot of the DLPC membrane simulated using solvent-free HCG methodology. The system size is 38 nm × 38 nm consisting of 5000 lipid molecules simulated for 200 ns (a) Top view. No structural correlation is visible. (b) Side view. Thermal undulations modes are clearly visible.

Table 3. Area Compressibility Modulus of the Three HCG Models and Comparison with Reference Data

K_A (dyn/cm)	DLPC	DOPC	DOPC/DOPS
HCG model	310.0 ± 6.0	254.0 ± 9.0	165.0 ± 12.0
experiment/all-atom MD	461 ± 92 (ref 64)	188.0−265.0 (refs 82 and 83)	

lation data is also made. The DOPC HCG model compares very well the experimental data.[82,83] The area compressibility modulus of DLPC has a value of 310.42 dyn/cm for the HCG model and is comparable to the DOPC model. This can be assumed to be a reasonable value since K_A is known to be very weakly dependent on membrane thickness and does not change much with chain length and level of saturation.[83] In absence of any definitive experimental data, it is compared with a recent all-atom simulation result (461 (±92) dyn/cm)[84] that is known to overestimate the K_A.[64] A much different value of K_A (165.0 dyn/cm) is calculated for the mixed DOPC/DOPS lipid system. The reduction in K_A is consistent with a recent experimental calculation that determines the mechanical properties (including K_A) of DOPC/DOPS liposomes using a micropipet-aspiration techniques[85] where the area compressibility modulus was found to be around 67 dyn/cm. The A_L in mixed DOPC/DOPS HCG system has higher fluctuations than those observed in pure DOPC or DLPC HCG models, which results in a large reduction in area compressibility, even though A_L is approximately similar to that seen in pure lipids. Bilayers with multiple lipid types have a higher propensity for spontaneous curvature due to compositional heterogeneities,[86] which may explain the higher flexibility of the mixed DOPC/DOPS HCG model.

4.2.2. Bending Modulus. The bending modulus is a critical membrane material property that can be defined as the resistance of a membrane deforming from its existing curvature to some other curvature. This rigidity is very important in studies involving cellular level mechanisms such as curvature inducing protein−membrane interactions, curvature mediated vesiculations, and cell division and fusion.[87,6,88,89] In this work, the bending modulus is calculated from membrane undulation spectrum using the Helfrich model.[90,91] The bilayer is assumed to be a surface $u(x,y)$ given as[91]

$$u(\mathbf{r}) = u(x, y) = \frac{1}{2}(z_{top}(x, y) + z_{bot}(x, y)) \quad (5)$$

where $r = (x,y)$ is the two-dimensional real space vector and z_{top} and z_{bot} are z-position of top and bottom leaflet, respectively, at any given r. The Fourier mode of the undulations height at any given time is denoted as $u(\mathbf{q})$ where $\mathbf{q} = (q_x, q_y)$ are the two-dimensional reciprocal space vectors and have the following relationship with $u(\mathbf{r})$.

$$u(\mathbf{r}) = \sum_{\mathbf{q}} u(\mathbf{q})\, e^{i\mathbf{qr}} \quad (6)$$

Based on the Helfrich model, the fluctuation spectra $S_u(q)$ can be expressed in the Fourier space as[92,91]

$$S_u(q) \equiv \frac{N}{2}\langle |u(q)|^2 \rangle = \frac{k_B T}{A_L(k_c q^4 + \gamma q^2)} = \frac{k_B T}{A_L k_c q^4} at \gamma = 0 \quad (7)$$

where N is the number of lipids, k_c is the bending modulus, and γ is the surface tension. Figure 12a shows the full fluctuation spectra of the HCG DLPC system in the large wavelength regime. The bending modulus is calculated by fitting the data to $\mathbf{q}^{-4}$. The value of bending modulus for the DLPC HCG model was found to be 6.65(±0.317) × 10^{-20} J, which is approximately equal to 15 k_BT. In general, the bending modulus of bilayer range[89] from a few k_BT to tens of k_BT and the predicted modulus for the model is within the range of the experimental value of 5.5 × 10^{-20} J.[69,76,84] The bending modulus for DOPC HCG model was estimated to be around (7.64 ± 0.671) × 10^{-20} (or ∼18−19 k_BT). This is in good agreement with the experimental data for DOPC where a value of 18.8 k_BT was reported using X-ray scattering[93] and a value of 21 k_BT was obtained from pipet aspirations.[83]

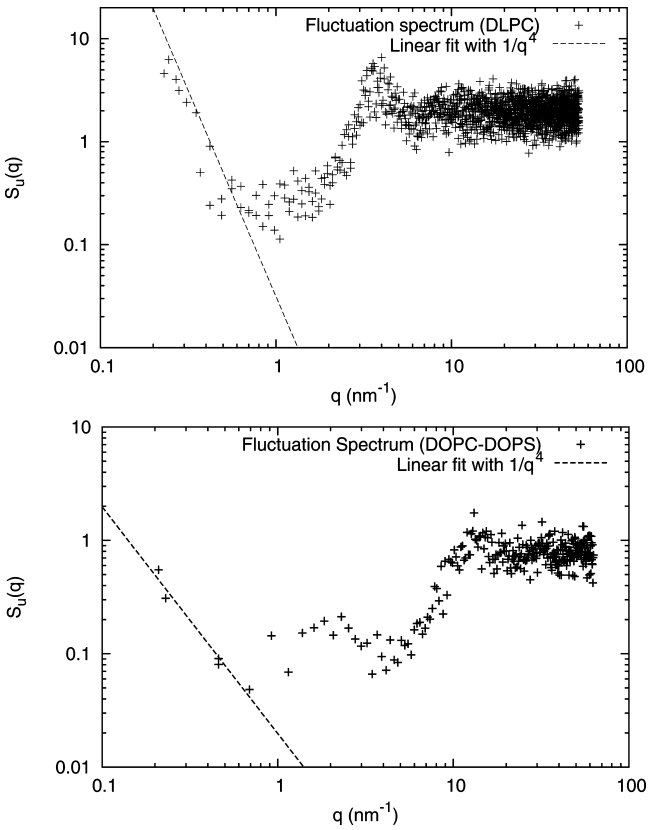


Figure 12. Full fluctuation spectra $S_u(q)$ for the (a) DLPC and (b) DOPC/DOPS HCG system versus wave vector **q**. The long wavelength undulation spectrum is used to calculate the bending modulus. The fitting is done to the theoretical q^{-4} line and is denoted as dashed line.

The bending modulus for a mixed DOPC/DOPS HCG model is calculated to be approximately equal to 7 k_BT based on the Helfrich model. This calculated bending modulus is comparable to estimates provided by experiments on similar lipid system.[86,85] Experiments[86,85] on DOPG/DOPS calculations estimated the bending modulus to be around 5 k_BT and were found to be heavily dependent on the composition.

The fluctuation spectrum for the mixed system is shown in Figure 12b. The fluctuation spectra profile is different from the spectra observed in pure lipid systems. A larger undulation (and extended protrusion)[91] zone is observed. The bending modulus for DOPC/DOPS lipids is lower than that of pure lipid

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

systems. The inclination for spontaneous curvature formation in the DOPC/DOPS system can be attributed to the significantly lower bending modulus as compared to pure lipid systems such as DOPC and DLPC.

4.2.3. Lateral Pressure Distribution. Pressure in a lipid bilayer model is usually characterized using the lateral pressure profile across the bilayer thickness. It is defined as the difference between the lateral and normal component of the pressure tensor in the lipid system.[94,95,92,96] Given that the diagonal elements of the pressure tensor are P_{xx}, P_{yy}, and P_{zz}, where the z direction is normal to the plane of the lipid surface, the lateral pressure across the thickness of the lipid bilayer is given as $P_{lat}(z) = P_{zz}(z) - 0.5[P_{xx}(z) + P_{yy}(z)]$. Lateral pressure can be directly related to the surface tension and the bending modulus of the bilayer systems.[92,96,97]

In this work, the algorithm for generating pressure profile is based on per-atom stress tensor calculation. The atomic-stress tensor $\mathbf{s}_{i\alpha\beta}$ for a given atom is computed based on the virial formulation as

$$\mathbf{s}_{i\alpha\beta} = -\frac{1}{V_i}\sum\left[\frac{1}{2}\sum_{j\neq i}(\vec{\mathbf{r}}_{ij}\otimes\vec{\mathbf{f}}_{ij}) + m_i\vec{\mathbf{v}}_{ij}\otimes\vec{\mathbf{v}}_{ij}\right] \quad (8)$$

where α and β take on values x, y, z to generate the six components of the symmetric stress tensor. Here, $\mathbf{r}_{ij}$ is the vector from the position of the i-th atom to the position of the j-th atom. The velocity $\mathbf{v}_i$ is the velocity vector of the i-th atom, and m_i is the atomic mass of i-th atom. The stress for each atom is due to its interaction with all other atoms in the simulation. V_i is the individual atomic volume. In this work, the atomic volume is calculated through Voronoi tessellation using the Voro++ tool.[98] Figure 13a shows the "atomic volume" for each CG site calculated using Voronoi tessellation. To measure the z-dependent pressure profile, the simulation box is divided into 1 Å thick layers perpendicular to the bilayer normal direction. The total stress in a given layer is calculated from average stress of atoms belonging to the layer. The average over time is used as the final stress tensor value at the mean z-location for that layer. The pressure tensor is obtained by taking the negative of the stress tensor.

The lateral pressure profile obtained from the CG simulation of DOPC HCG model is shown in Figure 13b. It is worth mentioning the lateral pressure profile calculated in this work using the virial formulation for the CG system gives an estimate of the relative lateral pressure across the thickness of the bilayer and in no way should the pressure values be taken to be exact because of the shortcomings of using a virial pressure formulation for coarse-grained systems. The profile shows positive pressures near the water−lipid interface and in the tail interior and a negative lateral pressure in the middle of the leaflet. This profile is qualitatively consistent with other lipid models.[92,96,40,45,48,44] The lateral pressure profile can be broadly divided into three distinct zones for analysis. The first zone marks the extreme left and right of the profile and represents the headgroup and solvent interface region of the all-atom lipid system. The lateral pressure is zero for isotropic water and sharply rises to about 250 bar in the region representative of the water−lipid interface. The positive lateral pressure in this zone reflects the repulsive interactions between headgroup sites that prevent the head groups from aggregating into a reverse-micelle-like structure, thereby disrupting the bilayer morphology. The second zone marks the middle of the leaflet where the lateral pressure steeply drops to as much as −600 bar indicating

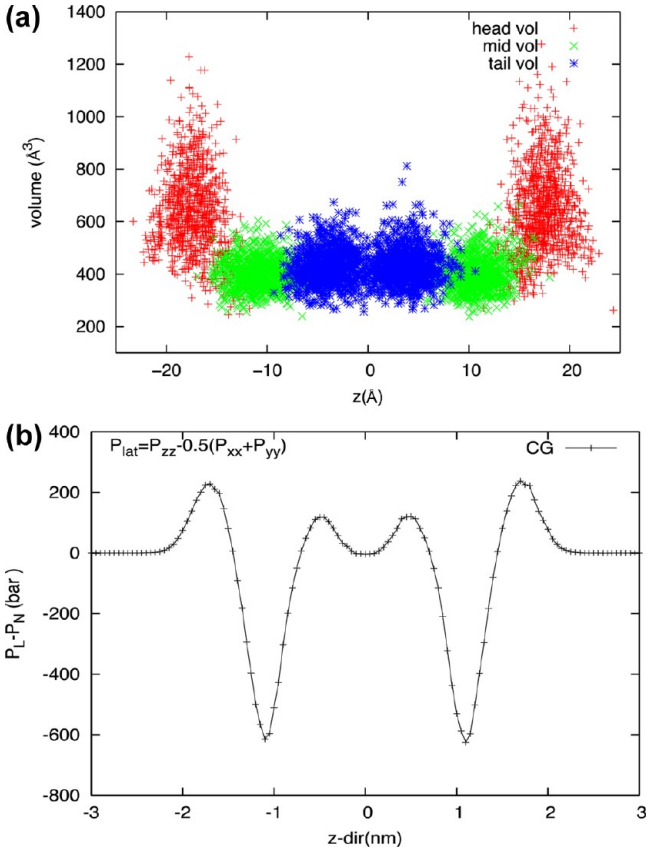


Figure 13. (a) Circumscribing volume of individual CG site calculated using Voronoi tessellation. (b) Lateral pressure profile for the HCG DOPC membrane.

a strong compression zone. This is the region in the lipid system where the hydrocarbon atoms past the hydrophilic head groups compete to reduce contact with water. The third zone is the center of the bilayer where lateral pressure rises up to 100 bar and settles to a nominally positive value in the middle. The positive values of lateral pressure in the core suggest that tails at the center of the bilayer tend to expand the system laterally. This tail−tail repulsion can be attributed to entropic resistance to compression at the center of the bilayer.

4.3. Liposome Simulation. HCG models for DOPC and DLPC lipid systems were tested with large liposome simulations (∼120 nm diameter) for at least 20 million time steps with a time step of 50 fs. Two snapshots from liposome simulations performed for pure DLPC HCG lipid models are shown in Figure 14. In both sets, the size of the liposome is 120 nm in diameter and it consists of 17 2316 lipids (516 948 CG sites). This CG system effectively represents the full liposome including the water atoms. In the first simulation, the liposome system was run for 20 million steps (time step of 50 fs). The liposome was stable for the duration of the simulation, as shown in Figure 14a. A second set of simulations on the liposome was done with initially distorted structure to test the resilience of the model to distortions. The initial distorted liposome was prepared by bringing together two liposomes in close proximity with each other and simulating for 4 million CG time steps. At the end of the simulation, both liposomes were distorted at the interface. One of them was selected for the resilience test, and the initial structure is shown in Figure 14b. The spherical shape was recovered from the distorted

dx.doi.org/10.1021/ct300751h | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

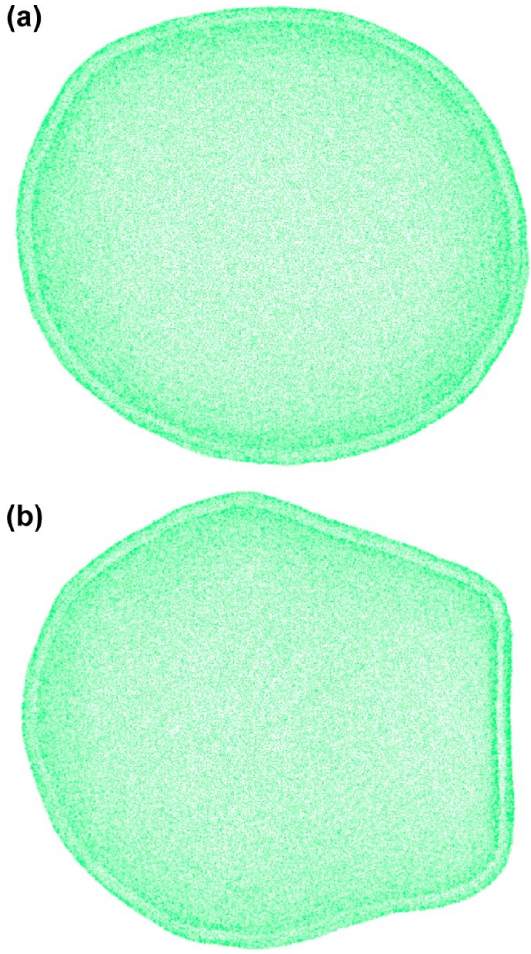


Figure 14. (a) CG simulation of liposome (DLPC), 120 nm in diameter, after 1 μs of CG simulation run with time step of 50 fs. (b) Distorted liposome modeled, which converged to spherical geometry after 20 ns of simulation time. The test was done to test the resilience of the model to deformation.

liposome within 10 million CG time steps. On a 512-processor machine, 150−200 ns of simulation time could be achieved in 24 h with the current model while running on LAMMPS MD software. Additional CG simulations were carried out to test the fidelity of the HC model for high curvature systems (smaller liposomes). The four smaller liposomes that were simulated had diameters of 10 nm, 20 nm, 40 nm, and 80 nm. The liposomes with diameters greater than 10 nm were found to be stable for 20 million time steps. It must be noted that these HCG models are not intended for small systems, which should presumably require a more highly resolved CG model to model accurately. However, the present simulations reflect the considerable advantage in terms of computational efficiency and the range of mesoscopic time and length scales that can be accessed by the current MS-CG based aggressive HCG lipid models.

5. CONCLUSION

A new HCG protocol based on the MS-CG methodology is presented in this paper to systematically develop solvent-free highly CG models for lipids. CG interaction parameters are obtained using the MS-CG method in the range where atomistic configurations are well sampled. In this hybrid approach, analytical functions are used for CG interactions in the sparsely sampled SR configuration space at atomic-level. This HCG methodology was applied to develop three-site CG models for DLPC, DOPC, and DOPC/DOPS lipid systems in this work. A four-site DOPC HCG model with two tails was also developed to show the flexibility of the modeling approach in terms of user defined CG resolution. The HCG model shows good agreement with atomic-level structural properties such as RDF and z-density despite the very aggressive nature of the coarse-graining. The model is also shown to reproduce critical biophysical membrane properties, including area compressibility, bending modulus, and lateral stress profile.

This model provides significant advantages in the simulation length and time scales that can be explored for important membrane−protein related biophysical studies. The scope of this model is demonstrated by successfully simulating a 120 nm diameter liposome for 1 μs (20 million steps with 50 fs time step). The protocol developed in this paper can be used to develop similar CG models of other lipids and lipid−mixture systems that can be used in the study of large membrane−protein systems in the future. As a part of future work, further modifications to the HCG model will also be made to explicitly represent the screened electrostatic interactions, especially for anionic lipids such as DOPS, dioleoylphosphatidylglycerol (DOPG), and phosphatidylinositol 4,5-bisphosphate/phosphatidyl inositol 3,4,5-triphosphate (PIP2/PIP3).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Interaction parameter tables for all the four systems studied in this work; tables for pairwise nonbonded interactions for DLPC, DOPC (3 sites), DOPC/DOPS, and DOPC (4 sites) models; bonded and angle potential; limits of developing the HCG model by fitting the MS-CG data to simple nonbonded analytical form (Figures S1, S2, and Table S5); user-defined parameter R_{SR} for various systems (Table S6). This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoth@uchicago.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health (NIH Grant No. R01-GM063796). The authors thank Mijo Simunovic and Dr. Gerrick Lindberg for their critical reading of the manuscript. Chun-Liang Lai and Dr. Frank Vazquez are acknowledged for insightful discussions. Computational resources were provided by the National Science Foundation through XSEDE computing resources (Grant No. TG-MCA94P017) specifically, the Texas Advanced Computing Center.

■ REFERENCES

(1) *Structure and Dynamics of Membranes I. From Cells to Vesicles*, 1st ed.; Lipowsky, R.; Sackmann, E., Eds.; *Handbook of Biological Physics*, Vol. *1A*; Hoff, A. J., Ed.; Elsevier: North Holland, 1995; p 1052.

(2) Alberts, B.; Johnson, A.; Lewis, L.; Raff, M.; Roberts, K.; Walter, P. *Mollecular Biology of the Cell*, 4th ed.; Garland Science: New York, 2002; p 1616.

(3) Karp, G. *Cell and Molecular Biology: Concepts and Experiments*, 5th ed.; John Wiley & Sons: Hoboken, NJ, 2007; p 864.

(4) Brown, H. A.; Marnett, L. J. Introduction to lipid biochemistry, metabolism, and signaling. *Chem. Rev.* **2011**, *111*, 5817−5820.
(5) Heuser, J. E.; Reese, T. S.; Dennis, M. J.; Jan, Y.; Jan, L.; Evans, L. Synaptic vesicle exocytosis captured by quick freezing and correlated with quantal transmitter release. *J. Cell Biol.* **1979**, *81*, 275−300.
(6) McMahon, H. T.; Gallop, J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **2005**, *438*, 590−596.
(7) Filmore, D. It's a GPCR world. *Mod. Drug Discovery* **2004**, *7*, 24−28.
(8) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *Biochim. Biophys. Acta* **2000**, *1469*, 159−195.
(9) Petrache, H.; Dodd, S.; Brown, M. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by ^{2}H NMR spectroscopy. *Biophys. J.* **2000**, *79*, 3172−3192.
(10) *The Structure of Biological Membranes*, 2nd ed.; Yeagle, P., Ed.; CRC Press: Boca Raton, FL, 2004; p 552.
(11) McGillivray, D. J.; Valincius, G.; Vanderah, D. J.; Febo-Ayala, W.; Woodward, J. T.; Heinrich, F.; Kasianowicz, J. J.; Lösche, M. Molecular-scale structural and functional characterization of sparsely tethered bilayer lipid membranes. *Biointerphases* **2007**, *2*, 21−33.
(12) Petrache, H.; Salmon, A.; Brown, M. Structural properties of docosahexaenoyl phospholipid bilayers investigated by solid-state ^{2}H NMR spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 12611−12622.
(13) Veatch, S. L.; Polozov, I. V.; Gawrisch, K.; Keller, S. L. Liquid domains in vesicles investigated by NMR and fluorescence microscopy. *Biophys. J.* **2004**, *86*, 2910−2922.
(14) Kučerka, N.; Tristram-Nagle, S.; Nagle, J. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **2006**, *208*, 193−202.
(15) Peter, B. J.; Kent, H. M.; Mills, I. G.; Vallis, Y.; Butler, P. J. G.; Evans, P. R.; McMahon, H. T. BAR domains as sensors of membrane curvature: The amphiphysin BAR structure. *Science* **2004**, *303*, 495−499.
(16) Gallop, J. L.; Jao, C. C.; Kent, H. M.; Butler, P. J.; Evans, P. R.; Langen, R.; McMahon, H. T. Mechanism of endophilin N-BAR domain-mediated membrane curvature. *EMBO J.* **2006**, *25*, 2898−2910.
(17) Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Clarendon Press: New York, 1989; p 385.
(18) *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, Binder, K., Ed.; Oxford University Press: New York, 1995; p 587.
(19) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.
(20) *Computational Modeling of Membrane Bilayers*, 1st ed.; Feller, S., Ed.; Current Topics in Membranes, Vol. *60*; Simon, S. A., Benos, D. J., Eds.; Vol. *60*; Academic Press: Burlington, MA, 2008; p 448.
(21) Ayton, G. S.; Voth, G. A. Systematic multiscale simulation of membrane protein systems. *Curr. Opin. Struct. Biol.* **2009**, *19*, 138−144.
(22) Marrink, S. J.; de Vries, A. H.; Tieleman, D. P. Lipids on the move: Simulations of membrane pores, domains, stalks, and curves. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 149−168.
(23) Tieleman, D.; Marrink, S. J. Lipids out of equilibrium: Energetics of desorption and pore mediated flip-flop. *J. Am. Chem. Soc.* **2006**, *128*, 12462−12467.
(24) Gurtovenko, A. A.; Vattulainen, I. Molecular mechanism for lipid flip-flops. *J. Phys. Chem. B* **2007**, *111*, 13554−13559.
(25) Leontiadou, H.; Mark, A. E.; Marrink, S. J. Molecular dynamics simulations of hydrophilic pores in lipid bilayers. *Biophys. J.* **2004**, *86*, 2156−2164.
(26) Sengupta, D.; Leontiadou, H.; Mark, A. E.; Marrink, S. J. Toroidal pores formed by antimicrobial peptides show significant disorder. *Biochim. Biophys. Acta* **2008**, *1778*, 2308−2317.
(27) Vernier, P. T.; Ziegler, M. J.; Sun, Y.; Chang, W. V.; Gundersen, M. A.; Tieleman, D. P. Nanopore formation and phosphatidylserine externalization in a phospholipid bilayer at high transmembrane potential. *J. Am. Chem. Soc.* **2006**, *128*, 6288−6289.
(28) Khavrutskii, I. V.; Gorfe, A. A.; Lu, B.; McCammon, J. A. Free energy for the permeation of Na^+ and Cl^- ions and their ion-pair through a zwitterionic dimyristoyl phosphatidylcholine lipid bilayer by umbrella integration with harmonic fourier beads. *J. Am. Chem. Soc.* **2009**, *131*, 1706−1716.
(29) Ayton, G. S.; Lyman, E.; Voth, G. A. Hierarchical coarse-graining strategy for protein−membrane systems to access mesoscopic scales. *Faraday Discuss.* **2010**, *144*, 347−357.
(30) Ayton, G. S.; Voth, G. A. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* **2010**, *99*, 2757−2765.
(31) Heuser, J. E.; Reese, T. S.; Dennis, M. J.; Jan, Y.; Jan, L.; Evans, L. Synaptic vesicle exocytosis captured by quick freezing and correlated with quantal transmitter release. *J. Cell Biol.* **1979**, *81*, 275−300.
(32) Lee, E.; Marcucci, M.; Daniell, L.; Pypaert, M.; Weisz, O. A.; Ochoa, G. C.; Farsad, K.; Wenk, M. R.; De Camilli, P. Amphiphysin 2 (Bin1) and T-tubule biogenesis in muscle. *Science* **2002**, *297*, 1193−1196.
(33) Sato, K.; Nakano, A. Dissection of COPII subunit-cargo assembly and disassembly kinetics during Sar1p-GTP hydrolysis. *Nat. Struct. Mol. Biol.* **2005**, *12*, 167−74.
(34) Brannigan, G.; Lin, L.; Brown, F. Implicit solvent simulation models for biomembranes. *Eur. Biophys. J.* **2006**, *35*, 104−124.
(35) Voth, G. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Taylor & Francis: Boca Raton, 2008.
(36) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. Multiscale modeling of emergent materials: Biological and soft matter. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869−1892.
(37) Berger, O.; Edholm, O.; Jahnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **1997**, *72*, 2002−2013.
(38) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. A coarse grain model for phospholipid simulations. *J. Phys. Chem. B* **2001**, *105*, 4464−4470.
(39) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: A coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(40) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(41) Ayton, G. S.; Voth, G. A. Hybrid coarse-graining approach for lipid bilayers at large length and time scales. *J. Phys. Chem. B* **2009**, *113*, 4413−4424.
(42) Izvekov, S.; Voth, G. A. Solvent-free lipid bilayer model using multiscale coarse-graining. *J. Phys. Chem. B* **2009**, *113*, 4443−4455.
(43) Lu, L.; Voth, G. A. Systematic coarse-graining of a multi-component lipid bilayer. *J. Phys. Chem. B* **2009**, *113*, 1501−1510.
(44) Wang, Z.-J.; Deserno, M. A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *J. Phys. Chem. B* **2010**, *114*, 11207−11220.
(45) Brannigan, G.; Philips, P.; Brown, F. Flexible lipid bilayers in implicit solvent. *Phys. Rev. E* **2005**, *72*, 011915(1−4).
(46) Cooke, I.; Kremer, K.; Deserno, M. Tunable generic model for fluid bilayer membranes. *Phys. Rev. E* **2005**, *72*, 011506(1−4).
(47) Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K. Coarse grained protein−lipid model with application to lipoprotein particles. *J. Phys. Chem. B* **2006**, *110*, 3674−3684.
(48) Sodt, A. J.; Head-Gordon, T. An implicit solvent coarse-grained lipid model with correct stress profile. *J. Chem. Phys.* **2010**, *132*, 205103(1−8).
(49) Noguchi, H. Solvent-free coarse-grained lipid model for large-scale simulations. *J. Chem. Phys.* **2011**, *134*, 055101(1−12).
(50) Drouffe, J.; Maggs, A.; Leibler, S. Computer simulations of self-assembled membranes. *Science* **1991**, *254*, 1353−1356.
(51) Farago, O. "Water-free" computer model for fluid bilayer membranes. *J. Chem. Phys.* **2003**, *119*, 596−605.
(52) Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.

(53) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114(1−11).
(54) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115(1−20).
(55) Izvekov, S.; Voth, G. A. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *J. Chem. Theory Comput.* **2006**, *2*, 637−648.
(56) Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site−site potential. *J. Chem. Phys.* **1981**, *74*, 3316−3319.
(57) Lu, L.; Voth, G. A. The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. *J. Chem. Phys.* **2011**, *134*, 224107(1−10).
(58) Izvekov, S.; Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **2005**, *123*, 134105(1−13).
(59) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Voth, G. A. Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B* **2007**, *111*, 4116−4127.
(60) Das, A.; Andersen, H. C. The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. *J. Chem. Phys.* **2009**, *131*, 034102(1−11).
(61) Larini, L.; Lu, L.; Voth, G. A. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.* **2010**, *132*, 164107(1−10).
(62) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *49*, 61−74.
(63) Reith, D.; Meyer, H.; Muller-Plathe, F. CG-OPT: A software package for automatic force field design. *Comput. Phys. Commun.* **2002**, *148*, 299−313.
(64) 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*, 7830−7843.
(65) Ulrich, A. S.; Watts, A. Molecular response of the lipid headgroup to bilayer hydration monitored by 2H-NMR. *Biophys. J.* **1994**, *66*, 1441−1449.
(66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(68) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(69) Kučerka, N.; Liu, Y.; Chu, N.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* **2005**, *88*, 2626−2637.
(70) Zhao, W.; Rug, T.; Gurtovenko, A. A.; Vattulainen, I.; Karttunen, M. Atomic-scale structure and electrostatics of anionic palmitoyloleoylphosphatidylglycerol lipid bilayers with Na+ counterions. *Biophys. J.* **2007**, *92*, 1114−1124.
(71) Pan, J.; Heberle, F. A.; Tristram-Nagle, S.; Szymanski, M.; Koepfinger, M.; Katsaras, J.; Kučerka, N. Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 2135−2148.
(72) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *Biochim. Biophys. Acta, Rev. Biomembr.* **2000**, *1469*, 159−195.
(73) Steve, P. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(74) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14* (33−8), 27−8.
(75) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(76) Kučerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys. J.* **2008**, *95*, 2356−2367.
(77) Izvekov, S.; Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys.* **2006**, *125*, 151101(1−4).
(78) Murarka, R. K.; Liwo, A.; Scheraga, H. A. Separation of time scale and coupling in the motion governed by the coarse-grained and fine degrees of freedom in a polypeptide backbone. *J. Chem. Phys.* **2007**, *127*, 155103(1−16).
(79) Izvekov, S.; Chung, P. W.; Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **2010**, *133*, 064109(1−16).
(80) Soheilifard, R.; Makarov, D. E.; Rodin, G. J. Rigorous coarse-graining for the dynamics of linear systems with applications to relaxation dynamics in proteins. *J. Chem. Phys.* **2011**, *135*, 054107(1−10).
(81) Braun, A. R.; Brandt, E. G.; Edholm, O.; Nagle, J. F.; Sachs, J. N. Determination of electron density profiles and area from simulations of undulating membranes. *Biophys. J.* **2011**, *100*, 2112−2120.
(82) Tristram-Nagle, S.; Petrache, H. I.; Nagle, J. F. Structure and interactions of fully hydrated dioleoylphosphatidylcholine bilayers. *Biophys. J.* **1998**, *75*, 917−925.
(83) Rawicz, W.; Olbrich, K. C.; McIntosh, T.; Needham, D.; Evans, E. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.* **2000**, *79*, 328−339.
(84) Poger, D.; Mark, A. E. On the validation of molecular dynamics simulations of saturated and cis-monounsaturated phosphatidylcholine lipid bilayers: A comparison with experiment. *J. Chem. Theory Comput.* **2009**, *6*, 325−336.
(85) Allen-Rodowicz, K.; Francisco, H.; Layton, B. Determination of the mechanical properties of DOPC/DOPS liposomes using an image procession algorithm and micropipette-aspiration techniques. *Chem. Phys. Lipids* **2010**, *163*, 787−793.
(86) Claessens, M. M. A. E.; van Oort, B. F.; Leermakers, F. A. M.; Hoekstra, F. A.; Cohen Stuart, M. A. Bending rigidity of mixed phospholipid bilayers and the equilibrium radius of corresponding vesicles. *Phys. Rev. E* **2007**, *76*, 011903(1−6).
(87) Marrink, S. J.; Mark, A. E. The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *J. Am. Chem. Soc.* **2003**, *125*, 11144−11145.
(88) Zimmerberg, J.; Kozlov, M. M. How proteins produce cellular membrane curvature. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 9−19.
(89) Derek, M. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophys. J.* **2007**, *93*, 3884−3899.
(90) Helfrich, W. Elasticity and thermal undulations of fluid films of amphiphiles. In *Liquids and Interfaces*; Charvolin, J., Joanny, J. F., Zinn-Justin, J., Eds.; Les Houches XLVIII 1988; Elsevier: Amsterdam, 1990; p 680.
(91) Brandt, E. G.; Braun, A. R.; Sachs, J. N.; Nagle, J. F.; Edholm, O. Interpretation of fluctuation spectra in lipid bilayer simulations. *Biophys. J.* **2011**, *100*, 2104−2111.
(92) Lindahl, E.; Edholm, O. Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys. J.* **2000**, *79*, 426−433.
(93) Pan, J.; Tristram-Nagle, S.; Kučerka, N.; Nagle, J. F. Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. *Biophys. J.* **2008**, *94*, 117−124.
(94) Marsh, D. Lateral pressure in membranes. *Biochim. Biophys. Acta, Rev. Biomembr.* **1996**, *1286*, 183−223.

(95) Cantor, R. S. Lateral pressures in cell membranes: A mechanism for modulation of protein function. *J. Phys. Chem. B* **1997**, *101*, 1723−1725.
(96) Lindahl, E.; Edholm, O. Spatial and energetic-entropic decomposition of surface tension in lipid bilayers from molecular dynamics simulations. *J. Chem. Phys.* **2000**, *113*, 3882(1−12).
(97) Marrink, S. J.; Mark, A. E. Effect of undulations on surface tension in simulated bilayers. *J. Phys. Chem. B* **2001**, *105*, 6122−6127.
(98) Rycroft, C. H.; Kamrin, K.; Bazant, M. Z. Assessing continuum postulates in simulations of granular flow. *J. Mech. Phys. Solids* **2009**, *57*, 828−839.