JCTC Journal of Chemical Theory and Computation

Multiscale Coarse-Graining of Mixed Phospholipid/ Cholesterol Bilayers

Sergei Izvekov and Gregory A. Voth*

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

Received November 30, 2005

Abstract: Coarse-grained (CG) models for mixed dimyristoylphosphatidylcholine (DMPC)/ cholesterol lipid bilayers are constructed using the recently developed multiscale coarse-graining (MS-CG) method. The MS-CG method permits a systematic fit of the bonded and nonbonded interactions and system pressure to trajectory and force data derived from an underlying reference all-atom molecular dynamics (MD) simulation. The CG sites for lipid and cholesterol molecules are associated with the centers-of-mass of atomic groups because of the simplicity in the evaluation of the forces acting on them from the atomistic MD data. Corresponding models with four-site and seven-site representations of the cholesterol molecule were also developed. The latter CG models differed by the bonding scheme of CG sites to represent intramolecular interactions. A one-site MS-CG model based on the TIP3P potential was used for water, with the interaction site placed at the molecular geometrical center, and the analytical fit of the model is presented. The MS-CG models were then used to conduct simulations in the constant NPT ensemble which reproduce accurately the structural properties as seen in the full all-atom MD simulation.

I. Introduction

Cell membranes are remarkably adaptive and durable molecule aggregates which are directly involved in a large variety of cecullar processes.[1,2] Due to the complexity of such membranes, most information on realistic biological membrane properties comes from experiment when available. However, with the continual increases in computer power, atomic-scale computer simulations of increasingly realistic membranes based on empirical force fields can provide valuable insight in the membrane structure and dynamics.[3−13]

At the present time, however, the all-atom molecular dynamics (MD) simulations of the membranes are limited to systems having several hundred lipid molecules on a time scale of about 100 ns. The extension of computer simulation of biological membrane systems to more biologically relevant time scales and system sizes requires a simpler treatment of the biological molecules of interest and the surrounding solvent molecules. The coarse-graining method is seen as one possible way to accomplish this task.[14−22] Numerous methods for coarse-graining and coarse-grained (CG) models have been reported in the past (for a recent review see, e.g., ref 23). In the framework of united atom-like CG models, which are next to all-atom models in length scale, small groups of atoms are represented as single interaction sites, thus preserving the molecular frame but with less resolution.

The most difficult stage in the implementation of such approaches is the construction of an effective force field to describe the interactions between the CG structural units. Most existing approaches for parametrization of effective CG potentials are not directly based on the underlying atomistic-scale forces which are available from, for example, all-atom empirical potential MD simulations. These approaches instead typically target the reproduction of the average structural properties seen in atomistic simulations, for example using an iterative adjustment of potential parameters starting from an approximation based on potentials of mean force[14,24] or the Inverse Monte Carlo technique.[25] They may also be parametrized to match thermodynamic properties.[14−16,21,22] These latter approaches can be computationally expensive

10.1021/ct050300c CCC: $33.50
Published on Web 02/18/2006

and may become rapidly less useful with an increase in the number of different types of interaction centers in the system.

The coarse-graining of potentials is further complicated due to the fact that the effective interaction between structural units is expected to be more dependent on the thermodynamic state compared to atomistic force fields. This is because the effective interaction between the CG structural units is defined by the average structure (e.g., average orientations, distances) within the complexes formed by those units in a particular phase. Rigorously speaking, the best CG potential should be a potential of mean force (PMF), which is a configurational *free energy* in a reduced phase space. Therefore, the structural properties, and thus the CG potentials, can in principle be sensitive to variations in temperature and other thermodynamic conditions due to both entropic and (anharmonic) energetic effects. A poor transferability of "coarser-grained" potentials may undermine the reliability of simulations in which the same CG potential is used to simulate different system conditions. Intuitively, the construction of a more transferable CG model may be improved by using a finer scale coarse-graining as has been observed in the modeling of proteins.[26,27] Of course, this comes at an increased computational cost. Another solution is to fit (or refit) CG potentials of a system under the same thermodynamic conditions for which the CG simulation is intended. This approach requires the availability of computationally efficient methods for the development of CG force fields as discussed below.

Recently, a significantly different approach for developing CG potentials from an underlying explicit atomistic molecular dynamics (MD) simulation has been presented.[28,29] This approach is called "multiscale coarse-graining" (MS-CG) because it provides a systematic way for coarse-graining the interactions from a reference all-atom MD simulation. This methodology is an extension of the force-matching (FM) approach developed in ref 30. The FM methodology enables one to obtain effective pairwise interactions, and a simpler variant of it was also successfully applied to construct all-atom models for several liquids from ab initio MD simulation data.[30,31]

The MS-CG method implements a FM procedure for the CG images from the reference atomistic trajectory/force data. It is natural to map atomic groups into a CG site through association of the CG site with the centers-of-mass (CM) because the force acting on the CM of the atomic groups can be straightforwardly evaluated from the atomistic MD data. The FM procedure, when applied to these data, should yield an approximation to the effective interactions between the CG sites. These interactions are a variational least-squares approximation to the average atomistic forces.[28,29]

The MS-CG method was first successfully applied to construct a CG model of the pure dimyristoylphosphatidylcholine (DMPC) lipid bilayer.[28] This model was able to accurately reproduce the structural properties of the system in constant NVT all-atom MD simulations. The FM method can also be extended to include a fit of the pressure in the all-atom simulation, with the resulting force field then being appropriate for MS-CG simulations under constant NPT conditions.[29]

In the present article, we report the application of the MS-CG method to a mixed DMPC/cholesterol lipid bilayer. Several MS-CG models are developed here that are different in the level of coarse-graining of the cholesterol sterol molecule and the underlying bonding scheme. The models are shown to capture many properties of the solvated bilayer and represent an important milestone in the coarse-graining of complex biological membrane systems. The applicability of the present MS-CG models is limited to preassembled bilayers. However, the MS-CG methodology permits an efficient parametrization of the interactions for simulation of other phases of phospholipid/cholesterol mixtures by fitting to an all-atom MD simulation of the phase of interest, which is generally expected to be better than a construction of a "universal" coarse-grained model. In general, the quality of an MS-CG model is determined by the length of the reference all-atom simulation as well as the accuracy of the underlying force field used in the reference MD simulation. Consequently, the present MS-CG force fields can be improved upon with the availability of more accurate atomistic mixed bilayer simulations.

This article is organized as follows: section 2 describes the MS-CG method used, and then section 3 reports applications of the method to the mixed lipid bilayer. This latter section starts with a presentation of general details of the FM fitting procedure from the underlying all-atom MD simulations and then presents the MS-CG models. The accuracy of the MS-CG models is also analyzed. In section 4, conclusions from this study are presented.

II. Coarse-Graining Methodology

The algorithmic development of the MS-CG method is presented elsewhere.[28,29] In the framework of the MS-CG approach, the atom−atom force $\mathbf{f}_i^{\mathrm{p}}(r_{ij})$ (acting from particle j on particle i) can be partitioned into a short-ranged part and a long-ranged Coulomb part as

$$\mathbf{f}_i^{\mathrm{p}}(r_{ij}) = -\left(f(r_{ij}) + \frac{q_i q_j}{r_{ij}^2}\right)\mathbf{n}_{ij} \tag{1}$$

where r_{ij} is the modulus of the vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ connecting two atoms, q_i is the partial atomic charge (subject to fit), and $\mathbf{n}_{ij} = \mathbf{r}_{ij}/r_{ij}$. The effective short-ranged force is fitted by cubic splines connecting a set of points $\{r_k\}$ ($r_{k_{max}}$ defines a cutoff radius), thus preserving continuity of the functions and their first two derivatives across the junction, such that

$$f(r, \{r_k\}, \{f_k\}, \{f_k''\}) = A(r, \{r_k\})\, f_i + B(r, \{r_k\})\, f_{i+1} + C(r, \{r_k\})\, f_i'' + D(r, \{r_k\})\, f_{i+1}''; \; r \in [r_i, r_{i+1}] \tag{2}$$

where A, B, C, and D are known functions of r, $\{r_k\}$, and $\{f_k\}$, and $\{f_k''\}$ are tabulations of $f(r)$ and its second derivative on a radial mesh $\{r_k\}$. A spline representation depends linearly on its parameters $\{f_k, f_k''\}$, which permits the use of force averaging over trajectories as a part of the least-squares fit.

In the first step of the MS-CG procedure, all sets of system configurations recorded along the trajectories of the reference MD simulation are partitioned into blocks, each containing L configurations. For each block, the reference total force

$\mathbf{F}^{\text{ref}}_{\alpha il}$, which acts on the ith interaction center of kind α in the lth configuration set of the block, and the same force predicted using the representation in eq 2 are required to be equal, which leads to the following equation

$$-\sum_{\gamma=nb,b}\sum_{\beta=1,K}\sum_{j=1,N_\beta}\left(f(r_{\alpha il,\beta jl},\{r_{\alpha\beta,\gamma,k}\},\{f_{\alpha\beta,\gamma,k}\},\{f''_{\alpha\beta,\gamma,k}\})+\frac{q_{\alpha\beta}}{r^2_{\alpha il,\beta jl}}\delta_{\gamma,nb}\right)\mathbf{n}_{\alpha il,\beta jl}=\mathbf{F}^{\text{ref}}_{\alpha il},\ i=1,...,N_\alpha,\ \alpha=1,...,K,\ l=1,L \quad (3)$$

with respect to $\{f_{\alpha\beta,\gamma,k}, f''_{\alpha\beta,\gamma,k}, q_{\alpha\beta}\}$, which are the force parameters subject to fit. In eqs 3, $\{\alpha il\}$ labels the ith site of kind α in the lth configuration of the block; $r_{\alpha il,\beta jl}$ is the distance between sites $\{\alpha i\}$ and $\{\beta j\}$ in the lth configuration; $q_{\alpha\beta}$ is a product of partial charges q_α, q_β of sites of kind α and β; and N_β and K are, respectively, the number of sites of kind β and the total number of site kinds in the system.

The MS-CG method also allows for a systematic separation of bonded and nonbonded forces.[29] This is important for the CG sites which have an overlap in regions of intra- and intermolecular motions, causing the FM force field in those regions to be a mixture of nonbonded and bonded components, which are otherwise impossible to separate. To aid in fitting explicitly the bonded forces, the additional index $\gamma=\{nb, b\}$, which indicates whether the site pair $\alpha il,\beta jl$ is bonded (i.e., $\gamma=b$), is introduced. This index distinguishes the f, f'' parameters for bonded and nonbonded interactions in eq 3. For bonded pairs, the Coulomb term is absent, which is enforced in eq 3 by the $\delta_{\gamma,nb}$ term which is unity if $\gamma=nb$ and zero otherwise. It is permissible to choose the spline mesh $\{r_{\alpha\beta,\gamma,k}\}$ to be different for different pairs $\{\alpha\beta\}$ of site kinds.

If L is large enough, eq 3 forms an overdetermined system of linear equations with respect to the fitting parameters. Standard equations which are linear with respect to $\{f_{\alpha\beta,\gamma,k}, f''_{\alpha\beta,\gamma,k}\}$ must also be included into eq 3 to ensure that the $f(r)$'s first derivative, $f'(r)$, is continuous across the boundary between two intervals.[29,32] The number of equations in eq 3 can also be increased with an increment equal to the number of interaction sites in each configuration set.

The method described above is general and can be straightforwardly applied to derive an effective force field for coarse-grained models. To do this, the reference atomistic trajectories and forces have to be reduced to the trajectories and forces for the structural units subject to coarse-graining. The most natural way to coarse-grain a set of atomic groups is to place the CG site at each group's center-of-mass (CM) since the force acting on it can be straightforwardly evaluated from the atomistic data. Consequently, the FM algorithm can be applied to determine coarse-grained interactions from fully atomistic MD trajectory data. The CG force field so obtained is then an approximation to the effective interaction, or pairwise PMF, between the CG interaction sites.

Generally, a CG site can be mapped into an atomic group in other ways with the reference force evaluated in the atomistic MD simulation. For example, a CG site may instead be associated with the geometrical center (GC) of the structure (i.e., the CM of the system assuming that all atoms have the same mass). The latter choice may be more preferable for atomic groups with highly uneven mass distribution (e.g., the water molecule). Of course, the MS-CG force field will be dependent on where the CG sites are placed, and the same choices of that placement may be better than others. The optimal choice of MS-CG sites for a given problem is the topic of ongoing research in our group.

As is shown elsewhere,[29] CG force fields can perform well in the reproduction of the bulk phase structural properties; however, they may fail to maintain the proper internal pressure in the system, and as a result the density may also be incorrect (typically too low) in constant pressure CG MD simulations. This behavior can be corrected in the MS-CG approach through the virial equations used to evaluate the pressure in the MD simulation.[29] Because the virial depends linearly on the atomic forces and the kinetic energy does not rely on the forces at all, the MS-CG force field can also be constrained to produce the correct pressure.[29] This task can be accomplished by adding to eq 3 the proper constraints which include the system virial.[29]

The FM force field derived by including the virial constraint depends explicitly on the instantaneous kinetic energy and therefore on the temperature in the reference atomistic simulation. This may define the transferability (or lack thereof) of the present MS-CG models to other temperatures. Thermodynamic properties which rely on the derivatives of the temperature (e.g., thermal expansion coefficient) may also be less accurate.

Other CG approaches have developed a CG force parametrization by fitting a few preselected macroscopic and average structural properties. Our method is fundamentally different as it systematically maps the underlying atomistic-level interactions into effective interactions between CG structural units. Hence, it is a "multiscale" coarse-graining approach. Importantly, for systems such as mixed phospholipid/cholesterol bilayers (described below) for which fitting CG potentials to preselected average properties is difficult at best, the MS-CG approach appears to offer some real advantages.

III. Results and Discussion

A. Molecular Dynamics Simulation Details. In the present work, the MS-CG method was employed to obtain a CG model of a mixed DMPC/Chol bilayer at 50 mol % concentration of the Chol. The effective force field describing interactions between sites in the CG representation of the bilayer was derived using trajectory and force data from a single all-atom MD simulation. The DMPC molecules were modeled using a united atom force field,[12] and Chol molecules were modeled with a modified AMBER force field.[8] For water, the rigid TIP3P model[33] was used. The temperature of the system was kept constant at $T=308$ K using the Nosé-Hoover thermostat with a relaxation time of 0.2 ps. The electrostatic interactions were calculated via the particle mesh Ewald (PME) summation, and the van der Waals interactions were cut off at 0.7 nm. The system of 6336 total atoms (32 lipids, 32 cholesterol molecules, and 1312 water molecules) was integrated with a time step of 2

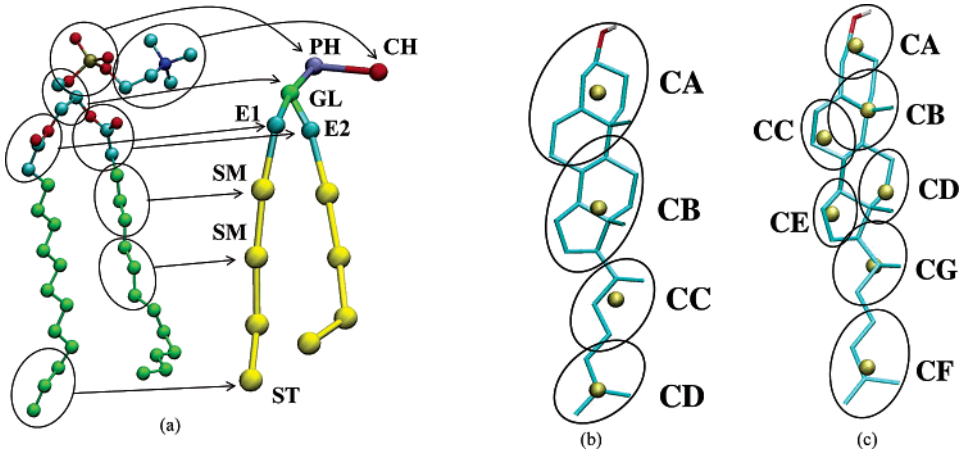

Figure 1. Coarse-grained representation of lipid and cholesterol molecules: panel (a) DMPC; panel (b) four-site cholesterol; and panel (c) seven-site cholesterol.

fs, and the initial structure of the DMPC/Chol bilayer taken from ref 34 was equilibrated for 4 ns. The reference properties of the atomistic system were obtained from the MD simulations under both constant NPT and NVT conditions. The equilibrium volume of the supercell was 92.029 nm^3 with an area per DMPC molecule of 0.440 nm^2 under the assumption of an effective area per Chol molecule of $\sim$0.336 nm^2.[35]

B. Coarse-Graining and Details of the Force-Matching. The MS-CG water molecules were mapped into a single CG interaction site associated with their GC instead of CM. As shown elsewhere,[28] with such a choice the FM force field produces a water structure which compares better to the atomistic reference MD system. The CG representation of a DMPC molecule is shown in Figure 1 (a) (also used in ref 28) and is similar to the models of Shelley et al.[14] and Marrink et al.[16] A coarse-graining of the acyl tail into a smaller set of spherical sites has the potential problem that could not fully describe trans−gauche isomerizations of the acyl chains. The trans−gauche conformations may result in a sharp change in the geometries of SM and ST CG groups in the all-atom MD simulation and thus may have an adverse impact on the quality of the MS-CG force field. On the other hand, the presence of cholesterol reduces the probability of gauche defects.[36]

For the Chol molecule, two CG representations were developed. In the first representation, the molecule was coarse-grained into four interaction sites. In the second representation, seven sites were used. The sites associated with each CG model of Chol are shown in Figure 1b,c. In the four-site representation, the tetracyclic ring structure is split between CA and CB sites. The CA site includes the polar hydroxyl group, while the CC−CD site pair represents the aliphatic side chain attached to the ring structure. For the model with four-site CG sterols, Table 1 summarizes the average atomistic radii of the groups forming the CG sites as evaluated from the same MD trajectory data used in the FM fit as well as the sum of the charges on the atomistic sites from which the CG site is formed. (It should be noted that in the final MS-CG models none of the CG groups have charges, i.e., they interact only through effective short-ranged forces.) In the four-site Chol representation, the arrangement

Table 1. Radii, $R_{CG} = \langle (r - r_{CM})_{max} \rangle$, and Total Charges, q_{CG}, of the underlying CG Sites in the All-Atom Model

site	R_{CG}, nm	q_{CG}, au	site	R_{CG}, nm	q_{CG}, au
CH	0.237	+1.4	ST	0.144	0
PH	0.165	−1.4	CA	0.392	−0.012
GL	0.138	+1.2	CB	0.266	0.009
E2	0.156	−0.7	CC	0.201	0.004
E1	0.153	−0.5	CD	0.160	−0.001
SM	0.140	0	WC	0.078	0

of the CG sites is close to linear. In the seven-site representation, three more sites were added to better reflect the planar geometry of the ring structure. The total number of DMPC plus Chol site pairs is 78 for the model with the four-site representation of the Chol molecule and 120 for the seven-site representation. Such a large variety of interaction sites makes it very difficult to utilize existing methods[14,16,25] for parametrization of the coarse-grained interactions. However, the MS-CG method is capable of treating such a complex system.

The intramolecular CG sites were linked by effective bonds, and two bonding schemes were used. In the first bonding scheme, which is referred to as "minimal", only consecutive sites were linked by bonds. In such a model, the force term describing bond bending was implicitly present in the FM nonbonded forces. In the second scheme, the next-nearest-neighbors were connected by bonds (e.g., CH-GL, PH-E1,2, etc., for DMPC and CA-CC, CB-CD, etc., for Chol). Such interactions are similar to a bond angle potential. The use of the bonds instead of angle potentials was necessary because the FM method can handle explicitly only site−site (either bonded or not) interactions. Therefore, the second bond scheme is referred to as the "bond angle" scheme. The use of the explicit separation of bonds in the FM procedure increased the number of interactions for the force-match. For example, in the model with the four-site Chol molecules and the "bond angle" scheme, the total number of interactions to match was 101.

For the purpose of force-matching, the atomistic MD simulation was carried out for 5 ns in a constant NVT ensemble using the equilibrium volume obtained from a constant NPT MD simulation. The trajectory, velocity, and

Table 2. Coefficients A_n of the Least-Squares Fit of TIP3P MS-CG One-Site Force Field Using the Expansion $f(r) = \sum_{n=2}^{16} A_n/r^n$ [a]

n	A_n	n	A_n
2	1121.994059416	10	1461833177752.
3	−100003.0777743	11	−2619871642943.
4	3956632.300647	12	1698105340916.
5	−91592923.27445	13	1534516677398.
6	1375454692.338	14	−940850812653.1
7	−14021182651.00	15	−2041415905133.
8	98418431541.42	16	−1742793289718.
9	−470296232562.7		

[a] Atomic units were used. A cut-off of 0.8 nm must be applied to this expansion.

force data were sampled with an interval of 1 ps so that the number of stored configurations was 5000. The instantaneous virial for each sampled configuration was also recorded. The velocity and virial data were necessary to implement the virial constraint.[29] Next, trajectories, velocity, and forces were transformed into corresponding data for the CG sites and then were used as an input into the FM algorithm with an explicit separation of bonds and virial constraint.[29] The FM force field was represented by a spline over a mesh with a grid spacing of approximately 0.01 nm for nonbonded and 0.0025 nm for bonded interactions. For nonbonded forces, the mesh was extended to 0.13 nm. The number of configurations used to build each of the overestimated systems of equations in eq 3 was $L = 3$.

As mentioned earlier, despite the fact that some composite CG sites in the underlying atomistic model are charged (see Table 1), it is possible to omit the Coulomb term in eq 3 if a large enough cutoff is used for the short-ranged interactions. For example, if a sufficiently large cutoff is used, the FM procedure with the Coulomb term explicitly present yielded effective charges on the CG sites to be much smaller compared to those inferred from the reference atomistic charges. This is a result of the inherent screening of Coulomb forces and the short-ranged forces then incorporating effectively the Coulomb interactions within their cutoff radius. The screening effects come from the water as well as the polar DMPC headgroups which are captured by the MS-CG force field. Therefore, the FM procedure was used with the Coulomb term removed in eq 3.

C. MS-CG Interactions. *1. Water Model.* The one-site CG representation of the TIP3P model as obtained from the FM to the all-atom MD simulation of the bilayer was tested on a bulk water system. The structural properties from the all-atom and MS-CG MD simulations of the 256 TIP3P molecules in the constant NVT ensemble and at ambient conditions were identical. However, the density of the CG water in the NPT MD simulation was slightly lower. By contrast, the FM (using the virial constraint) of the one-site water force field to the MD simulation of 256 TIP3P molecules at ambient conditions produced a force field which showed the same structural properties as one from FM to the bilayer system; however, the ambient density was correct. A polynomial representation of the MS-CG potential for this model is given in Table 2. A comparison of the water

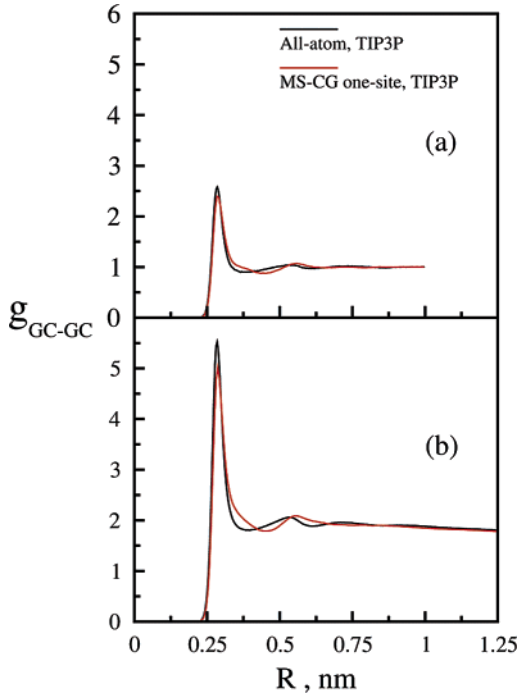

Figure 2. Comparison of the water GC-GC RDFs from the all-atom (black line) and MS-CG (red line) MD simulations of bulk water [panel (a)] and the bilayer system [panel (b)] using the MS-CG TIP3P model fitted to the bulk water data.

Table 3. Properties of TIP3P One-Site MS-CG CG Water Model[a]

property	TIP3P 1-site	TIP3P[b]	expt[c]
ρ (kg/m^3)	1004	1002	997.05
U^{pot} (kJ/mol)	−8.3	−41.1	−41.5
α_P (10^{-4} 1/K)	14.7	6.4	2.53
κ_T (10^{-5} 1/bar)	37.1	9.2	4.52
D_{diff} (10^{-9} m^2/s)	18.1	5.19	2.3

[a] Shown are ρ, density; U^{pot}, average configuration energy; α_P, thermal expansion coefficient; κ_T, isothermal compressibility; D_{diff}, self-diffusion coefficient. [b] References 33 and 44. [c] Reference 45 (25 °C).

geometrical center-geometrical center (GC-GC) RDFs from the bulk water and bilayer MD simulations is given in Figure 2.

Some other thermodynamic properties of the MS-CG water model are compiled in Table 3. The larger isothermal compressibility, κ_T, from the CG model is likely a consequence of the use of the virial constraint in the FM procedure which explicitly depends on the temperature in the reference MD simulation. A similar behavior has been reported for a one-site water model matched to ab initio MD simulation.[29]

The water−water (W−W) force field from the FM to the bulk water system with the virial constraint was used in all models of the bilayer reported here. However, in contrast to the bulk geometry, the all-atom water in the vicinity of the bilayer surface is "orientationally polarized", orienting its molecular dipoles toward the bilayer surface.[7,37] On the other hand, the correct pressure of the water seems to be a more influential factor in the formation of the bilayer properties even in MD simulations in the constant NVT ensemble. For example, the bilayer structural properties from the constant NVT MD simulation using the bulk MS-CG water force field

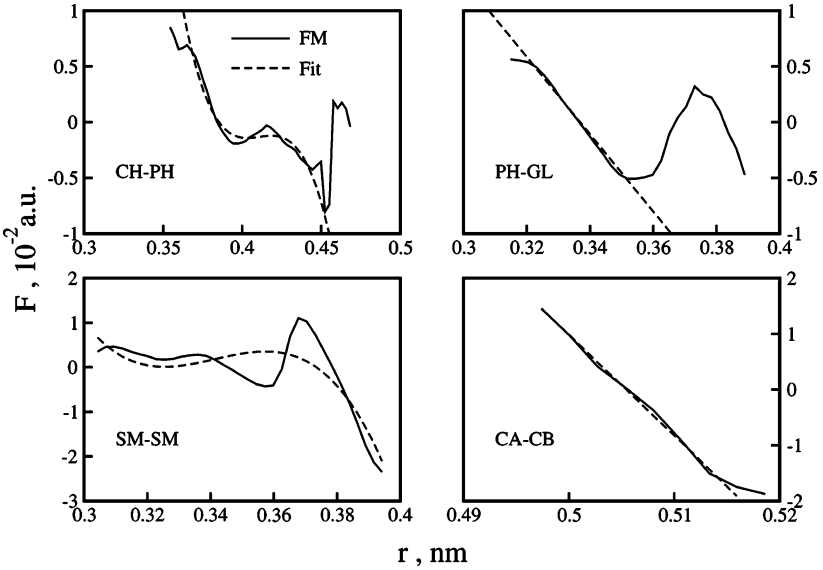

Figure 3. Effective pairwise bonded forces between selected coarse-grained interaction sites as a function of intersite separation (solid curve) and their analytical fit (dashed curve) used in the MS-CG MD simulations.

appeared to be slightly better compared to those properties if the MS-CG water force field from explicit FM to the bilayer system was used. It seems this behavior was due to a more correct pressure of the water on the bilayer surface from the bulk MS-CG water model.

2. DMPC and Chol MS-CG Models. Selected force profiles for bonded interactions for the models with the four-site representation of a Chol molecule are shown in Figure 3. The bonded forces of the model with the minimal bonding scheme were identical to the corresponding forces in the model with the bond angle scheme. As seen in a majority of the cases, the force deviates substantially from being harmonic (i.e., linear). In the MS-CG MD simulation, a least-squares approximation of the bonded forces either by harmonic or harmonic+quartic forces was used. A harmonic approximation was used, for example, to a lipid PH-GL bonded interaction, as it was found to be accurate enough. In those FM bonded forces for which the harmonic approximation obviously was not accurate, the quartic potential was instead used. A comparison of FM forces with their analytical fits is shown in Figure 3. An accurate description of the bonded forces is important for those parts of the molecules that have greater flexibility, such as the hydrocarbon chains in the DMPC molecule. To give an indication of the rigidity of the various bonds in the CG representations of the DMPC and Chol molecules, Table 4 shows parameters for the harmonic approximations to the bonded forces and associated frequencies (estimated as if the bonded pair were an isolated diatomic molecule). The low-lying spectra of bond frequencies also allows one to increase the time step in the MS-CG MD simulation compared to an all-atom simulation. The model with the minimal bonding scheme was deemed to be more accurate as the interactions between nearest-neighbor sites are included into nonbonded forces and, therefore, may be less influenced by errors introduced from analytical fits of the bonded forces.

Selected profiles of the MS-CG nonbonded forces for the model with the four-site representation of Chol are shown in Figure 4. The effective force field often shows substantial variations with distance in the vicinity of separations of the order of the sum of average radii of the CG sites (see Table 1). This is because at these distances the interaction in the all-atom MD simulation is most sensitive toward mutual orientations of the atomic groups representing the CG sites. For some of the site pairs, which were at large separations in the all-atom MD simulation, the unsampled core region was also significant. For example, for the CH−CD pair in the model with the four-site Chol, the r_{core} was ∼1.7 nm. Those DMPC-W and Chol-W interactions which have a large r_{core} value or are underconvergent/noisy might be improved by matching to the simulation of single DMPC or Chol molecules in the solvent. However, in the present study this approach was not used. The interaction within the r_{core} region

Table 4. Parameters of Harmonic Approximation to Bonded Forces and Corresponding Frequencies for the Model with the CG Chol Molecule of CG Four Sites[a]

bond	k_b, [kJ/(mol·Å^2)]	r_0, nm	ω, cm^{-1}
CH−PH (b)	0.184	0.385	132.5
CH-GL (a)	0.007	0.610	31.9
PH-GL (b)	0.187	0.337	168.5
PH-E2 (a)	0.022	0.435	51.1
PH-E1 (a)	0.038	0.661	67.4
GL-E2 (b)	0.554	0.282	315.9
GL-E1 (b)	0.237	0.326	206.4
GL-SM (a)	0.002	0.595	19.4
E2-SM (b)	0.365	0.374	254.7
E2-SM (a)	0.005	0.692	28.1
SM-SM (b)	0.203	0.345	204.3
SM-SM (a)	0.020	0.654	64.0
CA-CB (b)	0.871	0.506	235.0
CA-CC (a)	0.096	0.924	102.2
CB−CC (b)	0.648	0.426	266.1
CB−CD (a)	0.130	0.831	118.7
CC−CD (b)	0.278	0.424	206.5

[a] Labels (b) and (a) label "consecutive" and "angle" bonds, respectively.

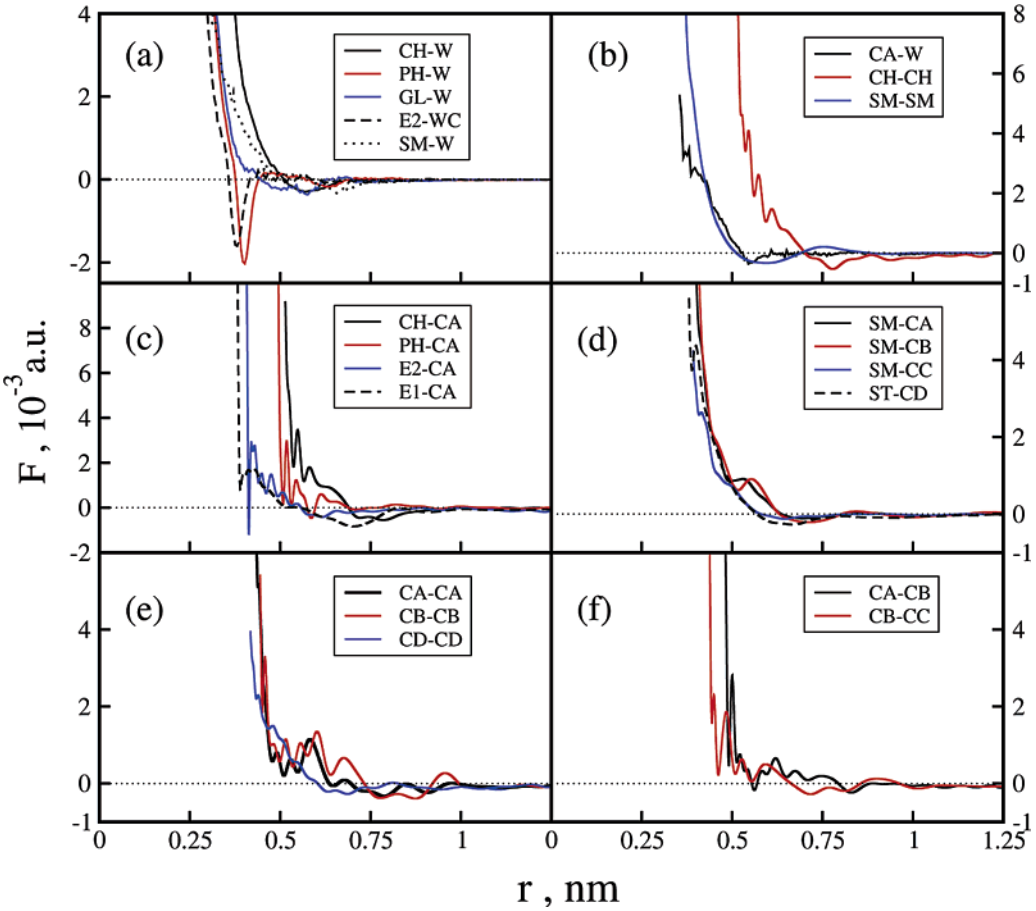

Figure 4. Effective pairwise nonbonded forces between selected coarse-grained interaction sites smoothed by a polynomial fit as a function of intersite separation. The approximate error bars in these curves are about 5−10% of the force value.

was therefore assumed to be weakly repulsive. Obviously, large r_{core} for certain pairs of CG sites limits the transferability of the model to other than bilayer geometries. Consequently, the models reported here can be only safely used to simulate preassembled bilayers.

The force profiles for some of the bilayer site−site pairs sometimes converged slowly in the FM procedure with some statistical noise. Such noise is not generally a source of difficulty in the MS-CG MD simulation because the force profiles can be further smoothed out using a least-squares polynomial fit of the FM spline data. However, such a fit in the present case gave somewhat worse properties. Therefore, the original spline fit was used to conduct the reported MS-CG MD simulations.

An inspection of the MS-CG effective forces also provides important insight into the nature of the interactions inside the bilayer. As seen from Figure 4(c), the most attractive interaction between lipid and Chol species occurs between E1 and CA sites. This is an interesting fact because there is experimental evidence of strong hydrogen bonding of the Chol hydroxyl group (belonging to the CA site), which acts as a proton donor with the carbonyl group of the sn-2 chain[38] (of which the E2 site is a part). MD simulations of DMPC/Chol bilayer at low sterol content have also shown that bilayer Chol can bind to lipid molecules via either hydrogen bonding or water bridges.[11,35] At a low Chol concentration (X_{Chol} = 12.5 mol %), the largest number of hydrogen bonds (both direct and through water bridges) occurs for carbonyl oxygen in the sn-2 chain;[35] however, as the Chol concentration increases (to X_{Chol} = 22 mol %), the sn-1 oxygen (which is in the E1 site) becomes a preferable place for the OH group of Chol to bind.[11] A similar behavior was suggested by NMR studies.[39] With an increase of the Chol concentration due to the "condensing effect", flexible acyl chains pack closely around the Chol molecules, limiting the ability of the latter to form H bonds with the sn-2 oxygen. The tendency of Chol molecules to orient their hydroxyl groups close to the carbonyl groups of the phospholipid and then to remain there was also observed in an MD simulation of DPPC/Chol bilayer.[40]

There is also a pronounced attractive part in the CH−CA force profile, while the PH−CA interaction is mostly repulsive. The CH is the second lipid CG site with which the CA site interacts most significantly. The attractive part of the CH−CA force is centered around 0.77 nm and is relatively broad. This correlates with MD simulation studies which pointed toward complexation of the choline moiety with the Chol hydroxyl group (the CA site) via formation of charged pairs.[11] The OH of Chol can bind to the oxygens of the phosphate group (PH site); however, the probability of formation of the hydrogen bonds at this location is much smaller compared to that of carbonyl groups (E1,2 sites).[11,35] The small negative dip in the PH−CA force profile might be a manifestation of the hydrogen bonding of the phosphate

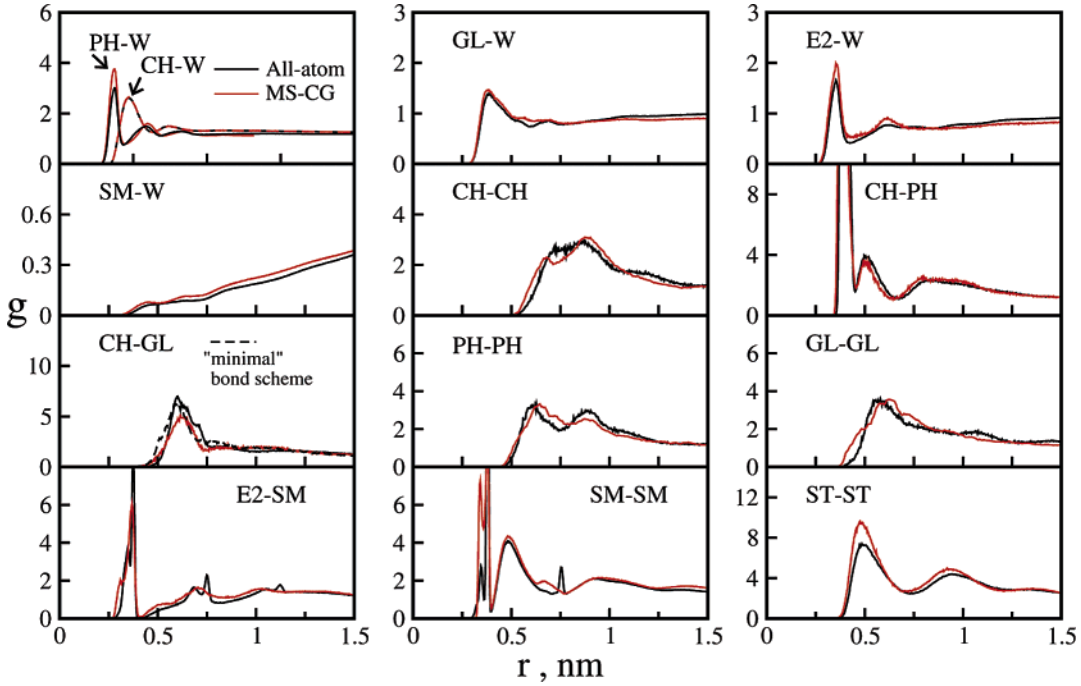

Figure 5. Selected DMPC−water and DMPC−DMPC site−site RDFs from the all-atom MD simulation (black lines) compared to those from the MS-CG MD simulation (red lines) using the four-site representation of the Chol molecule and the "bond angle" bonding scheme.

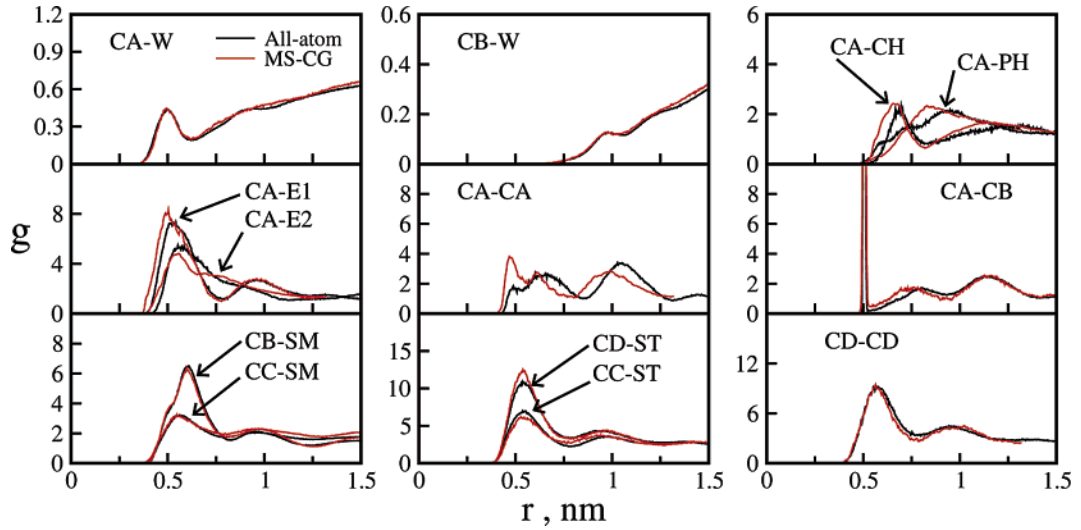

Figure 6. Selected Chol−water, Chol−DMPC, and Chol−Chol site−site RDFs from the all-atom MD simulation (black lines) compared to those from the MS-CG MD simulation (red lines) using the four-site representation of the Chol molecule and the "bond angle" bonding scheme.

oxygens and OH of Chol. The interactions of the Chol sites with acyl chain sites (SM, ST sites), which arise mostly from van der Waals forces, are only weakly attractive.

D. MS-CG Simulations of the Bilayer. To help validate the models, the system in the MS-CG MD simulation of the mixed bilayer was initially taken to be the same size as that of the all-atom MD simulation. At first, a 5 ns MD simulation for each MS-CG model was carried out using a constant NVT ensemble with time step of 0.005 ps. The MS-CG MD simulation for this system size was about 15 times faster. However, the gain in computational efficiency increases rapidly with an increase in system size, the latter (larger system sizes) being one purpose of CG simulations.

A comparison of the MS-CG RDFs from the model with the four-site CG Chol molecules with the bond angle scheme with their atomistic MD counterparts is given in Figures 5 and 6. The agreement is generally quite good. Some discrepancies in the bilayer site−site RDFs are attributed primarily to two factors. One of them is the analytical representation of the bonded forces, which as demonstrated in Figure 3 may deviate from the exact FM forces, thus influencing the subtle balance of the interactions in the MS-CG MD simulation. The second factor is the presence of sometimes large, unsampled regions of intersite separations. The most affected RDFs are those which involve solvated sites, and ways to improve this sampling are presently being

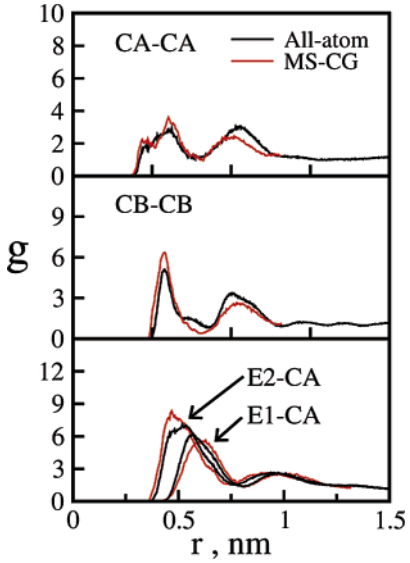

Figure 7. Selected site−site RDFs from the all-atom MD simulation (black lines) compared to those from the MS-CG MD simulation (red lines) using the seven-site representation of the Chol molecule and the "bond angle" bonding scheme.

explored. The use of the minimal bond scheme did not result in significant improvements or changes in the structure. For the pairs which were bonded by "angle" bonds in the bond angle scheme, the structure was slightly improved (e.g., the CH-GL pair, in Figure 5, panel CH-GL, dashed line).

An analysis of the CG structure of the bilayer is a convenient way to scrutinize the ordering in the bilayer because of the smaller number of degrees of freedom. The presence of a well pronounced first peak in DMPC-W and Chol-W RDFs (Figures 5 and 6) reflects the effective polarity of the respective CG sites. DMPC headgroup sites CH, PH, GL and ester groups E1,2 are solvated, and, as can be judged from a height of the first peak in the RDFs, the most strongly solvated site is PH followed by the CH, E2, GL, and E1 (RDF not shown) sites, respectively. The strong first peak in the PH-W RDF is a reflection of the fact that two phosphate oxygens are the most probable sites for the water to form hydrogen bonds with the DMPC.[35]

The good agreement of the CG and all-atom DMPC-W and Chol-W RDFs suggests that despite being nonpolar, the one-site CG water model adequately incorporates its interactions with the charged bilayer groups which arise from a polarity of the water molecules, e.g., hydrogen bonding with the PH and E1 sites. Nonpolarity of the SM and ST sites was manifested in an absence of the first peak in the SM-W (see Figure 5) and ST-W RDFs (the ST-W RDF is similar in shape to the SM-W RDF). As seen from Figure 6, the CA site is hydrophilic; however, it was solvated less compared to any of the lipid headgroup sites. The rest of the Chol CG sites are nonpolar.

As can be seen from Figure 7, the use of the seven-site representation of the Chol molecule improved the structure, presumably due to a better CG representation of the planar geometry of the Chol molecule. However, these improvements generally were not sufficient to consider the seven-site model to be universally superior compared to the four-site model, as the latter is computationally less expensive. The MS-CG models having a rodlike CG Chol molecular geometry of four interaction sites were already able to reproduce the bilayer structure at a sufficiently accurate level. Below is an analysis of the bilayer properties using the MS-CG model with the four-site Chol molecules with the bond angle bonding.

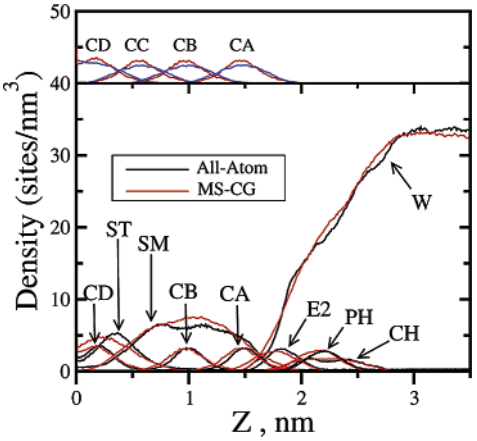

Figure 8. Comparison of the all-atom (black lines) and MS-CG (red lines) density profiles for several CG sites perpendicular to the bilayer. The insert compares z-density of Chol from smaller (red lines) and larger (blue lines) systems.

MS-CG MD simulations were next conducted using the constant NPT ensemble on the same size system. The simulation was 20 ns long. In Figure 8, the densities of the selected bilayer CG sites and water are shown, and the MS-CG model is again seen to capture the key structural properties of the bilayer accurately. The membrane thickness as estimated from the average distance between PH sites was 0.435 nm from the all-atom MD simulation compared to 0.422 nm from the MS-CG MD simulation. The area per DMPC molecule was 0.444 nm^2 (it was 0.440 nm^2 in the all-atom MD simulation). The z-density of Chol sites stretches to almost the same distance from the bilayer center as that of the acyl (SM, ST) sites. The front slope of the CA density follows closely the density profile of the SM sites.

Mutual arrangements of the CH and PH z-densities in Figure 8 and RDFs in Figure 6 indicate a tilt of the PH−CH bond toward the bilayer surface. The CH−CA and PH−CA RDFs suggest that the PH−CH bond points toward a neighboring CA group, forming the CH−CA charged pair. The first coordination number is 0.85 from the all-atom and 0.80 from the CG CH−CA RDFs. The CH−CA pair was most likely a solvent-separated pair as the position of the first peak in the CH−CA RDF at 0.69 nm exceeds the sum of 0.62 nm of the smallest CH−WC and CA-WC distances in the MS-CG MD simulation, as estimated from the corresponding RDFs. However, the formation of contact pairs is possible because the CH−CA separation reached a minimum at 0.51 nm. The Chol molecules tend to minimize exposure of their nonpolar sites to water molecules.[41] The CH sites shield the nonpolar parts of Chol (CA site represents a part of the first Chol ring) from the water that produced an additional driving force for the formation of the CH−CA pairs.

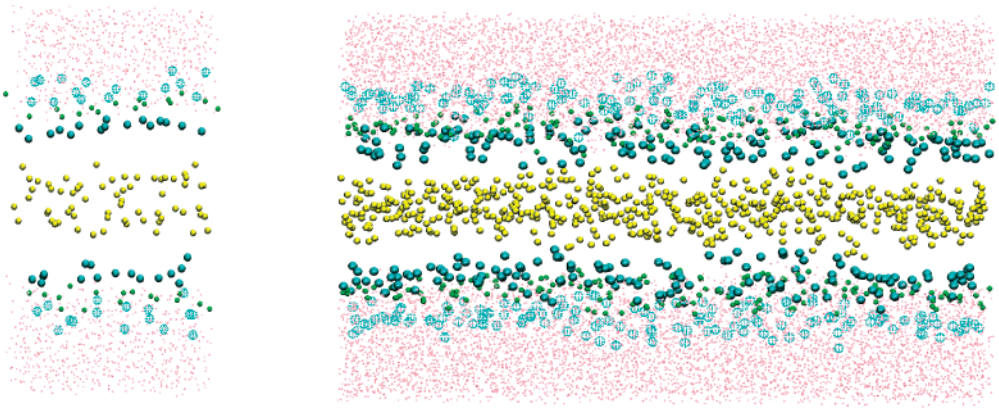

Figure 9. Side views of the bilayer with selected CG sites shown from the all-atom MD simulation [panel (a)] and the larger MS-CG MD simulation as described in the text (a segment of the supercell is shown) [panel (b)]. Referring to the CG sites in Figure 1, W: pink dots; CH: dotted spheres; GL: small green spheres; CA: large cyan spheres; and ST: small yellow spheres.

The MS-CG model was tested also on a system which was larger than the reference atomistic system by 4 times in the x and y directions (i.e., it contained 512 lipids, 512 cholesterol molecules, and 20 992 water molecules or 29 696 total sites). The simulation was initiated from the configuration obtained by a replication of the smaller system in the x and y directions four times and was integrated for a duration of 20 ns. Figure 9 presents snapshot images of the supercell with the selected sites shown from the all-atom MD simulation [panel (a)] and a segment of the simulation cell from the MS-CG MD simulation of the larger system. The z-density profiles from that simulation are similar to those of the smaller system; however, they exhibit a slight broadening which is most noticeable for Chol sites. Such a broadening might be partially attributed to undulations which are less pronounced in the smaller system due to finite size effects. The insert in Figure 8 compares the Chol z-density profiles systems of different sizes. The larger MS-CG mixed bilayer system was 50−100 times faster to integrate than the equivalent all-atom MD system of the same size depending on the choice of short-ranged cutoff and time step in the MD program for the MS-CG simulation.

The ordering in the bilayer was demonstrated by presenting the distributions of cosines of the angle between the PH−CH vector and the bilayer normal. In the all-atom MD simulation, the distribution of the PH−CH orientations has two well-defined maxima corresponding to orientations parallel and normal to the membrane. In pure lipid bilayer, the orientational distribution of the P−N vector is more flat.[8] A first maximum in the $\langle P_{\mathrm{PH-CH}}(\cos\theta)\rangle$ is attributed to the lipid molecules in close contact with Chol and corresponds to a charged CH−CA pair. The maximum at $\cos(\theta) = 1$ is due to the lipid molecules which are not associated (or associated less) with Chol molecules. These molecules prefer to orient their PH−CH bond outward. Several mechanisms have been suggested for such behavior,[8] depending on the concentration and packing of Chol molecules in the bilayer. For a Chol concentration of 50 mol %, probable mechanisms include a decrease in the area per lipid headgroup and the change in the solvation structure around the lipid headgroups as the average distance between lipid molecules in the Chol-contained bilayer increases. The distribution of PH−CH orientations from the MS-CG model is more flat with less structure from the parallel and perpendicular orientations to the bilayer. The distributions of the angles of acyl chain bonds with the bilayer normal are shown in Figure 10 (a).The MS-CG model shows broader distributions, suggesting more disorder.

In Figure 10(b), bilayer order parameters for bond orientations from the all-atom and MS-CG MD simulations are compared. The second-order parameter $P_2 = {}^1/_2\,(3\langle\cos^2(\theta)\rangle - 1)$ where θ is an angle between the bond and the bilayer normal, is shown. The value of $P_2 = 1$ indicates a perfect aligning of the bond pointing outward with the bilayer normal; $P_2 = -0.5$ means the bond pointing inward, and $P_2 = 0$ means a complete disorder in bond orientations. The PH-GL points predominantly into the bilayer interior. The linkage SM-E2 is in a more disordered state compared to the SM-E1 bond, likely as a result of a intensive sterol association with the E1 site. The SM-SM (II) bond, which is central in acyl chains, is in the most ordered state. The P_2 distribution is similar to those obtained by Marrink et al.[16] using their CG model of pure DPPC bilayer. The CB−CA vectors are well aligned along the bilayer normal, with an average angle $\theta = 9.8°$ from all-atom and $\theta = 13.0°$ from the MS-CG MD simulations.

The lateral diffusion coefficient, D, of the DMPC molecules was calculated from the slope of the diffusive part of the mean square displacement (MSD) of the molecular center-of-mass. The MSD from the MS-CG simulation is shown in Figure 11. The value $D = 1.1 \times 10^{-9}$ m^2/s was obtained for the DMPC diffusion, which is about four times larger than the experimental value,[42,43] $D = 3.0 \times 10^{-10}$ m^2/s for the ordered phase of DMPC/Chol at $X_{\mathrm{Chol}} = 33$ mol %. The diffusion of the CG water was $D = 11.2 \times 10^{-9}$ m^2/s, compared to the diffusion of the water in the all-atom MD simulation of $D = 2.74 \times 10^{-9}$ m^2/s. The dynamics of the MS-CG system is seen to be significantly (by about four times) faster than all-atom MD simulation, as is expected from the reduced number of degrees of freedom.[29] An advantage of this is a faster sampling of the configurational

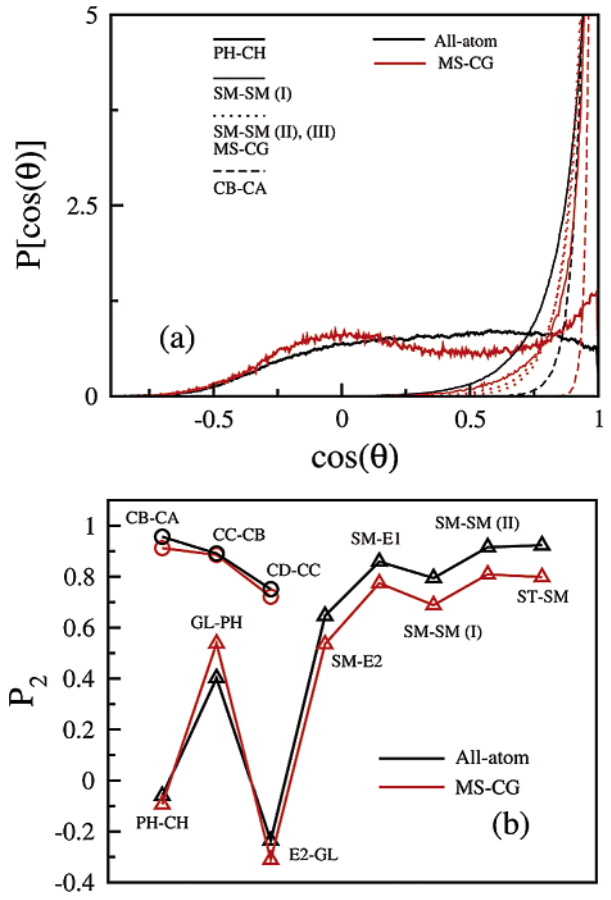

Figure 10. Panel (a): probability distribution of the angle between PH-CH, ST-SM, and CB-CA vectors and the bilayer normal from the reference all-atom (black lines) and MS-CG (red lines) MD simulations. Panel (b): P_2 order parameter of bonds with respect to bilayer normal from the all-atom (black line) and MS-CG (red lines) MD simulations.

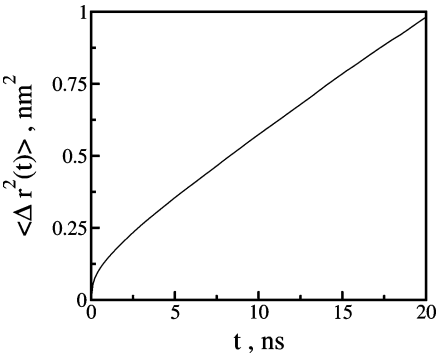

Figure 11. Center-of-mass mean squared displacement of DMPC molecules in the mixed bilayer from MS-CG MD simulation using the four-site CG Chol molecule.

space. However, the interpretation of the time scale in the MS-CG simulation is rather arbitrary.[29]

IV. Conclusions

Following our recent work on the application of the new MS-CG method to develop CG models of pure lipid bilayers,[28] we have applied the MS-CG method to construct a family of effective coarse-grained models for the mixed DMPC/Chol bilayer. The force-matching included a fit of the virial data, giving models that can be used for MS-CG MD simulations under constant NPT conditions. The models also explored different levels of coarse-graining of the cholesterol molecules and the intramolecular CG bonding scheme. The resulting MS-CG models were seen to predict the correct mixed bilayer structure, further demonstrating the effectiveness of the MS-CG method. Applications to more complex biomolecular systems are currently underway.

Acknowledgment. The research was supported by The National Institutes of Health (ROI GM63791). The computations in this project were provided in part by the United States National Science Foundation (NSF) cooperative agreement No. ACI-9619020 for the computing resources provided by the National Partnership for an Advanced Computational Infrastructure at the San Diego Supercomputer Center.

References

(1) Gennis, R. B. *Biomembranes: Molecular Structure and Function*; Springer: New York, 1989.

(2) *Lipid Bilayers: Structure and Interactions*; Katsaras, J., Gutberlet, T., Eds.; Springer: Berlin, 2001.

(3) Feller, S. E.; Pastor, R. W. *Biophys. J.* **1996**, *71*, 1350.

(4) Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C. *Biophys. J.* **1997**, *1331*, 235.

(5) Feller, S. E.; Yin, D.; Pastor, R. W.; MacKerell, J. A. D. *Biophys. J.* **1997**, *73*, 2269.

(6) Feller, S. E.; Venable, R. M.; Pastor, R. W. *Langmuir* **1997**, *13*, 6555.

(7) Tu, K.; Klein, M. L.; Tobias, D. J. *Biophys. J.* **1998**, *75*, 2147.

(8) Smondyrev, A. M.; Berkowitz, M. L. *Biophys. J.* **1999**, *77*, 2075.

(9) Smondyrev, A. M.; Berkowitz, M. L. *J. Chem. Phys.* **1999**, *111*, 9864.

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

(11) Pasenkiewicz-Gierula, M.; Róg, T.; Kitamura, K.; Kusumi, A. *Biophys. J.* **2000**, *78*, 1376.

(12) Smondyrev, A. M.; Berkowitz, M. L. *J. Comput. Chem.* **1999**, *20*, 531.

(13) Feller, S. *Curr. Opin. Colloid Interface Sci.* **2002**, *5*, 217.

(14) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464.

(15) Marrink, S.; Mark, A. *J. Am. Chem. Soc.* **2003**, *125*, 15233.

(16) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.

(17) Faller, R.; Marrink, S. J. *Langmuir* **2004**, *20*, 7686.

(18) Marrink, S. J.; Mark, A. E. *Biophys. J.* **2004**, *87*, 3894.

(19) Marrink, S. J.; Mark, A. E. *Biophys. J.* **2005**, *88*, 384A.

(20) Marrink, S. J.; Risselada, J.; Mark, A. E. *Chem. Phys. Lipids* **2005**, *135*, 223.

(21) Brannigan, G.; Brown, F. L. H. *J. Chem. Phys.* **2004**, *120* (2), 1059.

(22) Brannigan, G.; Philips, P. F.; Brown, F. L. H. *Phys. Rev. E* **2005**, *72* (1), 11915.

(23) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144.

(24) Meyer, H.; Biermann, O.; Faller, R.; Reith, D.; Müller-Plathe, F. *J. Chem. Phys.* **2000**, *113*, 6264.

(25) Murtola, T.; Falck, E.; Patra, M.; Karttunen, M.; Vattulainena, I. *J. Chem. Phys.* **2004**, *121*, 9156.

(26) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci.* **2003**, *100*, 10712.

(27) Zacharias, M. *Protein Sci.* **2003**, *12*, 1271.

(28) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.

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

(30) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.

(31) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 6573.

(32) De Boor, C. *Practical Guide to Splines*; Springer-Verlag: New York, 1978.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(34) Ayton, G.; Smondyrev, A. M.; Bardenhagen, S. G.; McMurtry, P.; Voth, G. A. *Biophys. J.* **2002**, *83*, 1026.

(35) Smondyrev, A. M.; Berkowitz, M. L. *Biophys. J.* **2001**, *80*, 1649.

(36) Scott, H. L. *Biophys. J.* **1991**, *59*, 445.

(37) Tobias, D. J.; Tu, K.; Klein, M. L. *Curr. Opin. Colloid Interface Sci.* **1997**, *2*, 15.

(38) Wong, P.; Capes, S.; Mantsch, H. H. *Biochim. Biophys. Acta* **1989**, *980*, 37.

(39) Guo, W.; Kurze, V.; Huber, T.; Afdhal, N. H.; Beyer, K.; Hamilton, J. A. *Biophys. J.* **2002**, *83*, 1465.

(40) Hofsäss, C.; Lindahl, E.; Edholm, O. *Biophys. J.* **2003**, *84*, 2192.

(41) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.

(42) Filippov, A.; Orädd, G.; Lindblom, G. *Biophys. J.* **2003**, *84*, 3079.

(43) Almeida, P. F. F.; Vaz, W. L. C.; Thompson, T. E. *Biophys. J.* **2005**, *88*, 4434.

(44) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

(45) *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, 20899; Linstrom, P. J., Mallard, W. G., Eds.; NIST: Gaithersburg, MD, 2003.

CT050300C