Hybrid Coarse-Graining Approach for Lipid Bilayers at Large Length and Time Scales

Gary S. Ayton and Gregory A. Voth*

Center for Biophysical Modeling and Simulation and Department of Chemistry, University of Utah, 315 S. 1400 E., Room 2020, Salt Lake City, Utah 84112-0850

A hybrid analytic−systematic (HAS) coarse-grained (CG) lipid model is developed and employed in a large-scale simulation of a liposome. The methodology is termed hybrid analytic−systematic because one component of the interaction between CG sites is variationally determined from the multiscale coarse-graining (MS-CG) methodology, whereas the remaining component utilizes an analytic potential. The systematic component models the in-plane center-of-mass interaction of the lipids as determined from an atomistic-level MD simulation of a bilayer. The analytic component is based on the well-known Gay−Berne ellipsoid-of-revolution liquid-crystal model and is designed to model the highly anisotropic interactions at a highly coarse-grained level. The HAS CG approach is the first step in an "aggressive" CG methodology designed to model multicomponent biological membranes at very large length and time scales.

1. Introduction

Arguably, one of the greatest challenges facing the field of biomolecular simulation and modeling of lipid bilayers is making the connection to real biological membranes. Real membranes are highly inhomogenous and include multiple lipids, cholesterol, and numerous proteins. Furthermore, in contrast to the original fluid mosaic model,[1] a new picture is emerging whereby the membrane is crowded, containing a large number of highly oligomerized proteins, and has a varying membrane thickness, as well as lipid spatial organization.[2,3] In some cases, upward of 20% or higher of the surface area of a membrane is occupied by proteins.[4] This scenario is quite far removed from model systems, for example, giant unilamellar vesicles (GUVs) consisting of a limited number of components.[5] Many experiments have utilized model membrane systems (see, e.g., refs 6 and 7 for two examples) rather than examining actual in vivo processes, as the inherent complexity of the real biological membrane does not allow for a fine control over various experimental conditions. Nonetheless, especially in terms of simulation and modeling, the challenge of moving toward examining real biological membranes and associated systems remains a critical priority.

From a simulation viewpoint, modeling real biological membranes presents a number of key challenges. First, as previously mentioned, real biological membranes contain multiple components that can include lipids, proteins, and nonlipid molecules such as cholesterol. Molecular dynamics (MD) simulations require that the interactions between every atom for each of the different components be rigorously specified within the MD force field. Here, the large number of components implies that a large number of interactions thus need to be specified, and even at the atomistic level, this can be a challenge, as a high level of accuracy in the simulation force fields might be required. Second, and perhaps more importantly, real biological membranes contain multiple length and time scales ranging from the atomistic (nanometers, nanoseconds) to the macroscopic (millimeters, milliseconds). As a result, the behavior of a small, isolated component of a biological membrane

can be quite different from that observed in the real membrane. Furthermore, MD simulation alone currently cannot even remotely access the longest length and time scales that are characteristic of real membranes (e.g., a cell with dimensions on the order of micrometers and time scales of milliseconds). As such, from the outset, multiscale simulation methodologies are required where the different scales are somehow systematically bridged or coupled.[8]

An emerging and core component of an overall multiscale simulation methodology for biological membranes is coarse-grained (CG) MD simulation.[8−47] This particular component bridges atomistic length and time scales with more field-based mesoscopic scales.[48−51] In fact, CG-MD has been used to examine emergent phenomena for relatively complex biomolecular systems at length and time scales that are currently out of reach of traditional atomistic-level MD.[8] For example, CG-MD has been employed to model both proteins[8,9,47,52,53] and lipid bilayers.[8−47] In the case of lipids, CG lipid models generally contain around 10−15 CG sites,[10−15] compared to 46 atoms in a united-atom lipid model[54] or over 100 atoms for a fully atomistic representation. Because the number of CG sites is less than the original number of atoms and the interactions between them are generally simpler and shorter range, a significant computational acceleration is possible.

It should be noted that the particular CG scheme that is chosen can be flexible. That is, both the number and location of the CG sites required to describe a particular molecule can, in principle, vary depending on the problem at hand. In cases where very large membranes are to be examined (but not so large that field-based mesoscopic models are required[48−51]), the degree of coarse graining might have to be quite aggressive. For example, in the modeling of a 200-nm-diameter liposome containing upward of half a million lipids, an entire lipid might have to be coarse-grained into one or two sites. It is also possible to develop so-called "solvent-free" CG membrane models,[17−21,23] where the hydrophobic−hydrophilic lipid−solvent interactions are subsumed into the new CG lipid interactions. This latter approach, in particular, can result in significant computational time savings, as a large part of the numerical effort is taken up accounting for the surrounding solvent and membrane solvent

---

* Corresponding author. E-mail: voth@chemistry.utah.edu.

interactions. This important aspect of the problem will be revisited shortly.

Simulations of liposomes and vesicles have been performed with CG simulation;[18,23,26–28,37,39,55] however, at this time, the largest systems studied have been small vesicles with diameters in the range of 40–60 nm[39] with typically less than 5000 lipids.[18,23,26,28] Liposomes of these sizes are too small to model most biologically relevant systems where the liposome diameter is upward of 200 nm.[6] Thus, it is important to explore even more highly coarse-grained lipid models such that these large liposome length scales can be achieved.

The aim of the present article is not to introduce yet another CG lipid model. As has been previously discussed, a multitude of "higher-resolution" CG lipids models have already been developed. The goal here is to develop a methodology wherein highly coarse-grained, low-resolution lipid models can be utilized at very large length scales while retaining at least a semiquantitative agreement with the underlying molecular-scale system. The methodology presented here is general and can potentially be applied to a number of CG lipid models[17,18,20,29] as a basis. However, the overall spirit of the present work is to start with a low-resolution CG lipid model that has as few sites as possible, as the difference between a single-site versus a 10-site CG model, for example, translates into a substantial increase in computational efficiency, thereby allowing the model to access much larger length scales. Moreover, as one begins to model larger-scale structures such as a 200-nm-diameter liposome, it is quite sensible that the resolution of the underlying CG model can become correspondingly lower.

An experimental example of a scenario in which a highly coarse-grained membrane model could be utilized is protein-mediated membrane remodeling,[6,7,47,48,56] where, for example, whole liposomes are remodeled by proteins and protein modules [e.g., the BAR (Bin/amphiphysin/Rvs) domain] into tubules.[6,48,57] Furthermore, experimentally, the process of tubulation often occurs over time scales much longer than microseconds.[58]

It is important to note that CG-MD shares with MD its dependence on an effective force field to define the interactions between CG "sites". It is not immediately obvious or trivial to determine these interactions, especially in multicomponent systems such as real biological membranes. In practice, many current CG models rely on a more "top-down" approach,[10,11,13,20,21] where the interactions between CG sites are designed from the outset to reproduce some desired macroscopic behavior. Such techniques, such as inverse Monte Carlo, have even been aggressively employed to construct a two-dimensional model for a dipalmitoylphosphatidylcholine (DPPC)/cholesterol bilayer that mapped the entire lipid (and cholesterol molecules) onto a single site associated with the center of mass of the molecules.[46] The two-dimensional constraint significantly simplified the interactions between the CG sites; because the CG lipids could not leave the plane of the membrane, only a pairwise-additive, radially symmetric interaction was required. The CG interactions incorporated, in some effective way, the out-of-plane lipid interactions. However, by definition, the model missed critical long-wavelength thermal undulations out of the plane of the membrane, and it would also be quite restrictive in the case of modeling liposomes and other closed surfaces that are more indicative of real biological membranes in cells.

Ideally, a CG model should be systematically derived from its underlying atomistic-level interactions. This is, in fact, one of the outstanding challenges in the field of CG simulation: to rigorously derive the force field for a CG model from the corresponding force field employed in a high-resolution atom-

istic-level system. It is especially desirable to construct a CG model entirely from atomistic-level information; in this case, all new emergent behavior that is observed is then predicted, rather than "tuned" or "engineered" based on empirical data. In this way, the behavior of the CG system can be traced back to the underlying atomistic-level interactions that were used in developing the CG model.

One possible approach to systematically develop CG models from atomistic-level information is the multiscale coarse-graining (MS-CG) methodology.[15,16,59–67] This approach relies on a variational principle[59,60] that determines the optimal CG force field for a given finite set of MD configurations (i.e., employing the Cartesian coordinates and forces of all the atoms for a large number of MD-generated configurations), along with a predefined CG force-field basis set that contains N_d parameters, $\phi = \{\phi_1,...,\phi_{N_d}\}$, that can be variationally determined from the available MD data.[59–61]

The process of variational fitting has been called "force matching"[15,59,60] and can be shown to guarantee the best fit of the available atomistic data to the CG force-field basis set. The scheme only requires that the CG force be linear in its coefficients,[60] and what results is a CG "fitted force field" whose degree of flexibility and/or complexity lies in the initial choice of the basis set.[60] In the case of nonbonded interactions, the MS-CG methodology can readily generate pairwise-additive radial force fields, where the radial dependence of the force field is modeled with, for example, spline coefficients[15,16,63–65] or even a delta-function basis.[62]

However, in the case of real biological membranes, a direct application of the MS-CG methodology can prove challenging because of the size and complexity of the initial MD simulation that is required to generate the MS-CG force field. Strictly speaking, the optimum force field obtained from the MS-CG methodology is reached only in the limit that the underlying MD simulation samples essentially all possible configurations. In that limit, all of the atoms that correspond to various CG sites can interact with all of the others such that a set of converged MS-CG force fields for that system can be generated. However, atomistic MD simulations of large biological membranes can sample only a relatively small set of configurations. For example, consider an atomistic simulation of a multicomponent lipid bilayer, with cholesterol and multiple embedded proteins. Over the duration of the typical MD simulation (perhaps 100 ns), some interactions are hardly, if ever, sampled, for example, a water–lipid tail interaction. Furthermore, within the duration of the MD simulation, long-time-scale processes, such as the formation of domain structures, protein-mediated lipid sequestering,[3] and other complex phenomena (e.g., membrane-bound-protein oligomerization[2]) might not be observed.

The option of running multiple, extended MD simulations in order to fully sample the entire system in many ways defeats the purpose of moving to a CG representation. Ideally, the atomistic MD simulations used to construct the MS-CG force field should be as small and efficient as possible. In fact, the full biological membrane system might be broken down into smaller subsystems for the explicit purpose of sampling key interactions. The overall coarse-grained configurational space can be thought of as a vast domain that is highlighted by small regions that are well-sampled by the initial MD simulations used to generate the MS-CG force field. However, it also contains large domains where little or no atomistic-level sampling and information were obtained. The MS-CG methodology can therefore be used with confidence to obtain CG models that

apply to those well-sampled domains, but some other approach is required in order to fill in the gaps between them.

The option explored in this article is to develop a hybrid CG model that has two components: The first component is a systematically obtained MS-CG force field that is employed for CG configurations that correspond to well-sampled MD configurations, whereas the second component relies on an analytic model to describe the poorly sampled configurations. This approach is denoted as a hybrid analytic–systematic (HAS) CG model. It is utilized in cases where the majority of the sampled CG configurations are related to well-sampled MD configurations, so that only periodically is the analytic component of the CG model employed. For multicomponent membranes, the full CG configuration space can then be thought of as being populated by "islands" of CG configurations that correspond to underlying well-sampled atomistic MD configurations that are connected by the analytic component of the HAS CG model. This type of approach affords a operational solution to limited MD sampling in large biological membranes, yet retains a strong degree of systematically obtained CG interactions, which are critical for the CG model to be bridged to the underlying atomistic scale.

The present implementation of the HAS CG approach is also an "aggressive" CG model that is capable of reaching significant length and time scales, for example, a reasonably sized liposome of 200-nm diameter and half a million lipids. The HAS CG methodology thus aims to combine systematically obtained interactions obtained by the MS-CG methodology with very computationally efficient analytic models, where, from the outset, an aggressive CG scheme is employed such that large biological membrane systems can be modeled. Specifically, in this work, a single-site, solvent-free, lipid model is explored; however, other options could be employed, as the approach is general.

For this application of the HAS CG approach to membranes, the systematic component of the CG lipid bilayer uses the MS-CG approach to develop an in-plane, single-site, CG membrane force field. There are a number of analytic models that can then be employed for the analytic component of the HAS CG model. One option is the so-called "shape-based" CG model for lipids.[47] Here, the lipid is modeled as a dimer with one bead representing the headgroup and the second bead representing the tail. Another possibility used here is the Gay–Berne (GB) liquid-crystal model,[68,69] which has the ability to model ellipsoids of revolution with varying aspect ratios and reduces to a standard Lennard-Jones (LJ) form when two identical ellipsoids are parallel and side-by-side. A single GB ellipsoid requires one site at its center to designate its Cartesian location, along with a unit vector to designate its orientation. This approach has been used to model a variety of liquid-crystal phases[70] and even lipids at a very reduced level of resolution.[35,36,38] For this work, the single-site GB analytic model is used because it ties in with the single-site MS-CG systematic component as previously introduced.

With this HAS CG approach, the GB interaction provides an anisotropic "core", and the interactions found from the MS-CG methodology supply systematically determined force details arising from the actual average in-plane interactions as observed in the atomistic MD simulation. The MS-CG methodology samples interactions over only a limited set of MD configurations (e.g., an MD trajectory of a stable bilayer in the liquid-crystal phase), where a large contribution to the total force acting on a lipid comes from the in-plane lipid interactions. As such, the GB core accounts for other interactions that were never fully sampled in the original MD trajectory. These include, for example, lipid–lipid interactions that might be observed in a completely isotropic lipid–water mixture.

This article describes an HAS CG lipid bilayer model that is employed at a very large length scale to simulate an entire liposome with a diameter on the scale of 200 nm. The next section introduces how analytic force fields such as the GB interaction can be combined with the MS-CG methodology to yield an HAS CG model and discusses the MS-CG variational principle as used in the present context. Section 3 gives the results of both the atomistic and CG simulations; section 4 provides a brief summary.

2. Methods

2.1. Hybrid Analytic–Systematic (HAS) Lipid Bilayer Model. The goal is to construct an HAS CG, single-site, solvent-free bilayer model by combining an analytic GB ellipsoidal liquid-crystal model with an MS-CG force field. The MS-CG force field is employed in specific configurations where the force-matching method can be applied with a good level confidence, whereas the GB force field is used otherwise. Specifically, the MS-CG force field gives the in-plane lipid center-of-mass (COM) interactions, whereas the GB component models the intermonolayer interactions, very close-range interactions, and any other out-of-plane interactions such that the bilayer stability is maintained in three dimensions. The parametrizations for the GB component are empirically determined from all of the available atomistic MD information on hand.

The resulting HAS CG model is a CG pair potential that depends not only on the Cartesian locations of the CG sites, $\mathbf{R}_\alpha$, where α is the label corresponding to that site, but also on the orientation of the CG site, given by a unit vector $\mathbf{e}_\alpha$. The orientational dependence is critical for modeling the anisotropy of the lipid–lipid interaction with a single site model. The final CG pair potential is given by

$$U_{IJ}(\mathbf{R}_I, \mathbf{R}_J, \mathbf{e}_I, \mathbf{e}_J) =$$
$$\begin{cases} U_{IJ}^{GB}(\mathbf{R}_I, \mathbf{R}_J, \mathbf{e}_I, \mathbf{e}_J) & R_{IJ} < R_{C,1} \\ U_{IJ}^{GB}(\mathbf{R}_I, \mathbf{R}_J, \mathbf{e}_I, \mathbf{e}_J) + \Delta U(R_{IJ}) & R_{C,1} \leq R_{IJ} \leq R_{C,2} \\ 0 & R_{IJ} > R_{C,2} \end{cases} \quad (1)$$

where $R_{C,1}$ and $R_{C,2}$ are inner and outer preset cutoffs, respectively; $R_{IJ} = |\mathbf{R}_I - \mathbf{R}_J|$; and $\Delta U(R_{IJ})$ is a spherically symmetric correction term that is found from

$$\Delta U(R_{IJ}) = U_{IJ}^{MS}(R_{IJ}) - U_{IJ}^{GB,A}(R_{IJ}) \quad (2)$$

Here, $U_{IJ}^{MS}(R_{IJ})$ is the MS-CG pair potential, and $U_{IJ}^{GB,A}(R_{IJ})$ is the GB interaction for a pair of perfectly aligned and in-plane GB ellipsoids. As such, the pair potential as given in eq 1 becomes the MS-CG pair interaction for two parallel and in-plane CG sites that are within the two inner and outer cutoffs. The hybrid nature of this approach is clear: The model gives exactly the MS-CG interaction when two CG particles are parallel and side-by-side. This particular lipid–lipid configuration reasonably maps over to many of the configurations sampled in the atomistic MD bilayer simulation where the MS-CG in-plane force field was originally calculated. Of course, whereas the MD simulation remains in the bilayer phase throughout the atomistic simulation, the CG model can potentially explore new regions of phase space. When this happens, the anisotropic GB component of the interaction in eq 1 takes over. Because the full CG model contains both a GB ellipsoid

component and an MS-CG part, the CG particles will be referred to as "CG lipids", versus, for example, GB ellipsoids.

Details of the MS-CG approach can be found elsewhere;[15,16,59-65] here, the focus is on the HAS approach used to combine the GB liquid-crystal model with the MS-CG force field. A detailed description of the HAS CG methodology is discussed in the following subsections and consists of four steps:

A. Instantaneously calculate the centers of mass (COM) of the lipids in a fully solvated atomistic lipid bilayer MD simulation.

B. Force match this system using the COM to give an in-plane, pairwise-additive MS-CG force field between the COM locations of the lipids for each monolayer.

C. Use the average vertical separation of the COM sites, along with the atomistic area per lipid, to construct an analytic GB model of the bilayer.

D. Combine the two force fields into one CG lipid model in which the MS-CG component is employed in configurations where the FM force field applies and the GB component is used otherwise as given in eq 1.

2.2. Molecular Dynamics Simulation. An initial MD simulation of a fully solvated dimyristoylphosphatidylcholine (DMPC) bilayer was performed. The system consisted of 64 lipids and was equilibrated for 6 ns under isothermal and isobaric conditions. Simulation details can be found elsewhere.[71] Subsequent MS-CG calculations were performed under conditions of constant volume V, constant number of atoms n, and constant temperature T (nVT) in order to map over to the MS-CG framework, which is currently developed for the nVT ensemble.[59,61] The Cartesian spatial coordinates of the n atoms are given by $\mathbf{r}^n \equiv \{\mathbf{r}_1,..., \mathbf{r}_n\}$. The center of mass (COM) of each lipid was calculated instantaneously via the mapping operator, $\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}^n) = \{\mathbf{M}_{\mathbf{R}1}(\mathbf{r}^n),..., \mathbf{M}_{\mathbf{R}N}(\mathbf{r}^n)\}$, which is the set of N CG mapping operators where each one separately is given by[59]

$$\mathbf{M}_{\mathbf{R}I}(\mathbf{r}^n) = \sum_{i=1}^{n} c_{Ii}\mathbf{r}_i \tag{3}$$

where, in this case, $c_{Ii} = P_{I,i}m_i/M_I$, $M_I = \sum_{i=1}^{n} P_{I,i}m_i$ and is the mass of CG site I, and $P_{I,i} = 1$ if atom i is part of CG site I and 0 otherwise. The associated CG sites are denoted by $\mathbf{R}^N \equiv \{\mathbf{R}_1,..., \mathbf{R}_N\}$ and are related to the mapping operator by the identity
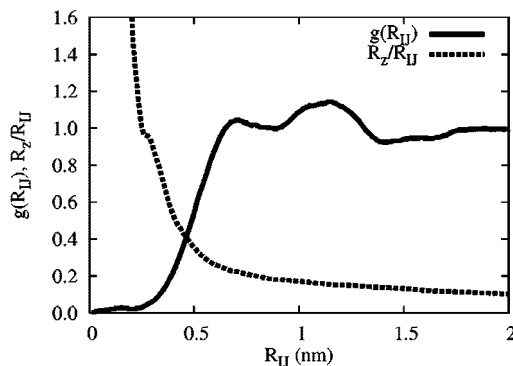
$$\int d\mathbf{R}^N \prod_{I=1}^{N} \delta(\mathbf{R}_I - \mathbf{M}_{\mathbf{R}I}(\mathbf{r}^n)) = 1 \tag{4}$$

The distinction between the CG sites and the mapping operator is subtle, but must be appreciated. The mapping operator gives an instantaneous measured COM of the lipids from an MD simulation; in turn, the CG sites describe the Cartesian locations of the CG lipids in the corresponding CG simulation. However, to keep the notation relatively simple, $\mathbf{R}^N \equiv \{\mathbf{R}_1,..., \mathbf{R}_N\}$ will sometimes be used to denote the COM location of the lipids in the MD simulation when appropriate. A snapshot of this instantaneous mapping is given in Figure 1, where it is clear from panels a and b that the COM resides near the middle of each lipid and that two distinct monolayers result. However, the COM locations are not perfectly in the plane and have a fair amount of variation in the z direction. Figure 1c shows a top-down view where, if anything, a complete lack of strong hexagonal symmetry is evident. In fact, some COM points
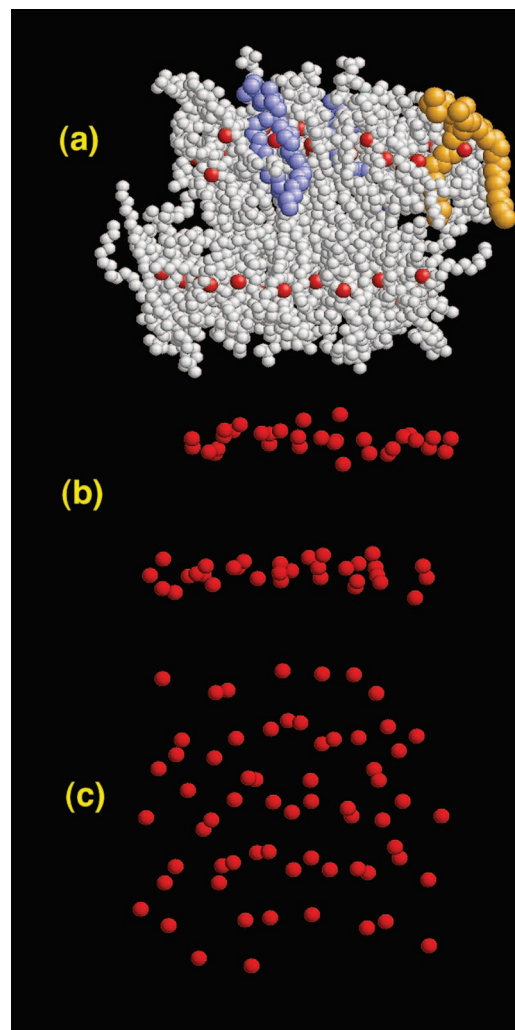
Figure 1. Simulation snapshots of the original atomistic DMPC system. (a) DMPC lipids with the CG sites shown as red spheres. The colored lipids highlight the two highly disordered tail conformations in the liquid-crystal phase. (b) Instantaneous locations of the COM CG sites corresponding to a. A distinct bilayer structure is observed. (c) Top-down view of the CG sites; very little structural correlation is observed, and some of the lipid COM CG sites are quite close together.

almost lie on top of each other. This behavior can be traced back to the COM mapping in eq 3; the lipid chains (as highlighted by the blue and orange lipids in Figure 1a) are quite disordered, which is typical of lipids in the liquid-crystal phase. As such, the COM mapping can result in sites that might reside at locations that are quite close, or even on top of, the COM of other lipid. The resulting COM structure of the full MD system gives two monolayers, where each monolayer has, at best, weak correlations. This is more fully realized in Figure 2, where the in-plane COM radial distribution function (RDF) is shown. The solid line in Figure 2 shows the RDF, averaged over the COM points in both leaflets of the bilayer separately. The small bump near zero corresponds to lipid COM points that are almost on top of each other and hence have a small in-plane radial component. The two peaks around 1 nm exhibit weak correlations, consistent with the observations in Figure 1. The size of the MD cell did not allow for the RDF to be examined out to greater distances. However, larger systems would start to exhibit undulation modes that could affect the planar projection of the RDF. The dotted line in the same figure is the average ratio of R_z to R_{IJ}, where $R_{IJ} = [(M_{RI_x} - M_{RJ_x})^2 + (M_{RI_y} - M_{RJ_y})^2]^{1/2}$, $R_z = |M_{RI_z} - M_{RJ_z}|$, and $\mathbf{M}_{\mathbf{R}\alpha} = M_{R\alpha_x}\hat{\mathbf{i}} + M_{R\alpha_y}\hat{\mathbf{j}} + M_{R\alpha_z}\hat{\mathbf{k}}$. As long as

Hybrid Coarse Graining of Lipid Bilayers

J. Phys. Chem. B, Vol. 113, No. 13, 2009 **4417**

Figure 2. Lipid COM radial distribution function (RDF) (solid line), along with the average of the ratio R_z/R_{IJ} (dark dotted line), where R_{IJ} was defined previously, $R_z = |\mathbf{M}_{\mathbf{R}I_z} - \mathbf{M}_{\mathbf{R}J_z}|$, and $\mathbf{M}_{\mathbf{R}I_\alpha}$ is the α Cartesian component corresponding to the COM of lipid I.

Figure 3. CG RDF. The solid line is as given in Figure 2, the dark-dotted line is the CG RDF using a constant-area two-dimensional system, and the light-dashed line is the RDF for the HAS CG system in three dimensions under zero surface tension.

R_z/R_{IJ} is less than 1, then the correlations in the RDF originate from lipids that are nearly planar; however, when R_z/R_{IJ} is greater than 1 (i.e., around $R_{IJ} < 0.3$ nm), the correlations arise from particles with a large R_z component. From Figure 2, it is therefore clear that the short-range correlations arise from lipid COM that contain large nonplanar components, whereas the longer-range correlations arise from nearly planar structures. The two peaks around 1 nm can be loosely characterized as arising from in-plane COM correlations.

2.3. In-Plane Force Matching of the Lipid Bilayer. The COM mapping introduced previously can be used to force match the bilayer system. For this study, the delta-function basis set was used,[62] and the procedure is briefly described here. A more detailed account can be found elsewhere[62] and in Appendix A. The effective MS-CG force for the Ith CG site can be expressed as a pairwise sum over all J CG sites within a range delineated by the inner and outer preset cutoff distances $R_{C,1}$ and $R_{C,2}$, respectively, as

$$\mathbf{F}_I^{MS}(\mathbf{R}^N;\phi) = \sum_{\substack{I \neq J \\ R_{C,1} \leq R_{IJ} \leq R_{C,2}}} \mathbf{F}_{IJ}^{MS}(R_{IJ};\phi) \tag{5}$$

where the MS-CG pair force, $\mathbf{F}_{IJ}^{MS}$, is linearly dependent on N_d coefficients, $\phi = \{\phi_1,..., \phi_{N_d}\}$. The optimal value of the coefficients is found from the minimization of the χ^2 residual given by[59−61]

$$\chi^2(\phi) = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |\mathbf{F}_I^{MS}(\mathbf{M}_{\mathbf{R}}^N(\mathbf{r}_t^n);\phi) - \mathbf{f}_I(\mathbf{r}_t^n)|^2 \right\rangle_t \tag{6}$$

where the additional t subscript denotes a specific set of atomistic configurations from the MD trajectory and

$$\mathbf{f}_I(\mathbf{r}^n) = \sum_{i=1}^{n} P_{I,i}\mathbf{f}_i \tag{7}$$

The instantaneous atomic force on atom i is given by $\mathbf{f}_i$, and the average in eq 6 is taken over all CG sites and configurations.

It can be shown[59,60] that the minimum value of the residual is found when $\mathbf{F}_I^{MS} = \mathbf{F}_I \equiv -\nabla_{\mathbf{R}_I}U(\mathbf{R}^N)$, that is, when the MS-CG force field is in fact the gradient of the exact potential of mean force, $U(\mathbf{R}^N)$, that governs the CG system.[59,61] In a
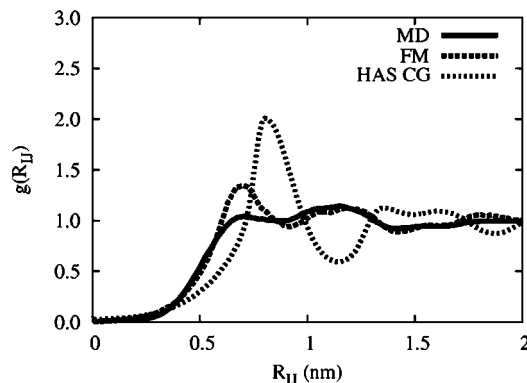
practical sense, the MS-CG force field reflects any degree of limited sampling in the underlying atomistic MD simulation. As discussed later, the analytic component of the interaction is designed to enhance and fill in the overall sampling range. In this study, the force is calculated only in the plane of the membrane. It might be possible to calculate the normal force, which would give the lipid tail−tail interaction. However, it is doubtful that either the head−head or head−tail cross interactions could be adequately sampled. Thus, the normal component is modeled through the analytic component. The two inner and outer cutoffs, $R_{C,1}$ and $R_{C,2}$, control the radial range to be used in the force-matching calculation. The inner cutoff, $R_{C,1}$, in particular, excludes specific data (i.e., data accumulated where the corresponding value of R_z/R_{IJ} is large or where the RDF is quite small) from the force-matching calculation. The outer cutoff, $R_{C,2}$, is determined from the size of the MD simulation and is given by the range of the RDF (in this case, 2 nm). The resulting MS-CG force field then gives the "best" force field subject to the range of inner and outer cutoffs employed. In previous work, the behavior of the CG potential at close range had to be interpolated.[15] This approach could also, in principle, be used here; however, as discussed below, the HAS CG approach circumvents this problem and adds additional flexibility to the CG model.

It was found that the MS-CG force field became increasingly unreliable (i.e., the condition number for the matrix inversion increased dramatically) when poorly sampled data obtained at close range was used. Therefore, an initial MS-CG force-field using the delta-function approach was found by including only data from 0.4 nm $< R_{IJ} <$ 2.0 nm. This discrete force field was then smoothed and directly employed in a two-dimensional CG simulation, similar to previous work,[46] to test the accuracy of this component of the HAS CG model relative to all-atom MD data. The resulting CG RDF is shown as the dark-dotted line in Figure 3. The slightly enhanced first peak is due to the sensitivity of the force field at close range and could be improved with a more elaborate treatment of the short-range interactions (i.e., versus simply omitting them in the force-matching calculation).

However, the end goal here is not to develop a purely two-dimensional force field, but to extend the model to three dimensions. As such, the focus now shifts to the analytic GB component of the HAS CG interaction.

2.4. Modified GB Lipid Bilayer Model. In this section, some details of the GB component of the HAS CG bilayer model are discussed. It should be noted that the GB interaction
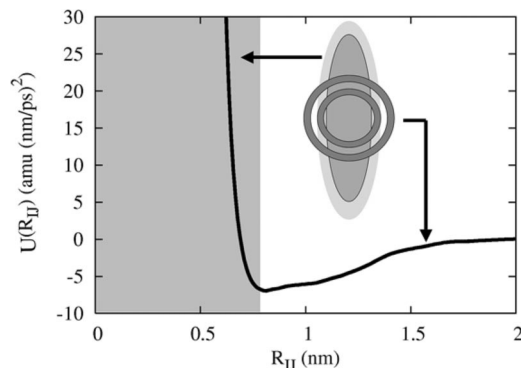
Figure 4. HAS CG potential. The inset shows a single CG particle, where the anisotropic component (the ellipsoid) corresponds to the core and the short-range spherically symmetric component comes from the MS-CG method and gives the attractive component of the interaction. The shaded region on the left corresponds to the potential arising from the superimposed anisotropic core region.

was modified such that it exhibits the symmetry of a lipid, i.e., the head is different from the tail. How this was done is described in Appendix B. The functional form of the GB pair potential is expressed as

$$U_{IJ}^{GB}(\mathbf{R}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = 4\varepsilon(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)(\varsigma_{ij}^{-12} - \varsigma_{ij}^{-6}) \quad (8)$$

where

$$\varsigma_{ij} = \frac{R_{IJ} - \sigma(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) + \sigma_0}{\sigma_0} \quad (9)$$

In this expression, $\hat{\mathbf{R}}_{IJ} = \mathbf{R}_{IJ}/R_{IJ}$ is the full three-dimensional vector. As mentioned previously, $\mathbf{e}_I$ is an orientation unit vector that lies along the long axis of the Ith GB ellipsoid; likewise for J. The quantity σ_0 is the width of the GB ellipsoid, and $\sigma(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)$ gives the "distance of closest approach" along $\hat{\mathbf{R}}_{IJ}$ for a pair of GB ellipsoids with orientations $\mathbf{e}_I$ and $\mathbf{e}_J$. See Appendix B for further details.

When two GB ellipsoids are constrained to be perfectly aligned and side-by-side, eq 8 reduces to

$$U_{IJ}^{GB,A}(R_{IJ}) = 4\varepsilon_0\left[\left(\frac{\sigma_0}{R_{IJ}}\right)^{12} - \left(\frac{\sigma_0}{R_{IJ}}\right)^6\right] \quad (10)$$

which is just the standard Lennard-Jones (LJ) interaction, where the additional superscript A refers to the fact that the GB ellipsoids are perfectly aligned and side-by-side. This is exactly the interaction that appears in eq 2. In this form, the interaction contains only an energy parameter, ε_0, as well as a fundamental unit of length, σ_0, and depends just on the pair distance between I and J, R_{IJ}. If the approximation is made that the lipids are perfectly aligned, then this form of the potential can be used in combination with the MS-CG force as described in subsections 2.2 and 2.3. It should be noted that explicitly including the angular dependence would dramatically increase the complexity of the MS-CG mapping. Initially setting σ_0 amounts to predefining the "size" of the CG particle, and it is empirically found from a knowledge of the atomistic area per lipid and the phase behavior of the GB ellipsoid system. However, this aspect falls under combining the GB and the MS-CG interaction and is discussed next.

2.5. HAS CG Lipid Bilayer Model. The first step in constructing the full HAS CG lipid model is to introduce a GB core; this is shown graphically in Figure 4 as the ellipsoid of revolution. This component of the model accounts for the highly anisotropic shape of a lipid. Equation 8 is employed with the following parametrization scheme: The fundamental unit of length, σ_0, was chosen based on the atomistic area per lipid, A_L, as found from the atomistic MD simulation ($A_L \approx 0.62$nm^2 per lipid), combined with the known density of the fluid phase of a GB ellipsoidal system ($\rho^* = A_L^{-1}\sigma_0^2 \approx 0.85$). A value of $\sigma_0 = 0.72$ nm was therefore chosen. The length of the ellipsoid was chosen to be $3\sigma_0$. The motivation for this 3:1 ratio draws on the extensive previous data on 3:1 GB ellipsoidal systems and the fact that this aspect ratio is known to give stable liquid-crystal and smectic phases.[69] This aspect ratio also reasonably models the bilayer thickness. Other approaches could also be employed; however, this approach reasonably combines the behavior of the atomistic and CG systems. The value of ε_0 was obtained based on an estimate of the MS-CG potential well depth and the requirement that the raw GB core should result in a stable bilayer. Values between 12 and 14 amu (nm/ps)^2 gave reasonable results. The other parameters in the GB model were as commonly employed.[69]

At this point, the nature of the HAS CG approach might seem rather ad hoc. However, it should be restated that the full HAS CG lipid model is designed such that the MS-CG component takes over and, in fact, dominates for specific configurations as expressed in eq 1. As such, the GB component is mainly being utilized in configurations that were never fully sampled in the initial MS-CG force-matching calculation.

The resulting full pair potential for two perfectly aligned and in-plane CG lipids is shown in Figure 4 and graphically depicts the pair interaction as given in eq 1. The core region is shaded, whereas the MS-CG component gives the critical attractive interaction. When two CG lipids are aligned and in-plane, the attractive component is exactly that determined from the MS-CG calculation. When they are in other orientations, the full pair potential contains an anisotropic GB component as well as the spherically symmetric MS-CG part. However, because the MS-CG part is fairly short-range (and, in fact, is zero beyond $R_{C,2}$), the end-to-end interaction of two CG lipids is entirely given by the GB component. As such, the sum of the two contributions can be illustrated by the inset in Figure 4; the GB component gives an analytic model for the anisotropic part of the system, whereas the MS-CG component gives a spherically symmetric part that comes fully into effect when two CG lipids are aligned and in-plane.

3. Results

The CG simulation results are separated into two sets of simulations: a relatively small square "patch" membrane ($N = 5000$ CG sites) and a much larger 190-nm-diameter liposome ($N = 380\,000$ CG sites). The smaller system is employed to measure a number of key membrane properties, whereas the liposome simulation is used to demonstrate the feasibility of large-scale CG simulation employing the HAS CG approach. The HAS CG pair potential as represented in Figure 4 and eq 1 is employed throughout. An $N = 1250$ CG site "raw GB system" that employs an interaction as in eq 1 but with $\Delta U(R_{IJ}) = 0$ is also employed for comparison when appropriate.

3.1. $N = 5000$ CG Lipid System. A square patch of CG bilayer was constructed using a total of $N = 5000$ CG lipids, with each monolayer containing half of the CG lipids. This CG bilayer system is equivalent to a united-atom lipid model with
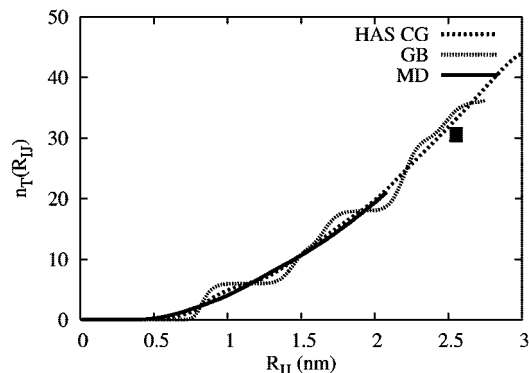
Figure 5. Accumulated average number of lipids in one of the membrane leaflets for the $N = 5000$ CG site HAS membrane (dotted), the raw GB (hatched), and the COM atomistic MD simulation (solid line). The solid square is the estimated number of lipids around a central one based on the atomistic area per lipid.

TABLE 1: Key CG Simulation Parameters

parameter	symbol	units	value
GB width	σ_0	nm	0.72
GB length	h	nm	2.16
Temperature	T	K	308
time step	δt	ps	0.01
GB mass	m	amu	678
GB energy	ε_0	amu (nm/ps)^2	12−14

230 000 sites or a fully all-atom model with 588 000 lipid atoms and around 1.5 million total atoms with water solvent included. Multiple simulations were performed under zero surface tension for up to 200 ns. Key simulation parameters are listed in Table 1. First, the in-plane RDF for the CG lipid model is shown as the light-dashed line in Figure 3. Clearly, the inclusion of the GB core in the potential in three dimensions has altered the short-range structure as compared to the original atomistic (solid line) and two-dimensional MS-CG (dark-dotted line) results. However, the new correlations are not that strong, and importantly, under zero surface tension, the correct area per lipid is recovered. It is to be expected that some degree of accuracy in the structural correlations will be lost in aggressive coarse-grained models such as this one.

In Figure 5 is shown the accumulated number of CG lipids, $n_T(R_{IJ})$, around a central lipid in one of the monolayers. The solid line is the original atomistic result and is directly related to the area per lipid in the atomistic MD simulation. The dotted line is the CG lipid result using the HAS CG potential as given in eq 1. The accumulated CG lipid density around a central CG lipid is seen to track well to the atomistic result. Furthermore, it continues to accumulate density correctly. The solid square is the predicted number of CG lipids based on the atomistic area per lipid. Because the CG lipid simulation is under zero surface tension, it could easily elect to adopt a structure such that the CG area per lipid was quite different from the atomistic result. The raw GB result is also shown, and it exhibits pronounced oscillations because of the strong hexagonal correlations in that system. Thus, at short range, the HAS interaction modulates the overstructured GB system such that it much better reproduces the atomistic MD result.

It should be noted that the CG area per lipid as calculated from the ratio of the CG simulation cell area to half the number of CG sites was found to be $A_L^{CG} = 0.59 \pm 0.2$ nm^2 per lipid which is slightly less than the MD value of $A_L = 0.62$ nm^2 per lipid; this is due to undulation modes that begin to emerge in the larger system (cf. Figure 6). These undulation modes are
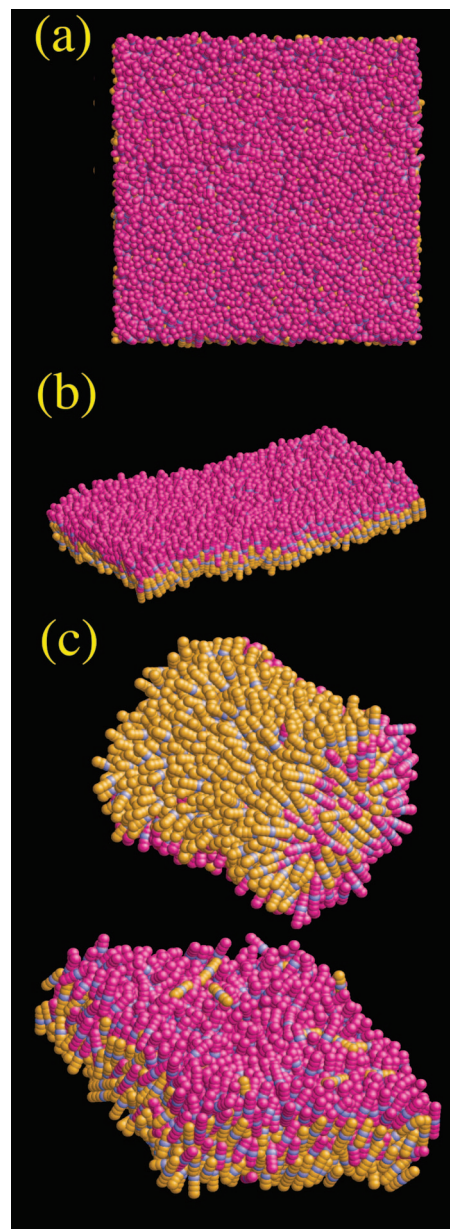
Figure 6. Snapshots of the $N = 5000$ CG site square-patch HAS CG bilayer. (a) Top-down view, in which very little distinct structural correlation is evident. (b) Side view, in which thermal undulations are clearly visible. (c) Two different self-assembled structures coming from an $N = 2048$ CG site system. Both bicellar and bilayer structures were observed depending on the volume of the system.

clearly evident in Figure 6b where the CG lipids are rendered as spherocylinders with an aspect ratio of $h/\sigma_0 = 3$. The top view in Figure 6a also shows a distinct lack of hexagonal correlations, similar to the atomistic structure in Figure 1. It should be noted that, when only the GB interaction was used in eq 1 [i.e., with $\Delta U(R_{IJ}) = 0$], the CG bilayer actually froze into a tightly packed hexagonal lattice. It is clear, therefore, what the MS-CG component is doing, above and beyond the analytic GB core of the HAS CG model: it is bringing back critical interactions that actually "melt" the otherwise solid membrane and, accordingly, give it a structure much more reminiscent of the original atomistic MD membrane. Furthermore, returning to Figure 3, it is clear that, in the CG lipid system, the close-range structure (i.e., configurations where lipids are close to being on top of each other) actually re-emerged in the CG system despite the inclusion of the GB
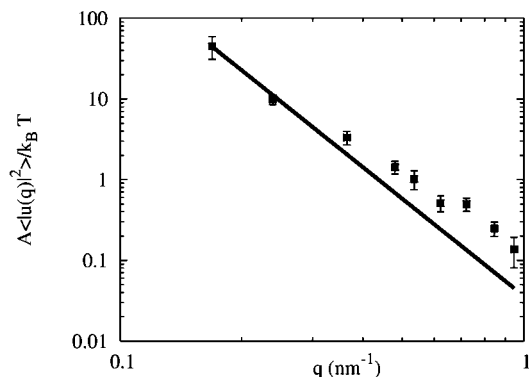
Figure 7. Undulation spectrum for an $N = 5000$ CG site square patch of HAS (solid squares) CG membrane under zero surface tension. Error bars are from block averages. The bending modulus for the HAS membrane using the smallest q values was found to be $k_c = (4.6 \pm 0.2) \times 10^{-20}$ J, while the modulus for the raw GB membrane was significantly larger at $k_c = 8.0 \times 10^{-20}$ J. For comparison, the experimental bending modulus for DMPC is $k_c = (3-6) \times 10^{-20}$ J.

core. When compared to the snapshots in Figure 6, the disordered nature of the system is consistent with the correlations in the RDF.

A key part of the HAS CG model relies on bridging the MS-CG force field to the GB interaction under the ideal configuration where the CG lipids are perfectly aligned and in a plane. The validity of this approximation can be tested by examining the orientational order of the CG lipid bilayer. If the CG lipids are highly aligned, then a large part of their interaction is coming from the systematically determined MS-CG component of the overall interaction. This can be measured via the P_2 order parameter, defined as $P_2 = \langle {}^1\!/_2 (3 \cos^2 \theta - 1) \rangle$, where θ gives the angle between the symmetry axis of the CG lipid and the director, $\mathbf{d}$.[72] Values of the P_2 order parameter around $0-0.2$ indicate basically an isotropic system, whereas values around 0.8 indicate a highly orientationally ordered phase. Here, the P_2 order parameter was found to be $P_2 = 0.79 \pm 0.004$, indicating that the CG lipids were quite aligned. For reference, the same quantity calculated from the original MD simulation was found to be $P_2 = 0.76 \pm 0.010$ (where the lipid orientation vector was found from the corresponding radius of gyration tensor). When combined with the fact that a stable bilayer structure is observed, it is reasonable to conclude that a large part of the CG lipid interaction is directly coming from the systematically force-matched MS-CG component.

Self-assembly has been observed in a number of CG models, resulting in bicelles, vesicles, and planar bilayers.[19,20,26,27,37,44] Although the current study simulates liposomes at much larger length scales, it is interesting to examine the self-assembly properties of the single-site HAS CG model. A low-density system ($N/V = 0.2$ nm^{-3}) of $N = 2048$ HAS CG particles was initially annealed at 400 K. This isotropic configuration ($P_2 \approx 0.02$) was then used at various other densities up to $N/V = 0.27$ nm^{-3}. The resulting self-assembled structures are shown in Figure 6c. Bicelles as well as bilayers are seen to spontaneously form. When only the raw GB interaction was employed, then only lamellar bilayers spontaneously formed.

A critical membrane material property is the bending modulus, k_c, which can, in principle, be found in the region of small wavevector, q, from membrane undulation fluctuations,[20,29,73,74] although care must be taken in interpreting the resulting undulation spectrum.[19,20,22,25] Recent studies have shown that wavevectors with $(q\sigma_0)^2 \lesssim 0.1$ need to be accessed in order for the $1/q^4$ scaling to be observed.[19] Furthermore, it has been shown

that membrane protrusions can also alter the undulation spectrum.[22] Here, the boundary conditions of the CG lipid system are such that a square patch of membrane is bound in the xy plane, and thermal undulations, $u(\mathbf{r},t)$ occur in the z direction. Under these boundaries and for $(q\sigma_0)^2 \lesssim 0.1$, the bending modulus can be measured using the undulation spectrum of $\langle |u(q)|^2 \rangle = \langle u(\mathbf{q},t) u(-\mathbf{q},t) \rangle$, where $\mathbf{q}$ is a wavevector, $u(\mathbf{q},t)$ is the Fourier mode of the undulation height $u(\mathbf{r},t)$, and

$$u(\mathbf{q}, t) = \frac{1}{A} \int dA\, u(\mathbf{r}, t) \exp(i\mathbf{q} \cdot \mathbf{r})$$

$$u(\mathbf{r}, t) = \sum_{\mathbf{q}} u(\mathbf{q}, t) \exp(-i\mathbf{q} \cdot \mathbf{r}) \tag{11}$$

Equipartition under zero surface tension gives

$$\langle |u(q)|^2 \rangle = \frac{k_B T}{A k_c q^4} \tag{12}$$

Figure 7 shows the quantity $\langle |u(q)|^2 \rangle A / k_B T$ versus $q = |\mathbf{q}|$. For reference, the experimental measurement of the bending modulus for DMPC is $(3-6) \times 10^{-20}$ J.[75,76] The resulting low-q CG lipid undulation magnitudes are quite reasonable. The bending modulus was estimated by fitting the undulation spectrum according to eq 12 using only those wavevectors that satisfies $(q\sigma_0)^2 \leq 0.1$. The bending modulus for the HAS membrane was found to be $k_c = 4.6 \pm 0.2 \times 10^{-20}$ J, whereas the modulus for the raw GB membrane was significantly larger at around $k_c = 8 \times 10^{-20}$ J. Deviations from the $1/q^4$ scaling are expected to contribute at higher wavevectors, and a more complex analysis is required in this regime.[22] It should be noted that the bending modulus from the atomistic MD system was not determined as it was too small to access low-wavelength undulation modes, with a lowest q value of about 1 nm^{-1}. Thus, the bending modulus of the CG lipid membrane is entirely a predicted quantity and results from the HAS CG model proposed in eq 1. It should be noted that tests with different values of ε_0 had little or no effect on the bending modulus. This results from the design of the HAS CG model and the fact that that most of the in-plane forces arise from the MS-CG component and thus appear to dominate the membrane's material properties. However, the GB component is responsible for "holding" the structure of the CG lipid bilayer together in a nontrivial way. Interestingly, even though the atomistic MD system that was used for force matching was too small to bend, the resulting CG model, when extended to larger system sizes, resulted in a bending modulus that was consistent with the underlying MD DMPC model from which it was originally derived (assuming that the atomistic DMPC model would give the correct bending modulus for a large enough system). The undulation dynamics of this solvent-free CG model would, of course, miss critical hydrodynamic dampening due to interactions of the membrane with the surrounding viscous solvent. This behavior is characteristic of any solvent-free model where the membrane essentially exists in a vacuum with effective interactions between the lipids mimicking the effects of the solvent. A much more involved model would be required to correctly incorporate membrane hydrodynamics.

The area expansion modulus, K_A, defined via $(\partial\gamma/\partial A)_T = K_A/A^*$, where γ is the surface tension, A is the area, and A^* is the average area found under a starting state of zero surface

Hybrid Coarse Graining of Lipid Bilayers

J. Phys. Chem. B, Vol. 113, No. 13, 2009 **4421**
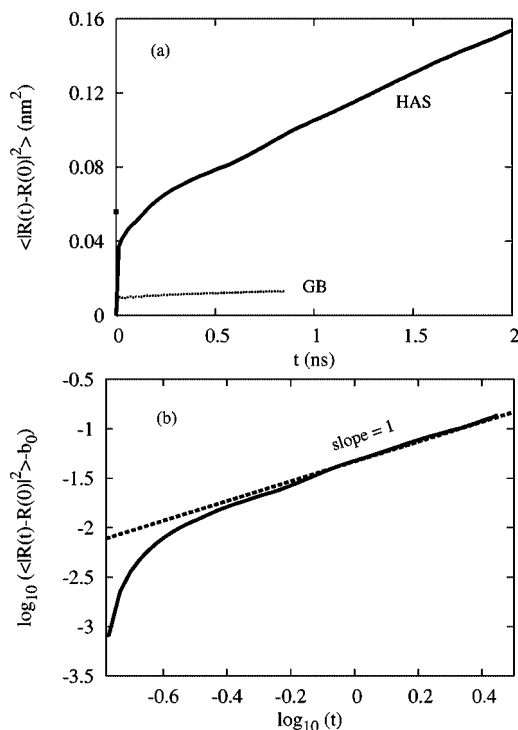
Figure 8. (a) Two-dimensional mean square displacement (MSD), $\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle$ for the HAS (solid line) and raw GB (dotted) systems. The corresponding diffusion coefficient is found from a linear fit, $\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle = 4Dt + b_0$, of the MSD for $t > 0.8$ ns for the HAS bilayer and is $D = 1.2 \times 10^{-7}$ cm²/s, compared to the experimental estimate of 1.10×10^{-7} cm²/s. The $t = 0$ intercept for this fit is $b_0 = 0.057$ nm², as shown by the small square. The corresponding raw GB bilayer is frozen. (b) Plot of $\log_{10}[\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle - b_0]$ versus $\log_{10}(t)$, which is predicted to have a slope of unity at long times if the linear fit proposed in (a) holds.

tension,[77] was found to be ~138 mJ/m², in good agreement with experimental value of 145 mJ/m².[76] It should be noted that the area compressibility modulus of the small atomistic system was estimated at ~158 mJ/m². Apparently, the larger CG system under zero surface tension, with the appearance of soft thermal modes, slightly modifies the atomistic result. Given that K_A can vary by as much as a factor of 6 for different membranes, it is encouraging that the present HAS model can incorporate this property at such a high degree of coarse-graining.

Figure 8a shows the two-dimensional mean square displacement (MSD), $\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle$, which is related to the two-dimensional diffusion coefficient, D, by $\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle = 4Dt$. Results for both the HAS model and the raw GB membrane are shown. It appears that the HAS CG bilayer is indeed fluid, with a diffusion coefficient of around 1.2×10^{-7} cm²/s, which is actually not far from the experimental value of around 1.10×10^{-7} cm²/s.[76] The diffusion constant was found from a linear fit to the data above $t = 0.8$ ns. In contrast, the raw GB membrane is frozen, indicating once again that the effect of the HAS interaction is to melt the otherwise frozen GB bilayer. Figure 8b shows the quantity $\log_{10}(\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle - b_0)$ versus $\log_{10}(t)$, which is expected to have a slope of 1 when $\langle|\mathbf{R}(t) - \mathbf{R}(0)|^2\rangle \approx 4Dt$ at long times. Indeed, this behavior appears to recovered for $t > 0.8$ ns, indicating that the dynamics is diffusive. However, as has been noted previously, comparing CG and atomistic dynamical quantities only really has meaning when the correct frictional and drag forces are incorporated into the CG scheme;[78] in this case, the fact that the system appears fluid (i.e., the MSD increased linearly) is sufficient. (Frictional forces would tend to decrease the diffusion constant and bring it into

even better agreement with the experimental value.) We also note that we have made no attempt to "rescale" time in these CG simulations.

3.2. CG Simulation of a Liposome. With the in-plane fluid nature and the membrane bending modulus confirmed from the $N = 5000$ simulation, a much larger HAS CG liposome simulation was performed. In this case, the liposome was placed in the center of a $220 \times 220 \times 220$ nm³ simulation box, and the initial liposome diameter was 190 nm. A total of 379 858 CG lipids were employed; the outer leaflet contained 195 312 CG lipids, and the inner leaflet contained 184 546 lipids. This HAS CG simulation is equivalent to a united-atom lipid model with 17 million atoms or about 1.1×10^9 atoms when a water solvent and an all-atom lipid model are employed. The difference between the two leaflets ensured that the CG lipid density in each monolayer was the same. The simulation was performed in a solvent-free mode over 512 processors using our in-house code TANTALUS.[79] It was observed that about 2 ns of simulation time could be achieved over 10 clock hours even with this moderate level of parallelization. Given that fully solvated CG simulations at these length scales are not currently possible, this type of scaling is encouraging; scaling over 1000 processors is entirely possible with a more advanced parallelization scheme. The aim of these simulations was to assess the stability of the liposome and to ensure that its membrane structure was similar to that observed with the smaller $N = 5000$ simulation.

Panels c and d of Figure 9 show snapshots of the liposome after 60 ns of simulation. The first two panels, a and b, give an estimate of the relative length scales of the original atomistic system (as in Figure 1a) and the $N = 5000$ CG lipid system (as in Figure 6), respectively, as compared to a closeup of the liposome surface. When the scale bars are compared relative to the entire liposome, it becomes clear how much of a jump in length scale has been achieved in the HAS CG simulation; the original MD system would be not much more than a small spot on the liposome surface. After 60 ns, the liposome structure remains intact, except for visible thermal undulations that have appeared all over the surface. The final average diameter of the liposome was 192 ± 2 nm, where the error arises from thermal undulations. The slightly larger diameter results from the CG lipids not being perfectly aligned at finite temperature; thus, the liposome swells slightly. The final area per CG lipid evaluated over the surface area of the liposome and averaged over the two leaflets was found to be 0.60 ± 0.02 nm²/CG lipid, indicating that the local structure in the liposome simulation was consistent with that in the smaller $N = 5000$ HAS CG study.

At this point, at least the shorter-time stability of the HAS CG liposome over a reasonable simulation time scale is confirmed. Longer simulations will be performed in the future when more complex scenarios (e.g., protein-mediated membrane remodeling) are to be examined. The feasibility of the HAS CG approach is therefore established, and the overall approach looks promising in systems where aggressive coarse graining is required.

It should be noted that after 60 ns of CG simulation time, about 60 lipids had swapped leaflets (out of the ~190 000 CG lipids per leaflet). It might be possible to pursue a more thorough lipid flip-flop rate analysis;[19] however, given the ambiguity of what CG time actually means, the limited statistics, and the fact that some lipids actually leave the liposome, the finding that lipids do seems to swap leaflets over long time scales is taken to be encouraging. A rough estimate for the lipid flip-flop rate is about 5×10^{-6} ns^{-1}, consistent with other CG models.[19]
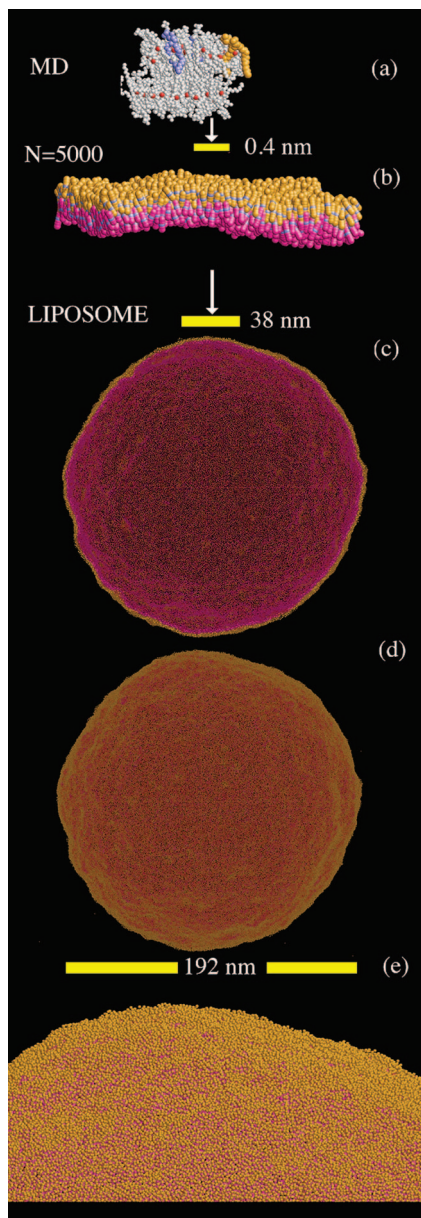
Figure 9. Relative length scales eventually reaching the HAS CG liposome with $N = 379\,858$ total CG sites. (a) Original MD system as in Figure 1. (b) Snapshot of the $N = 5000$ HAS membrane. (c) Slice of the HAS liposome after 60 ns of simulation. (d) Entire surface of the HAS liposome after 60 ns of simulation. The lipids are colored as in Figure 6. The different scales are shown in the yellow scale bars. (e) Closeup in which the liposome surface is clearly visible. The area per CG lipid evaluated over the surface area of the liposome was found to be 0.60 ± 0.02 nm^2/CG lipid. The lipid flip-flop rate was estimated at around 5×10^{-6} ns^{-1}, based on a count of the number of lipids that had moved from one leaflet to the other during the course of the simulation.

A question arises as to whether 60 ns of CG simulation time is sufficient to equilibrate the liposome. As previously noted, CG time does not correspond to real time; the highly accelerated lipid flip-flop rate supports this notion and suggests that the system has effectively sampled longer time scales. Ideally, the duration of the simulation should be such that deformation modes corresponding to all accessible wavelengths of the liposome are fully sampled.[80−82] Of course, the long-wavelength undulation modes evolve much more slowly than the fast ones, and it is reasonable to assume that, for a system of this size, the long-wavelength modes have not yet been fully sampled in

the simulations presented here. At very large length scales, e.g., vesicles, the undulations due to thermal fluctuations in fact become subvisible and act more as a reservoir for increases in the apparent surface area of the vesicle.[83] However, within the context of specific biomolecular processes, e.g., membrane remodeling via BAR domains, simulations of the duration reported in this article could be employed to examine to examine the early stages of protein-induced membrane remodeling.

4. Conclusions

The aim of this article was to develop an aggressively coarse-grained, yet systematically obtained CG lipid model. A CG approach was therefore presented in which MS-CG with force matching was utilized to generate a CG force field that dominates in certain configurations that are well-sampled at the atomistic level. An analytic force field based on the Gay−Berne liquid-crystal model was then used to model the anisotropic core of the CG lipid, and the final hybrid analytic−systematic (HAS) CG force field was obtained by combining the two components.

The resulting HAS CG lipid model is able to reproduce key bilayer quantities such as the bending modulus and lipid diffusion, all within a solvent-free CG simulation methodology that affords a substantial computational speedup as compared to the original atomistic MD model. When extended to much longer length scales (e.g., a liposome), the HAS CG lipid model can be employed to simulate a liposome. Here, a 60-ns CG simulation of a liposome with a starting diameter of 190 nm was performed as a demonstration of the method. The bilayer structure of the liposome remained intact, including the emergence of thermal undulation modes.

The HAS CG methodology proposed here will become an integral part of a more fully enriched CG model in the future aimed at examining real biological membranes and processes such as membrane remodeling.

Appendix A: Normal Solution and Delta Function Basis

Begin with the residual as given in eq 6 where $\mathbf{F}_{IJ}^{MS}(\mathbf{R}_I, \mathbf{R}_J, \phi)$ is given by

$$\mathbf{F}_{IJ}^{MS}(\mathbf{R}_I, \mathbf{R}_J, \phi) = \sum_{d=1}^{N_d} \phi_d \delta_C (R_d - R_{IJ}) \hat{\mathbf{R}}_{IJ} \quad (A1)$$

and $\delta_C = 1$ if $R_d \leq R_{IJ} < R_{d+1}$ and 0 otherwise. A set of linear equations is found from setting $\partial \chi^2 / \partial \phi_{d'} = 0$ for each $d' = 1,...,$ N_d. It should be noted that the $d = 1$ bin at $R_{C,1}$ could be set at any desired location. For example, $R_{C,1}$ could be set at the distance corresponding to the first nonzero entry in the RDF (see Figure 2). It could also be set at other distances, and a different solution would then be found for that set of coefficients.

Appendix B: Anisotropic Gay−Berne Lipid Bilayer Model

The Gay−Berne (GB) liquid-crystal model has been discussed in detail elsewhere;[68,69] here, an anisotropic version is presented where the two "ends" of the GB ellipsoids interact differently. By breaking the symmetry of the GB model, molecules such

Hybrid Coarse Graining of Lipid Bilayers

J. Phys. Chem. B, Vol. 113, No. 13, 2009 **4423**

as lipids, which have headgroups at one end and hydrocarbon chains at the other, can be modeled.

The GB interaction energy for a pair of ellipsoids I and J is given by eq 8. The quantity σ_0 is the width of the GB particle, and $\sigma(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)$ gives the distance of closest approach along $\hat{\mathbf{R}}_{IJ}$ for a pair of GB particles with orientations $\mathbf{e}_I$ and $\mathbf{e}_J$ and is given by

$$\sigma(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = \sigma_0\left\{1 - \frac{\chi}{2}\left[\frac{(\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I + \hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_J)^2}{1 + \chi\mathbf{e}_I\cdot\mathbf{e}_J} + \frac{(\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I - \hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_J)^2}{1 - \chi\mathbf{e}_I\cdot\mathbf{e}_J}\right]\right\}^{-1/2} \quad \text{(B1)}$$

where $\chi = (\kappa^2 - 1)/(\kappa^2 + 1)$ and $h = \kappa\sigma_0$. Here, h is the length of the GB particle along the symmetry axis. This component of the GB interaction models the ellipsoidal shape of the molecule. The energy component, $\varepsilon(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)$, can be easily modified to break the symmetry of the GB model; one scheme is presented here. The energy component is given by

$$\varepsilon(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = \varepsilon_0(1 - \chi^2)^{1/2}\varepsilon_1(\mathbf{e}_I, \mathbf{e}_J)[\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)]^2$$
$$\varepsilon_1(\mathbf{e}_I, \mathbf{e}_J) = [1 - \chi^2(\mathbf{e}_I\cdot\mathbf{e}_J)^2]^{-1/2} \quad \text{(B2)}$$

where, in order to make the ends of two GB particles interact differently, the last term, $\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J)$, can be slightly modified to

$$\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = 1 - \frac{\chi'}{2}\left[\frac{(\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I + \hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_J)^2}{1 + \chi'\mathbf{e}_I\cdot\mathbf{e}_J} + \frac{(\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I - \hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_J)}{1 - \chi'\mathbf{e}_I\cdot\mathbf{e}_J}\right] \quad \text{(B3)}$$

and the last term is no longer squared (as it usually is in a regular GB interaction model). To see how this works, first consider two GB particles that are side-by-side and parallel: In this case, $\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = 1$. Now consider two GB particles end-to-end and with $\mathbf{e}_I\cdot\mathbf{e}_J = 1$; $\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = (1 - \chi')/(1 + \chi')$. With $\mathbf{R}_{IJ} = \mathbf{R}_I - \mathbf{R}_J$, a configuration with $\mathbf{e}_I\cdot\mathbf{e}_J = -1$ and $\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I = 1$ has $\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = 1/(1 + \chi')$, whereas a configuration with $\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I = -1$, has $\varepsilon_2(\hat{\mathbf{R}}_{IJ}, \mathbf{e}_I, \mathbf{e}_J) = (1 + 2\chi')/(1 + \chi')$. The commonly employed value of χ' is given by $\chi' = [(\varepsilon_{ss}/\varepsilon_{ee})^{1/2} - 1]/[(\varepsilon_{ss}/\varepsilon_{ee})^{1/2} + 1]$, where $\varepsilon_{ss}/\varepsilon_{ee}$ gives the ratio of the well depths for two side-by-side and end-to-end particles and is typically given a value between 3 and 5.[68,69] A value of 5 was chosen here to match previous studies.[69] Thus, this broken-symmetry GB model gives the deepest end-to-end well depth when $\mathbf{e}_I\cdot\mathbf{e}_J = -1$ and $\hat{\mathbf{R}}_{IJ}\cdot\mathbf{e}_I = -1$. In terms of a lipid bilayer, this configuration corresponds to a configuration in which the lipid tails are in contact and the headgroups are at opposite ends.

References and Notes

(1) Singer, S. J.; Nicolson, G. L. *Science* **1972**, *175*, 720.
(2) Engelman, D. M. *Nature* **2005**, *438*, 578.
(3) McLaughlin, S.; Murray, D. *Nature* **2005**, *438*, 605.
(4) Dupuy, A. D.; Engelman, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2848.
(5) Veatch, S. L.; Keller, S. L *Biophys. J.* **2003**, *85*, 3074.
(6) Peter, B. J.; Kent, H. M.; Mills, I. G.; Vallis, Y.; Butler, P. J. G.; Evans, P. R.; McMahon, H. T. *Science* **2004**, *303*, 495.
(7) Gallop, J. L.; Jao, C. C.; Kent, H. M.; Butler, P. J. G.; Evans, P. R.; Langen, R.; McMahon, H. T *EMBO J.* **2006**, *25*, 2898.

(8) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192.
(9) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press/Taylor and Francis Group: Boca Raton, FL, 2009.
(10) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464.
(11) Marrink, S. J.; deVries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.
(12) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; deVries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.
(13) Stevens, M. J. *J. Chem. Phys.* **2004**, *121*, 11942.
(14) Faller, R.; Marrink, S. J. *Langmuir* **2004**, *20*, 7686.
(15) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
(16) Izvekov, S.; Voth, G. A. *J. Chem. Theor. Comp.* **2006**, *2*, 637.
(17) Farago, O. *J. Chem. Phys.* **2003**, *119*, 596.
(18) Cooke, I. R.; Kremer, K.; Deserno, M. *Phys. Rev. E* **2005**, *72*, 011506.
(19) Cooke, I. R.; Deserno, M. *J. Chem. Phys.* **2005**, *123*, 224710.
(20) Brannigan, G.; Brown, F. L. H. *J. Chem. Phys.* **2004**, *120*, 1059.
(21) Brannigan, G.; Lin, L. C. L.; Brown, F. L. H. *Eur. Biophys. J.* **2006**, *35*, 104.
(22) Brannigan, G.; Brown, F. L. H. *Biophys. J.* **2006**, *90*, 1501.
(23) Cooke, I. R.; Deserno, M. *Biophys. J.* **2006**, *91*, 487.
(24) Reynwar, B. J.; Illya, G.; Harmandaris, V. A.; Muller, M. M.; Kremer, K.; Deserno, M. *Nature* **2007**, *447*, 461.
(25) Harmandaris, V. A.; Deserno, M. *J. Chem. Phys.* **2006**, *125*, 204905.
(26) Lyubartsev, A. P. *Eur. J. Biophys.* **2005**, *35*, 53.
(27) Markvoort, A. J.; Pieterse, K.; Steijaert, M. N.; Spijker, P.; Hilbers, P. A. J. *J. Phys. Chem. B* **2005**, *109*, 22649.
(28) Markvoort, A. J.; van Santen, R. A.; Hilbers, P. A. J. *J. Phys. Chem. B* **2006**, *110*, 22780.
(29) Goetz, R.; Gompper, G.; Lipowsky, R. *Phys. Rev. Lett.* **1999**, *82*, 221.
(30) Goetz, R.; Lipowsky, R. *J. Chem. Phys.* **1998**, *108*, 7397.
(31) Kumar, P. B. S.; Gompper, G.; Lipowsky, R. *Phys. Rev. Lett.* **2001**, *86*, 3911.
(32) Shillcock, J. C.; Lipowsky, R. *NIC Symp. Proc.* **2001**, *9*, 407.
(33) Shillcock, J. C.; Lipowsky, R. *J. Chem. Phys.* **2002**, *117*, 5048.
(34) Shillcock, J. C.; Lipowsky, R. *Nat. Mater.* **2005**, *4*, 225.
(35) Michel, J.; Orsi, M.; Essex, J. W. *J. Phys. Chem. B* **2008**, *112*, 657.
(36) Orsi, M.; Haubertin, D. Y.; Sanderson, W. E.; Essex, J. W. *J. Phys. Chem. B* **2008**, *112*, 802.
(37) Yamamoto, S.; Maruyama, Y.; Hyodo, S. *J. Chem. Phys.* **2002**, *116*, 5842.
(38) Ayton, G.; Bardenhagen, S.; McMurtry, P.; Sulsky, D.; Voth, G. A. *J. Chem. Phys.* **2001**, *114*, 6913.
(39) Risselada, H. J.; Mark, A. E.; Marrink, S. J. *J. Phys. Chem. B* **2008**, *112*, 7438.
(40) Marrink, S. J.; Risselada, H. J.; Mark, A. E. *Chem. Phys. Lipids* **2005**, *135*, 223.
(41) Risselada, H. J.; Marrink, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17367.
(42) Knecht, V.; Marrink, S. J. *Biophys. J.* **2007**, *92*, 4254.
(43) Marrink, S. J.; Mark, A. E. *J. Am. Chem. Soc.* **2003**, *125*, 11144.
(44) Marrink, S. J.; Mark, A. E. *J. Am. Chem. Soc.* **2003**, *125*, 15233.
(45) Murtola, T.; Falck, E.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2007**, *126*, 075101.
(46) Murtola, T.; Falck, E.; Patra, M.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2004**, *121*, 9156.
(47) Arkhipov, A.; Yin, Y.; Schulten, K. *Biophys. J.* **2008**, *95*, 2806.
(48) Ayton, G. S.; Blood, P. D.; Voth, G. A. *Biophys. J.* **2007**, *92*, 3595.
(49) Brown, F. L. H. *Biophys. J.* **2003**, *84*, 842.
(50) Lin, L. C.-L; Brown, F. L. H. *Biophys. J.* **2004**, *86*, 764.
(51) Lin, L. C. L.; Brown, F. L. H. *Phys. Rev. E* **2005**, *72*, 011910.
(52) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. *J. Chem. Theory Comput.* **2008**, *4*, 819.
(53) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.
(54) Smondyrev, A. M.; Berkowitz, M. L. *J. Comput. Chem.* **1999**, *20*, 531.
(55) Hiroshi, N.; Gerhard, G. *J. Chem. Phys.* **2006**, *125*, 164908.
(56) Blood, P. D.; Voth, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15068.
(57) Takei, K.; Slepnev, V. I.; Haucke, V.; De Camilli, P. *Nat. Cell Biol.* **1999**, *1*, 33.
(58) Masuda, M.; Takeda, S.; Sone, M.; Ohki, T.; Mori, H.; Kamioka, Y.; Mochizuki, N. *EMBO J.* **2006**, *25*, 2889.
(59) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Anderson, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(60) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
(61) Ayton, G. S.; Noid, W. G.; Voth, G. A. *MRS Bull.* **2007**, *32*, 929.

(62) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116.

(63) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(64) Wang, Y.; Izvekov, S.; Yan, T.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 3564.

(65) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289.

(66) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B*, manuscript submitted.

(67) Thorpe, I.; Zhou, J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 13079.

(68) Gay, J. G.; Berne, B. J. *J. Chem. Phys.* **1981**, *74*, 3316.

(69) Brown, J. T.; Allen, M. P. *Phys. Rev. E* **1998**, *57*, 6685.

(70) De Miguel, E.; Rull, L. F.; Chalam, M. K.; Gubbins, K. E. *Mol. Phys.* **1991**, *74*, 405.

(71) Ayton, G.; Smondyrev, A. M.; Bardenhagen, S.; McMurtry, P.; Voth, G. A. *Biophys. J.* **2002**, *82*, 1226.

(72) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, U.K., 1987.

(73) Lindahl, E.; Edholm, O *Biophys. J.* **2000**, *79*, 426.

(74) Marrink, S. J.; Mark, A. E. *J. Phys. Chem. B* **2001**, *105*, 6122.

(75) Kucerka, N.; Liu, Y.; Chu, N.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F *Biophys. J.* **2005**, *88*, 2626.

(76) Lipowsky, R.; Sackmann, E. *Structure and Dynamics of Membranes*; North-Holland: Amsterdam, The Netherlands, 1995; Vol. 1A.

(77) Feller, S. E.; Pastor, R. W. *J. Chem. Phys.* **1999**, *111*, 1281.

(78) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2006**, *125*, 151101.

(79) Ayton, G. S.; Voth, G. A. *Int. J. Multiscale Comput. Eng.* **2004**, *2*, 291.

(80) Drouffe, J.-M.; Maggs, A. C.; Leibler, S. *Science* **1991**, *254*, 1353.

(81) Gompper, G.; Kroll, D. M. *J. Phys.: Condens. Matter* **2000**, *12*, A29.

(82) Milner, S. T.; Safran, S. A. *Phys. Rev. A* **1987**, *36*, 4371.

(83) Rawicz, W.; Olbrich, K. C.; McIntosh, T.; Needham, D.; Evans, E. *Biophys. J.* **2000**, *79*, 328.