Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints

Teemu Murtola
Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FI-02015 Espoo, Finland
and Helsinki Institute of Physics, Helsinki University of Technology, P.O. Box 1100, FI-02015 Espoo, Finland

Emma Falck
Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign,
Urbana, Illinois 61801

Mikko Karttunen
Department of Applied Mathematics, The University of Western Ontario, 1151 Richmond Street North,
London, Ontario, Canada N6A 5B7, Canada

Ilpo Vattulainen
Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, FI-02015 Espoo, Finland;
Helsinki Institute of Physics, Helsinki University of Technology, P.O. Box 1100, FI-02015 Espoo, Finland;
Institute of Physics, Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland;
Memphys-Center for Biomembrane Physics, Physics Department, University of Southern Denmark,
Campusvej 55, DK-5230 Odense M, Denmark

The authors introduce a coarse-grained (CG) model for a lipid membrane comprised of phospholipids and cholesterol at different molar concentrations, which allows them to study systems that are approximately 100 nm in linear size. The systems are studied in the fluid phase above the main transition temperature. The effective interactions for the CG model are extracted from atomic-scale molecular dynamics simulations using the inverse Monte Carlo (IMC) technique, an approach similar to the one the authors used earlier to construct another CG bilayer model [T. Murtola *et al.*, J. Chem. Phys. **121**, 9156 (2004)]. Here, the authors improve their original CG model by employing a more accurate description of the molecular structure for the phospholipid molecules. Further, they include a thermodynamic constraint in the IMC procedure to yield area compressibilities in line with experimental data. The more realistic description of the molecular structure of phospholipids and a more accurate representation of the interaction between cholesterols and phospholipid tails are shown to improve the behavior of the model significantly. In particular, the new model predicts the formation of denser transient regions in a pure phospholipid system, a finding that the authors have verified through large scale atomistic simulations. They also find that the model predicts the formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the original model and the experimental phase diagram. However, the domains observed here are much more distinct compared to the previous model. Finally, the authors also explore the limitations of the model, discussing its advantages and disadvantages. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2646614]

I. INTRODUCTION

Membranes have a central role in many cellular processes such as signaling, recognition, and transport.[1–3] The overall properties of membranes are determined by phenomena at a range of time and length scales, starting from the atomistic and molecular regimes where intra- and intermolecular interactions play a role, up to mesoscopic and macroscopic regimes, where collective motions such as undulations, phase separation, and membrane fusion are important.

The present understanding of membranes is largely based on experimental studies.[1–3] At the same time, experiments have been complemented by theoretical and computational approaches.[1,4–6] Thanks to the interplay between the two approaches, a more detailed understanding of membranes and their biological relevance is emerging.

The standard method for computational studies of membranes in the atomistic regime is classical molecular dynamics (MD).[4–7] It provides detailed information on the structure and dynamics of individual molecules, as well as on intermolecular interactions. The main limitation is the computational load: with currently available computing resources, typical lipid systems that are being studied contain of the order of 10^3 lipid molecules, have linear sizes of a few tens of nanometers, and cover a time of the order of 100 ns or less.[8–10]

However, many interesting phenomena in lipid membranes occur at much longer time and length scales. Such phenomena include domain formation, membrane fusion, and phase changes. A particularly interesting issue is the formation of lateral domains in multicomponent bilayers, of which there is extensive experimental evidence.[1,11] Domain

126, 075101-1

formation is also central in the topical issue of lipid rafts,[12,13] although membrane proteins can also play a significant role in this case. The rafts have been suggested to participate in a wide range of cellular processes such as trafficking and sorting of proteins. Recent atomistic simulations of lipid rafts[8,9] and other lipid bilayer systems including cholesterol and other sterols[10,14–16] have provided insight into the atomic- and molecular-scale interaction mechanisms of the raft-associated lipids. However, to reach the length and time scales necessary to study large-scale phenomena, one has to resort to coarse-grained models that employ effective interactions and simplify the description of the underlying system.[6,17–19]

Coarse graining of any system consists of two stages. First, one has to define the coarse-grained degrees of freedom for representing the system. Typically, this is done using intuitive or heuristic arguments, with the aim to find the molecular degrees of freedom that are most relevant for the processes at hand. Recently, Murtola *et al.*[20] have proposed that this could be done somewhat more systematically by analyzing atomistic simulation data through self-organizing maps to find the most relevant conformations. Second, having defined the coarse-grained representation, one has to determine the effective interactions used in the coarse-grained model. The second issue has been studied rather extensively during the last decade (see below).

A significant amount of interest has been devoted to the development and applications of semi-atomistic membrane models.[19,21–25] The basic idea of these models is to describe small groups of atoms as single interaction sites. Several models have been proposed, differing mostly in the way the interactions are constructed (see, e.g., Ref. 19 for a review). Recently, a versatile model constructed by Marrink *et al.*[22] has been successfully applied to various studies of lipid systems.[26–30] This model is parametrized using thermodynamic quantities such as partition coefficients between water and oil. The models constructed by Voth and co-workers[23–25] represent a somewhat different approach and are discussed separately below.

An alternative approach is to use simpler theoretical models that include only the most relevant degrees of freedom.[31–39] Zuckermann and co-workers[31–35] have developed an off-lattice model where the molecules are described by hard-core particles with spin-type internal degrees of freedom. With a particular choice of interaction parameters, they have designed models whose phase diagrams are in agreement with the experimental ones for phospholipid bilayers containing cholesterol or lanosterol.[31–34] In addition, lateral diffusion in the bilayers has been successfully described.[35] Other approaches include field-theoretical approaches that have been applied to study phospholipid/cholesterol bilayers[36] and phase separation in mixed phospholipid bilayers.[37] Further, theoretical models for ternary mixtures of phospholipids and cholesterol have been shown to produce qualitatively correct phase diagrams.[38] Our previous model,[39] which is the starting point for the present work, is discussed separately below.

With most of the above approaches, the main difficulty is the proper construction of the interactions. A conceptually attractive alternative would be to construct the interactions by matching selected properties of the coarse-grained model to more detailed simulations of the same system. Several approaches have been proposed based on this idea, such as force matching[40–42] and structure matching.[39,43–46]

The force matching method of Voth and co-workers[41,42] uses the configurations and forces from a detailed simulation to perform a least-squares fit to obtain the coarse-grained forces. In the context of biological systems, the method has been successfully applied to constructing semiatomistic models of pure phospholipid bilayers[23] and phospholipid/cholesterol bilayers.[24] In addition, it has been applied to model a membrane protein in atomistic detail within a semi-atomistically modeled bilayer.[25]

Several other methods have been proposed based on fitting the interactions to the average structural properties of the detailed simulations.[39,43–46] For dense systems, iterative approaches, such as inverse Monte Carlo (IMC),[45,46] are needed if an accurate reproduction of the structure is desired.[47] The IMC approach has been used by Lyubartsev *et al.* to describe ions solvated in water,[45,46,48] to study the binding of different ions to DNA,[49] and to model self-assembly and the behavior of phospholipids in water.[50] Elezgaray and Laguerre have also used a structural optimization method to construct a coarse-grained model for a pure phospholipid bilayer.[51] Shelley *et al.*[52] employed a similar approach for determining some of the interactions in their model (see also Ref. 19).

Previously, we have applied the IMC approach to construct a very simple model for a phospholipid bilayer containing dipalmitoylphosphatidylcholine (DPPC) and cholesterol.[39] This system was selected since DPPC is one of the most studied phospholipids, and cholesterol is the most important sterol molecule found in the plasma membranes of eukaryotes. Further, the system displays an interesting phase behavior[53] characterized by three main phases (see Fig. 1). It has been suggested that the system might form cholesterol-rich domains[54] or "complexes" of phospholipids and cholesterol.[55,56]

In the previous model,[39] we described each molecule by a single, pointlike particle moving in a two-dimensional plane. The effective interactions were derived using the IMC method to reproduce the radial distribution functions (RDFs) calculated from extensive atomistic molecular dynamics simulations of the system.[14] The model was shown to predict the formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the expected two-phase coexistence.

In this article, we wish to extend the previous study to a more complex model. This attempt is motivated by several reasons. Most importantly, we wish to improve the structural representation of the DPPC molecules since the molecules are typically not radially symmetric as they were in the earlier model. In the new model, we still describe each cholesterol molecule by a single particle, but each phospholipid molecule is now described by three particles: one for the head group and one for each tail. Intuitively, in addition to being better able to describe the geometry of the DPPC molecules, such a description should better capture the structure
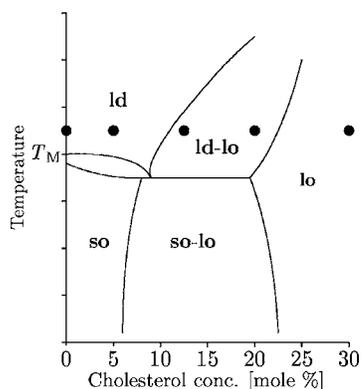
FIG. 1. Sketch of experimental phase diagram for DPPC/cholesterol system (Ref. 53) (reproduced from Ref. 39). At high temperatures and low cholesterol concentrations, there is a liquid-disordered (ld) phase, which is a fluid phase characterized by lipid acyl chains with a low degree of conformational order. When the temperature is lowered, the system goes through the main phase transition at $T_M \approx 314$ K to a solid-ordered (so) phase. The so phase is a solidlike phase in which the acyl chains are conformationally ordered and there is translational order in the plane of the bilayer. Finally, at high cholesterol concentrations, there is a liquid-ordered (lo) phase, characterized both by a high degree of acyl chain ordering and the lack of translational order. At intermediate cholesterol concentrations there are wide ld-lo and so-lo coexistence regions. The dots represent the concentrations at which the atomic-scale molecular dynamics simulations have been performed.

of possible phospholipid-cholesterol complexes. This is because one of the most important effects of the cholesterol molecules is to order the *nearby* tails; i.e., it may not necessarily order both tails of a given molecule in an identical fashion. The ability to describe the interactions better could also have an effect on the transferability of the interactions, i.e., on how well the interactions derived at a particular thermodynamic state point can be transferred to other conditions, e.g., concentration. However, the simplicity of the present model probably still places severe limits on transferability. Nonetheless, systematic studies of the effects of model complexity on the behavior can yield important insight into the coarse-graining process in general. Understanding the improved model also gives a good basis for subsequent improvements, particularly for including the conformational states of the tails in the model.

With the more complex model, we find that simple IMC gives unphysical values for the thermodynamic properties of the coarse-grained model. In particular, the area compressibility and surface tension of the system are not realistic, which may be related to the fact that the solvent is described implicitly through the effective interactions. In this work, we therefore account for these effects by adding thermodynamic constraints to the inverse Monte Carlo procedure to fix the area compressibilities to values found in experiments. We also discuss some other improvements to the standard IMC procedure to increase the speed of the method.

As a starting point for our study, we use the same MD simulations as in Ref. 39.[14] We construct the effective interactions for our coarse-grained (CG) model using the IMC method; i.e., the CG model reproduces the RDFs of the atomistic model. In addition, as mentioned above, experimental area compressibilities are used in the parametrization of the interactions.

We first validate the coarse-grained model by comparing its behavior to that of the atomistic model. Using the constructed coarse-grained model, we then employ Monte Carlo simulations to study the structural behavior of DPPC/cholesterol bilayers at large length scales at cholesterol concentrations ranging from 0% to 30%. The high degree of coarse graining allows us to easily study systems with linear sizes of the order of 100 nm in the plane of the membrane.

We find that the changes to the model are indeed advantageous and that the coarse-graining approach can provide new insights into the large-scale behavior of membrane systems. In particular, we observe regions of higher density of DPPC tails in the liquid-disordered (ld) phase, i.e., in the 0% and 5% cholesterol concentrations. This unexpected observation, not seen with the simpler model,[39] prompted us to look at this issue in more detail by atomistic simulations. Such phenomenon was indeed observed in the atomistic simulations as well, and the results are reported in a separate publication.[57] We also find that the model predicts the formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the simpler model and the experimental phase diagram.[39,53] The differences between the different domains are larger with the improved model, indicating that the better description of the molecular structure is necessary to accurately describe the system. In addition, we explore and discuss the limitations of the model and discuss the general advantages and limitations of the present approach. Finally, we point out some possibilities for future studies and applications.

II. INVERSE MONTE CARLO METHOD

A. Iterative refinement of potentials

The central idea of the inverse Monte Carlo is to iteratively adjust the effective interactions to match the target radial distribution functions more and more precisely.[45,46] The potentials are represented by a piecewise constant grid approximation, and the RDFs are calculated with the same accuracy. Let us denote the values of the potentials in each of these bins with V_α, and let S_α be the number of particle pairs that have that particular interaction energy. Clearly, $\langle S_\alpha \rangle$ is related to the RDF g_α in the same bin: $\langle S_\alpha \rangle = g_\alpha N_p A_\alpha / V$, where N_p is the number of pairs in the system, A_α is the volume of the bin, and V is the total volume of the system.

During each iteration, the derivatives of $\langle S_\alpha \rangle$ with respect to the V_β can be calculated for all pairs (α, β) using statistical mechanics relationships.[45,46] These can then be used to express the changes in $\langle S_\alpha \rangle$ to the first order in terms of changes in V_β as

$$\Delta \langle \boldsymbol{S} \rangle = A \Delta \boldsymbol{V}, \quad A_{\alpha\beta} = \frac{\partial S_\alpha}{\partial V_\beta} = -\frac{\langle S_\alpha S_\beta \rangle - \langle S_\alpha \rangle \langle S_\beta \rangle}{k_B T}, \quad (1)$$

where $\boldsymbol{S}$ and $\boldsymbol{V}$ are the vectors formed from the values S_α and V_β, respectively. This relationship may then be used to calculate the interactions for the next iteration. The procedure is repeated until convergence is achieved.

Conceptually, bonded interactions can also be added to the IMC procedure. However, in this case matrix A has a zero eigenvalue for each bonded interaction. The physical

reason is that the bonded interactions can only be determined up to a constant, and thus the eigenvectors for the zero eigenvalues correspond to situations where a constant is added to one or more bonded potential. Because of the zero eigenvalues, Eq. (1) does not have a unique solution. A simple remedy is to constrain the sum of the changes of each bonded potential to be zero. This constraint can be easily added to Eq. (1): if $Av=0$, $\|v\|=1$ (i.e., v is a normalized eigenvector of A with a zero eigenvalue), then the solution to the equation $(A+vv^T)x=b-vv^Tb$ satisfies $Ax=b$ and $v^Tx=0$.

B. Initial interactions

The initial convergence of the process depends on the initial guess for the effective interactions. Typically, the potential of mean force is used $(V_\alpha^{(0)}=-k_BT \ln g_\alpha)$.[45,46] This works for simple systems, and was also used in this work. However, the more complex the model is, the worse the performance of this initial guess and the convergence of the iteration get. In these cases, some regularization of the procedure is needed (see below). For more complex systems, some other way of constructing the initial potentials may also become necessary. This issue is discussed in more detail in Sec. V.

C. Regularization

As noted above, some regularization of the procedure may be needed if the initial guess for the interaction is poor. The simplest possibility is to multiply the ΔS in Eq. (1) by a small factor r between 0 and 1 to make the changes small enough such that the linear approximation is valid.[46] The value of r can then be increased after a better approximation has been constructed to obtain faster convergence.

Regularization can also be used to speed up the calculations considerably: in the original form, the IMC requires good sampling of the four-particle correlation function $\langle S_\alpha S_\beta \rangle$, which is very time consuming with a large number of effective interactions. If this quantity contains too much noise, the noise is transferred to the potentials and may lead to instabilities in the iteration. This is particularly a problem with bins where the RDFs are small but nonvanishing because their sampling requires extensive simulations.

To reduce the effect of noise in matrix A, we first replace the problem of solving Eq. (1) with the optimization problem of minimizing,

$$\min_{\Delta V} f(\Delta V), \quad f(\Delta V) = \|A\Delta V - \Delta S\|^2. \quad (2)$$

We then add a regularization term to this minimization problem; i.e., instead of minimizing $f(\Delta V)$, we minimize

$$\widetilde{f}(\Delta V) = f(\Delta V) + a\|R\Delta V\|^2, \quad (3)$$

where a is a regularization parameter and R is a regularization operator. The simplest forms of R are a diagonal matrix and the second derivative operator. In these cases, the regularization either limits the magnitude of changes in the potentials or prefers changes that keep the potentials smooth. The best value for a can be selected by trial and error, and it can also be changed during the iteration. Such a regulariza-

tion enables us to use fewer Monte Carlo steps during the initial phase of the iteration and hence to speed up the calculation considerably. When the iteration has converged within the numerical accuracy of the Monte Carlo simulations, the number of steps can be increased to refine the effective interactions further.

D. Surface tension constraint

To minimize finite-size effects to the effective interactions, the simulations during the IMC procedure should be carried out with a system that is identical in size to the system from which the target RDFs were determined.[45] In some cases, the effective potentials produced in this way do not generalize to larger systems. For example, in the present study the effective interactions for the largest cholesterol concentrations are too attractive to maintain a reasonably uniform density in the system. Instead, larger systems form dense clusters separated by empty space. Such behavior is clearly unphysical and may result from the absence of explicit effects of water in the model.

Such condensation effects can be characterized by the surface tension of the coarse-grained model. We define the surface tension γ of the coarse-grained model as

$$\gamma = \frac{1}{V}\left(\langle E_{\text{kin}} \rangle + \frac{1}{2}\left\langle \sum_{i<j} f_{ij}r_{ij} \right\rangle \right), \quad (4)$$

where $\langle E_{kin} \rangle = Nk_BT$ is the average kinetic energy in the system and the latter term is the virial. Here, N is the number of particles, V is the area of the system, and f_{ij} and r_{ij} are the force and the distance, respectively, between particles i and j. If this quantity is close to zero or negative in the simulations on small systems, then the system may not be stable for larger system sizes. This is indeed the case for the highest cholesterol concentrations.

Situations where thermodynamic properties, particularly the pressure, of the coarse-grained model do not match the underlying atomistic model have also been encountered in other coarse-graining approaches.[42,43] Proposed solutions to the problem include iterative adjustment of the pressure followed by reoptimization of the interactions[43] and imposing additional constraints on the instantaneous virial produced by the effective interactions.[42]

In this work, we first tried to implement a simple, iterative adjustment of γ to a desired value in the spirit of Ref. 43. However, the IMC procedure seems to favor a certain value of γ, and hence reoptimization of the structure is not possible if one wants to maintain the surface tension. Hence, we have implemented a constrained version of the IMC procedure, where γ is fixed to a given value. In this version, we solve the minimization problem [Eq. (2)] (or the regularized version) at each iteration with the additional constraint that γ should be equal to the value we specify. A rather wide range of γ values can be imposed without significantly altering the resulting RDFs (the differences are generally not visible on the resolution of, e.g., Fig. 2). However, the potentials do change rather significantly. This shows that although, in principle, the effective potentials corresponding to a given set of RDFs are unique,[58] in practice there can be substantial vari-
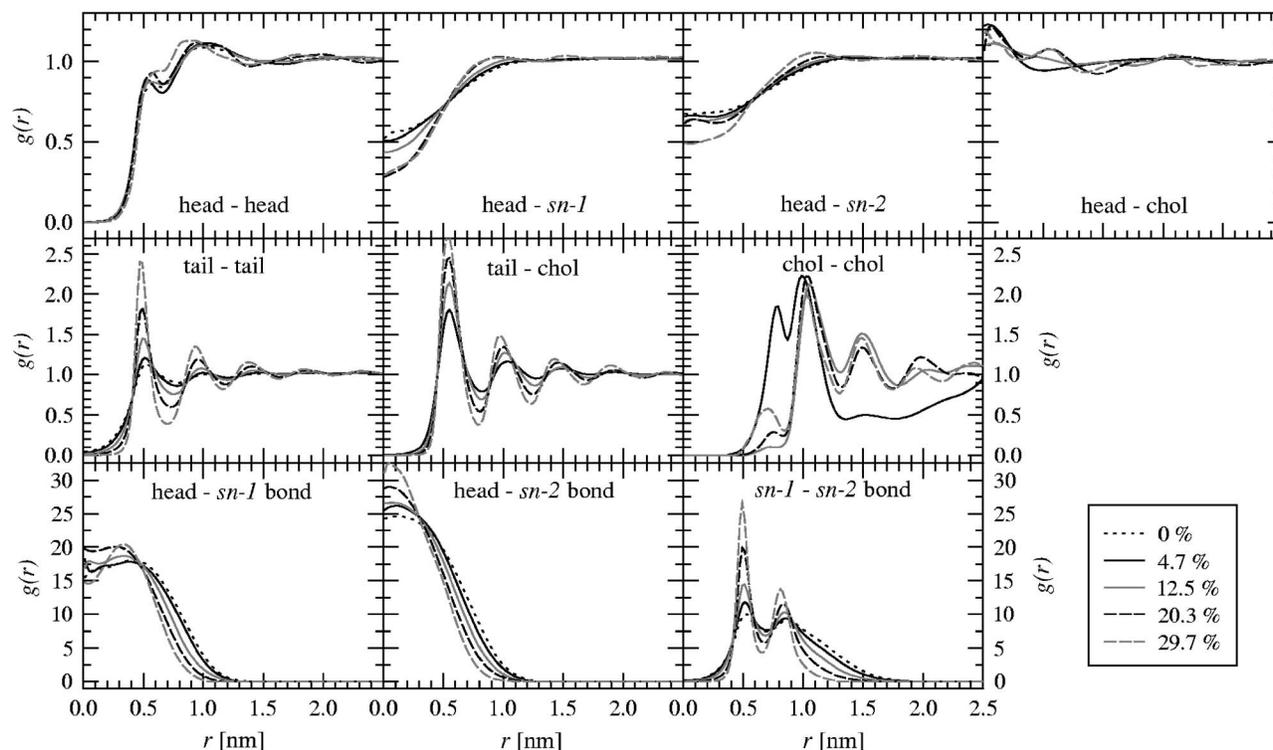
FIG. 2. Radial distribution functions calculated from MD simulations for different pairs of CG structural units. Pairs that have virtually identical RDFs have been merged and only the average RDF is shown. The RDFs have been calculated from the CM positions of the structural units, projected to the plane of the bilayer.

ability in the potentials without significant changes in the RDFs. This aspect is discussed briefly in Secs. III C and V.

One should note that the surface tension defined by Eq. (4) cannot be directly related to the surface tension in the atomistic simulations. This is because the effective potentials are, in general, volume dependent. Hence, the correct value of γ is not necessarily the same as the surface tension in the atomistic simulations, which has been proposed to be zero in equilibrium.[59] There has also been discussion about the value of surface tension that should be used in atomistic simulations of small systems.[60] Because of these considerations, the value of γ has to be fixed using other quantities. In this work, we have used the area compressibility of the bilayer to this end. The details are discussed in Sec. III.

III. MODEL CONSTRUCTION

A. General

The primary goal of the present work is to study the large-scale structural properties of the bilayer, so the degree of coarse graining must be high. Our earlier model[39] contained one particle per molecule, and here we wish to explore whether a somewhat more detailed description would give new insight into the behavior of the system. Hence, we still describe each cholesterol molecule by a single particle, but each DPPC molecule is now described by three particles: one describes the head group and the glycerol parts, and the two others describe the two tails. Each particle in the model describes the center-of-mass (CM) position of the corresponding group of atoms. The tails are not fully equivalent because of the molecular structure of the DPPC. Hence, the system

contains four types of particles and ten different nonbonded pairs. Of these, the tail-tail pairs are very similar (see below), and we use the same interaction for these pairs. In addition, the two tail-cholesterol interactions are assumed to be identical. As a result, the model has seven different nonbonded interactions and three bonded interactions. We include bonded interactions for all pairs in the DPPC molecule, but we do not have any bending potentials. As in Ref. 39, the system is projected onto a two-dimensional plane and the solvent degrees of freedom are integrated out.

The assumptions implicit in the model are the same as in Ref. 39 and are discussed in more detail there. In the model, we assume that the two leaflets of the bilayer interact only weakly, and hence we focus on one of them. Further, we neglect the out-of-plane fluctuations of the bilayer and focus on the lateral organization only. These assumptions allow us to construct a two-dimensional model. We also assume that the interactions between the different kinds of particles can be adequately described by pairwise and radially symmetric effective potentials.

All the effective interactions are constructed using the IMC method. This ensures that the coarse-grained model reproduces the RDFs of the atomistic model. The RDFs describe the structure of the system at short range and contain information about, e.g., the occurrence of different types of particles in the neighborhood of others. They can also be used to characterize the phase behavior of the system. Hence, the coarse-grained model should at least qualitatively reproduce the structure and the phase behavior of the original, atomistic model.

As our focus is on structural properties, we use Monte

Carlo simulations to calculate the equilibrium properties of the system. The simulations are conducted in the canonical ensemble, and the area per molecule is fixed to the average area per molecule in the underlying MD simulations.

B. Results from MD simulations

The first step in constructing the coarse-grained model is to obtain the RDFs for the different pairs of structural units. These can be calculated from atomic-scale MD simulations of the system. We have used the same MD simulation data as in Ref. 39, and only a brief summary is given here. More details can be found in Ref. 14. The simulated systems consisted of 128 macromolecules, i.e., DPPCs and cholesterols, and 3655 water molecules. The simulations were performed with 0, 3, 8, 13, and 19 cholesterol molecules per monolayer, corresponding to approximate cholesterol concentrations of 0%, 5%, 13%, 20%, and 30%, respectively. The studied concentrations are indicated by dots in Fig. 1. Each system was simulated for 100 ns, and the linear sizes in the plane of the bilayer were between 5 and 7 nm. The simulations were performed in the NpT ensemble at $T=323$ K using the GROMACS simulation package.[61]

The RDFs calculated from the MD simulations are shown in Fig. 2. The RDFs have been calculated for the CM positions of the different structural units, projected to the plane of the bilayer. The two monolayers have been considered separately, and the RDFs have been averaged over the monolayers. The first 20 ns of the MD data was regarded as equilibration as in Ref. 39. The RDFs were calculated up to $r=2.5$ nm, which is less than half the linear system size for all studied concentrations. The errors of the RDFs can be estimated to be of the order of a few percent, with somewhat higher errors at low cholesterol concentrations for the RDFs involving cholesterol.

The RDFs for all different combinations of tail particles are virtually identical. Similarly, the tail-cholesterol RDFs are nearly identical irrespective of which tail is being considered. Hence, only the average RDFs are shown for these pairs in Fig. 2. For the head-tail RDF there is a small difference between the *sn*-1 and *sn*-2 chains, and thus they are shown separately. For the nonbonded RDFs, all bonded pairs, i.e., pairs within the same molecule, have been excluded.

For all concentrations, the RDFs indicate liquidlike behavior: there are broad peaks at short length scales and the RDFs approach unity at larger r. This confirms that we are probing the liquid region of the phase diagram where the system is either in the ld and liquid ordered (lo), or coexisting ld and lo phases (see Fig. 1).

The RDFs show more or less systematic changes as the cholesterol concentration is increased. This is most clearly visible for the tail-tail, tail-cholesterol, and the bonded RDFs. In all these cases, increased cholesterol concentration leads to sharper peaks, meaning a more well-defined structure. In addition, more peaks appear at higher cholesterol concentrations, meaning that the range of the ordering is increased slightly. Similar behavior can also be seen in the head-head and head-tail RDFs, but here the changes are
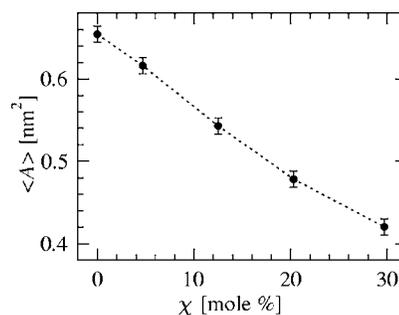

FIG. 3. Average area per molecule as a function of cholesterol concentration, calculated from MD simulations (Ref. 14).

much smaller, particularly for the head-head RDF. For the cholesterol-cholesterol RDF, the 5% case is different from the rest, but for the other concentrations similar conclusions hold. It is also interesting to note that as the cholesterol concentration is increased, a new peak appears at small r, indicating a change in the neighborhood structure. Finally, the changes in the head-cholesterol RDF are not as systematic. Nonetheless, the increase of the range of the ordering as well as sharper peaks can also be seen in this case.

Due to the fact that the RDFs have been calculated for the projected CM positions of the particles, not all RDFs approach zero at the origin. This is the case for the head-tail, head-cholesterol, and tail-tail RDFs. For the first two cases, it is easy to imagine a configuration where the CM of the flexible head group is essentially on top of a nearby tail or cholesterol particle. For the tail-tail case, the flexibility of the tails would allow the CM positions to overlap. This is also supported by the fact that the overlap decreases with increasing cholesterol concentration, i.e., when the chains become more ordered and rigid.

In addition to the RDFs, we need to obtain the area per molecule for each cholesterol concentration to fix the area of the simulation box in the canonical Monte Carlo simulations. These have been published previously,[14] and they are shown in Fig. 3 for reference. The area decreases monotonically with increasing cholesterol concentration, which is in agreement with other simulation studies[62,63] and related experiments.[55]

C. Effective potentials

Based on the RDFs in Fig. 2, we have constructed the effective interactions for our CG model using the IMC method described in Sec. II. As noted above, all the tail-tail pairs are similar to each other, as are the tail-cholesterol pairs. Hence, only one interaction was used for all the tail-tail interactions, and similarly a single interaction was used to describe the tail-cholesterol pairs. The cutoff of the interactions was set to 2.5 nm. This is the largest cutoff that is allowed by the constraint that the RDFs may only be calculated up to half the linear size of the simulation box.

As described in Sec. II, it is necessary to impose an additional constraint on the surface tension of the CG model to obtain physically feasible interactions for the highest cholesterol concentrations. Although this solves the immediate problem, the next question is how to select the value for the
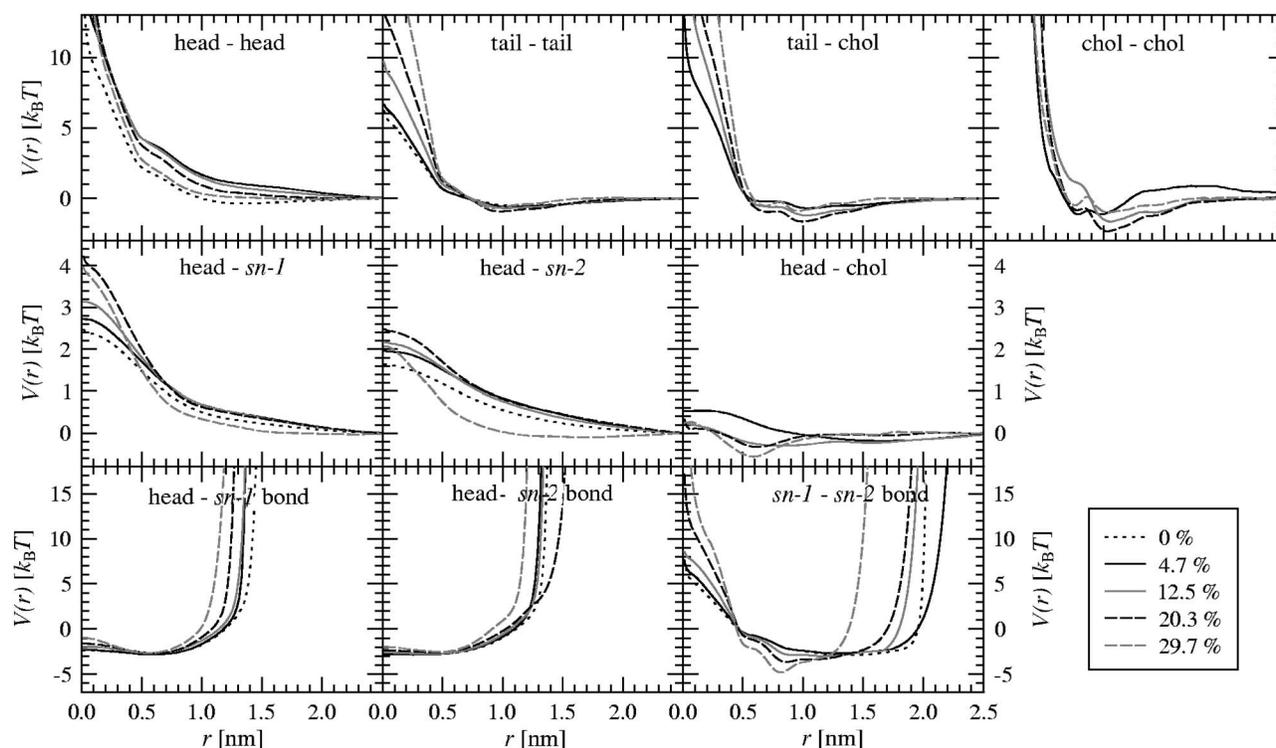
FIG. 4. Effective interaction potentials for different pairs of coarse-grained particles and different cholesterol concentrations. The bottom row shows data for intramolecular interactions through corresponding effective interactions.

surface tension γ. In this work, we have used the experimental area compressibilities of DMPC/cholesterol bilayers[64] to obtain a sensible value for γ. The experiments indicate that the area compressibility for a single-component DPPC system above T_M could be about 150 dyn/cm. The compressibility increases for increasing cholesterol concentration and reaches around 600 dyn/cm for 40% concentration (for details see Ref. [64]).

Since the area compressibility is a macroscopic quantity, it cannot be reliably calculated from the small systems used in the IMC. Hence, we employ a two-step iterative process to obtain γ. The basic idea is the following: we first select a value for γ and use the IMC to obtain the interactions. We then perform a simulation for a larger system and use a block analysis method[65,66] to extrapolate the area compressibility to an infinite system. We then adjust the value of γ and continue until the area compressibility is of the right order of magnitude. In practice, we do not require quantitative agreement since the process is computationally rather costly, and we are not exactly at the experimental conditions. Further, this analysis does not account for the (unknown) density dependence of the effective interactions. The qualitative results are also the same irrespective of the precise value of γ. For these reasons, we are content with γ values that reproduce the right order of magnitude for the area compressibilities and also reproduce the general trends of the area compressibility as a function of cholesterol concentration.

Due to the large-scale heterogeneity of the system at 13% and 20% concentrations (see Sec. IV), the extrapolation of the area compressibility was not computationally feasible for these systems. Hence, we first determined the γ values for the pure DPPC system and the 30% concentration case.

The above process gives $\gamma = 20$ SU (simulation units) (1 SU ≈ 4.46 dyn/cm) for the pure DPPC and $\gamma = 50$ SU for the 30% concentration. Since the value $\gamma = 50$ SU gives a reasonable value for the compressibility in the 5% concentration as well (approximately halfway between the pure DPPC and the 30% concentration values), we have used this surface tension for all the systems containing cholesterol.

The RDFs and the effective potentials were pre- and postprocessed as in Ref. [39] with the exception that no shifting functions were used in the present study. The RDFs were smoothed using a spline-fitting algorithm[67] before feeding them to the IMC algorithm, and the same algorithm was used to smooth the resulting effective potentials. In addition, the effective interactions in the regions where the RDFs vanish were fitted with a power-law form such that the potentials and their first derivatives were continuous. In the present study, no shifting was used because most of the interactions approach zero smoothly at the cutoff, and hence there are no large discontinuities.

Figure 4 shows the final effective potentials for different pairs of particles and different cholesterol concentrations. Due to the high degree of coarse graining, the effective potentials are soft. In particular, the interaction energies between the head and the other particles are only of the order of a few k_BT. This is understandable because the region that the head groups occupy is more or less separate from that of the other particles. The only exception is the case between cholesterol-cholesterol pairs, where we find interactions to be hard (see Fig. 4). This results from the steroid moiety of cholesterol being very rigid, and the cholesterol molecules standing approximately in an upright position along the bi-

layer normal. As a consequence, the CM positions of cholesterol molecules projected onto the bilayer plane cannot overlap.

The potentials change more or less systematically as the cholesterol concentration increases. The head-head potential becomes gradually less repulsive as the cholesterol concentration is reduced, while the shape of the potential remains the same. The pure DPPC interaction is an exception: it is somewhat less repulsive than the others, but the shape is still the same. The main reason for this deviation is probably the difference in the selected surface tensions. When surface tension is varied, it mostly affects the interactions involving the head particle, and the other interactions hardly change. The head-tail interactions become more repulsive with an increasing cholesterol concentration, in particular at small r. The 30% case is less repulsive than the others at larger r, but at smaller r it is the most repulsive. The head-cholesterol interaction is very weak, and it becomes gradually attractive as the cholesterol concentration is increased. In addition, it develops a clear minimum at higher cholesterol concentrations.

The tail-tail, tail-cholesterol, and cholesterol-cholesterol interactions differ from the other interactions in that they have an attractive region at intermediate r. The form of these interactions is also more complex. These interactions behave similarly to each other: at short ranges they become systematically more repulsive as cholesterol is added, and at intermediate ranges they become more and more attractive until 20% concentration, and the 30% case is again more repulsive.

The bonded interactions change only little with the cholesterol concentration. Generally, they become steeper with an increasing cholesterol concentration, giving the molecules a more rigid shape. This is in agreement with the RDFs in Fig. 2. One should note that the appearance of the interactions at large r depends on the details of the power-law fit, and no conclusions should be based on it. However, the precise shape of the interaction in this region does not matter as long as it is attractive enough because the RDFs are practically zero in this region.

A comparison of the interactions in Fig. 4 with those in Ref. 39 yields valuable information about the different models and the behavior of the coarse-graining method. First, we note that the interactions are roughly similar. Both are soft, and the forms of the interactions are similar (similar amount of detail, etc.). In particular, the cholesterol-cholesterol interactions are very similar. An interesting difference can be seen when comparing the behavior of the interactions as a function of cholesterol concentration. In the present case, the concentration dependence seems to be significantly weaker. This can be explained by the reduced complexity of the CG particles; i.e., fewer degrees of freedom are integrated into the effective interactions. As noted in the Introduction, it may also be that the ability of the model to better represent the fundamental interacting blocks reduces the concentration dependence of the interactions.

As noted in Ref. 39, the results are not sensitive to the details of how the interactions are constructed. We have verified that this is also the case for different values of the regu-
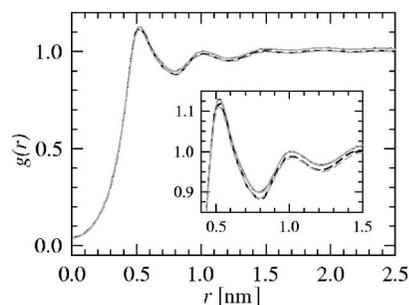

FIG. 5. Comparison between RDFs from CG model (gray lines) and atomistic MD simulations (black lines) for a pure DPPC system. The solid lines are for small systems (64 molecules per monolayer) and the dashed lines for larger systems (576 molecules for MD, 16 384 molecules for CG). Only the tail-tail RDF is shown, but the results are similar for the other cases. The inset shows a zoomed view of the first peaks.

larization parameters and different surface tensions. The reported results do not change qualitatively if the surface tension is varied within a reasonable range, although some details such as heights of different peaks may slightly change.

D. Validation

The coarse-grained model can be validated by comparing it to the atomistic simulations. By construction, the model reproduces the short-range structure of the underlying simulations. We have also verified that the RDFs do not significantly change when the system size is increased from the one that was used for constructing the interactions. These results are demonstrated for the pure DPPC system in Fig. 5: the RDFs from the CG and MD simulations are nearly identical. The results are similar for other concentrations and other pairs of CG particles. Figure 5 also shows the RDFs calculated from a larger MD simulation of the same system.[57] These show that the RDFs for both the MD and CG simulations change similarly when the size of the system is increased. The RDFs also have the characteristic structure of a fluid, confirming that the system is in a fluid state. All these results are in agreement of the findings of the earlier model.[39]

IV. BEHAVIOR AT LARGE LENGTH SCALES

The coarse-grained model can be used to study the behavior of the bilayer system at much larger length scales than those accessible by more detailed simulation methods. The following Monte Carlo studies were mostly conducted for systems whose linear size was 16 times that of the original molecular dynamics simulation. This means that the linear system size was in the range of 80–110 nm, depending on the cholesterol concentration. Such systems contained 39 424–49 152 particles, and reasonable statistics could be obtained within a few days on a standard desktop computer. In Ref. 39, it was estimated that with the simpler model, the computational gain over MD simulations was eight orders of magnitude. From this, we can estimate that the present model is around seven orders of magnitude faster than MD, since
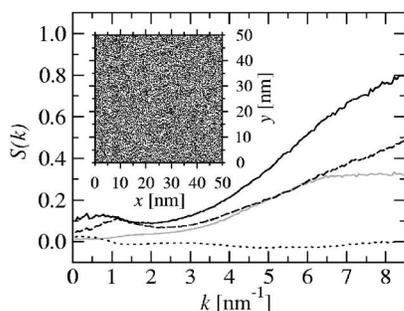
FIG. 6. Static structure factors for pure DPPC system. The solid black line shows the total structure factor, and the other lines show the head-head (solid gray line), tail-tail (dashed line), and head-tail (dotted line) structure factors. The inset shows a snapshot of the system. In the snapshot, only the tail particles are shown. The CG simulations were run with 16 384 molecules, i.e., with 16 times the linear size of the original 128 lipid system.

the increased number of particles per molecule and the increased density of particles increase the computational cost slightly.

A. Pure phospholipid bilayer

The long-range structure of the system is conveniently characterized by the static structure factor defined by

$$S(\mathbf{k}) = \frac{1}{N} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \exp\{-i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_i)\} \right\rangle, \quad (5)$$

where N is the number of particles, $\mathbf{r}_i$ are the positions of the particles, and $\mathbf{k}$ is a vector in the reciprocal space. The structure factor can also be calculated for a subset of particles by restricting the summations to these particles. The structure factors are related to the RDFs by Fourier transforms, and hence in principle carry the same information. However, the long-range structure is more conveniently characterized using the structure factors.

Figure 6 shows the circularly averaged static structure factors for the pure DPPC system. In addition to the total structure factor calculated over all pairs of particles, structure factors calculated over the head-head, head-tail, and tail-tail pairs are shown. The inset shows a snapshot of the system, displaying the locations of the tail particles. The head particles are not shown for clarity.

The total static structure factor in Fig. 6 has an interesting feature at the small k values: there is a clearly visible, broad peak around $k \approx 1$ nm^{-1}. A more careful look at the different components of the structure factor reveals that the peak is completely due to the tail-tail structure factor. Such a peak indicates that the tail density should have some density variation with a characteristic length scale around $l \approx 6$ nm. An examination of the snapshot in Fig. 6 indeed shows such small variations. The feature does not change with longer equilibration or larger system size, and it is thus not likely to be an artifact of the simulation. In addition, a similar peak can be seen in the 5% cholesterol concentration [see below and Fig. 7(a)].

The presence of such a feature in the coarse-grained simulations prompts us to look more carefully at the underlying atomistic simulations to check whether such density differences are present there as well. The length scale associated with the density variation is similar to the size of the bilayer in the MD simulations, so only limited information can be gained from the atomistic simulations. However, the atom-scale MD simulations for a small bilayer of 128 lipids were found to show density variation reminiscent of the behavior of the CG model despite the limited size of the simulation box.

Importantly, the above result for transient domain formation in a single-component lipid bilayer was indeed first found by the CG model. In a way, we considered this as a prediction which we wished to test through more extensive atomistic simulations for a considerably larger bilayer system. For this purpose, we employed a fully hydrated single-component DPPC bilayer containing 1152 lipid molecules. Results of that study have been reported and discussed elsewhere.[57] Here, it suffices to note that the larger atom-scale MD systems do show variations in the area per tail that are similar to that observed in the CG model. The MD system also allows us to assess the ordering of the tails at the different densities. Such a study indicates that the denser regions also display higher ordering and that the system has highly ordered dynamic domains. There are also some differences between the MD and the CG results: in the MD system, $S(k)$ does not have a clear peak but a steady increase towards smaller k values, and the peak is much higher. Reasons for the former can include the relatively small size of the MD simulation box and, in part, effects from the cutoff of the effective interactions (see below). For the latter, it may be the case that in the smaller MD simulations the density fluctuations do not have enough space to develop fully. Also, the lack of description of the ordering of the lipid tails may limit the ability of the CG model to describe the phenomenon.

The position of the observed peak at small k is close to the size of the original MD simulation box for a system with 128 DPPCs. However, the peaks for the pure DPPC and the 5% cholesterol systems are virtually identical, although the MD simulation boxes are of different sizes [albeit the difference in the linear sizes is of the order of 3%–4% (see Fig. 3)]. We have also confirmed that the peak in $S(k)$ occurs at the same position if the coarse graining is done using the RDFs from the larger MD simulations employing 1152 lipids reported in Ref. 57 (see below). This strongly indicates that the size of the original simulation box is not the origin of the peak. It also means that the small size of the underlying atomistic simulations cannot by itself explain the differences observed between the larger MD and CG simulations.

The position of the peak is of the same magnitude as twice the cutoff of the effective interactions. To study this, we have constructed several different interactions with different cutoffs. As the starting point, we used the RDFs calculated from the last 5 ns of the 1152 lipid PME simulation in Ref. 57. Cutoffs of 2.5, 3.0, and 5.0 nm were used.

When the cutoff of the interactions is increased, the position of the peak moves towards smaller k and its height increases slightly (figure not shown). When the cutoff is increased from 2.5 to 5.0 nm, the peak moves from $k \approx 1$ nm^{-1} to $k \approx 0.7$ nm^{-1}. The latter is significantly larger than half the k value for the 2.5 nm cutoff, showing that the
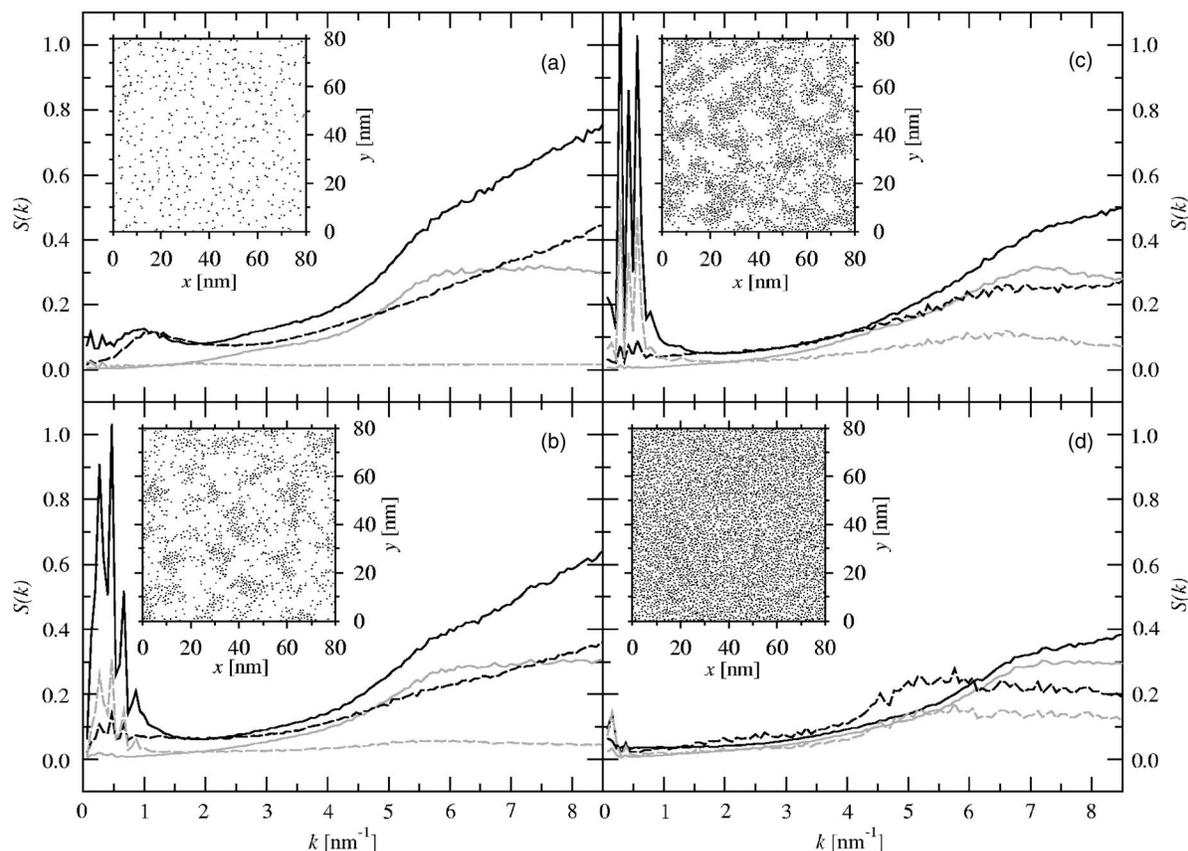
FIG. 7. Static structure factors for different cholesterol concentrations: (a) 5%, (b) 13%, (c) 20%, and (d) 30%. The solid black line shows the total structure factor, and the other lines show the head-head (solid gray line), tail-tail (dashed black line), and cholesterol-cholesterol (dashed gray line) structure factors. The other pairs are not shown for clarity. The insets show snapshots of the systems. In the snapshot, only the cholesterol particles are shown. For the 30% concentration, the peak at the smallest k values is probably due to very slow equilibration of the system.

position of the peak is not related simply to the cutoff. Some insight into the results can be gained from the way the effective interactions are determined. Due to the IMC process, the effective interactions reproduce the atomistic RDFs up to the cutoff. Since the structure factors are related to the RDFs by a Fourier transform, also the atomistic structure factors are reproduced up to $k_c = 2\pi/r_c$, where r_c is the cutoff (assuming that there are no correlations at these wavelengths beyond r_c). Hence, the onset of the peak on the high k side is determined more and more precisely as the cutoff is increased, which also affects the shape of the peak, as well as its allowed position. In all cases, the largest fluctuations form at length scales longer than the cutoff, although in the 5.0 nm case the peak is already starting to rise within the cutoff. These results demonstrate that although the qualitative results of the CG simulations are robust, care should be taken before interpreting the results quantitatively.

It is interesting to note that no indication of density variation was observed with the simpler model in which a single particle was used to describe one DPPC molecule.[39] The above results show, and the atomistic simulations in Ref. [57] confirm, that this effect is best characterized by its effect on the structure of the tail region. Hence, the simpler model, where the tails are not treated separately, may miss crucial structural information that is needed to obtain the effect. This highlights the effect that the degrees of freedom selected for the CG model can have on the properties of the model.

B. Organization of cholesterol

Figure 7 shows the static structure factors for the systems containing cholesterol. The figures show the total static structure factors and the head-head, tail-tail, and cholesterol-cholesterol structure factors for different cholesterol concentrations. The other pairs are excluded for clarity. In addition, a snapshot of the cholesterol organization is shown for each concentration. As the cholesterol concentration is increased, the system equilibrates more and more slowly (all the simulations were started from an independent random configuration). This is likely to be the reason for the small peak in the structure factors at the smallest k for the 30% concentration: the height of the peak is still slowly decreasing.

Both the snapshots and the structure factors clearly show that the distribution of cholesterol is not uniform at 13% and 20% concentrations. In contrast, no indication of such nonuniform distribution is seen for the 5% or 30% cases. Such formation of cholesterol-rich and cholesterol-poor domains at 13% and 20% is in line with the phase diagram of the system:[53] at the intermediate concentrations we should expect that the system displays coexisting ld and lo phases (see Fig. 1). Although we cannot directly assess the ordering of the tails within our model, we can argue as follows that the cholesterol-rich regions correspond to the lo phase. The local ordering effect of the cholesterol on the nearby phospholipid tails has been clearly demonstrated by atomistic

simulations.[14,63,68,69] Hence, the nearby cholesterols cause the tails in the cholesterol-rich regions to be more ordered. Further, the cholesterol-rich regions have more tails per a certain area than other parts of the system, which also indicates that the tails are more ordered.

For the systems containing cholesterol, it is much more difficult to observe domains in the atomistic systems, compared to the pure DPPC. This is due to the fact that the cholesterol molecules need to diffuse to the domains instead of just altering the local packing. Hence, for the domains to form, very long simulations would be needed, in particular since the addition of cholesterol slows down diffusion.[14] However, the above argument about the local ordering effect of cholesterol also gives indirect support for the differences in the area per lipid between the different domains. Namely, higher ordering of the lipid tails has been shown to lead to a smaller area per lipid.[14,70] Hence, the cholesterol-rich areas should also have a smaller area per lipid in atomistic simulations, if we could observe them.

Long correlation times of structures at large length scales make it computationally very demanding to determine the precise shape of the $S(k)$ at small k values (see below), even with the coarse-grained model. Hence, we limit our discussion to the more qualitative aspects. The position of the peak in the 13% and 20% concentrations corresponds to length scales of the order of 10 nm or more, which is also the size of the domains seen in the snapshots. As expected from the snapshots, the peak is strongest in the cholesterol-cholesterol $S(k)$, but it is also visible in all the other structure factors. However, it is relatively small for the pairs containing head particles. Hence, the cholesterol-rich domains are also characterized by a higher density of the tail particles and a slightly higher density of the head particles.

The length scales associated with the domains are comparable to the size of the simulation box, even in the CG simulations. Hence, the current simulations do not allow us to make any definite conclusions about the size of the domains. Further, the shape of the $S(k)$ peak can change between independent runs, suggesting that the correlation times for the largest structures are very long. Hence, a significant amount of computer resources (or, alternatively, more efficient Monte Carlo moves) would be needed to determine the shape of the peak accurately. For the same reason, a reliable quantitative analysis of the cholesterol distribution would require significantly longer simulations. In the light of the results for the pure DPPC system, we would also need to check whether the effect of the potential cutoff affects the exact results. Since we are focusing here on more qualitative aspects, such simulations have not been performed for the present study. However, the value of the $S(k)$ in Figs. 7(b) and 7(c) seems to be significantly lower for the smallest k values than for the maximum of the peak, indicating that the size of the simulation box may not be the limiting factor for the size of the domains in the present simulations.

We also note that the 5% case shows a peak at $k \approx 1$ nm^{-1} very similar to the pure DPPC case (Fig. 6). For the higher concentrations, no evidence of such a peak is found. This indicates that such density variations are a specific property of the ld phase and do not occur in the lo phase. This is perhaps not surprising because the lo phase is much more densely packed, and thus there is less space to allow for such variation. The lack of the peak for concentrations where coexistence is expected could be explained by the fact that the coexisting domains seem to be of the same order of magnitude in size as the tail density variation. Hence, the formation of denser and more ordered domains in the ld phase could be suppressed by the presence of a nearby region of the lo phase.

It is again interesting to compare the behavior of the current model to the simpler model in Ref. 39. Qualitatively, the conclusions are identical with the exception of the 5% peak. However, with the present model the cholesterol-poor regions are virtually depleted of cholesterol, which is a much stronger effect than that seen with the simpler model. This is also reflected in the height of the peaks in the $S(k)$. The difference may be attributed to the ability of the new model to better describe the tail-cholesterol interactions, which leads to a stronger preference for domain formation.

V. DISCUSSION

The results for the pure DPPC bilayer provide a prime example of the benefits of the present coarse-graining approach. The constructed model allows us to easily study the structure of the system at larger length scales and to focus on the most important aspects of the system. With the coarse-grained model, the observed denser domains are easy to spot, and this knowledge can then be used to guide subsequent studies of the system. The results also provide additional confidence in that the qualitative results given by this approach, such as the formation of cholesterol-rich domains, are indeed correct. As for the interpretation of the quantitative results, one has to be more careful. As shown for the pure DPPC case, the cutoff of the interactions, among other factors, can have an effect on the quantitative behavior.

Based on the present simulations, it is difficult to make quantitative conclusions about the sizes of the domains. This makes it also difficult to determine what the exact physical mechanism behind the formation of domains is. For the pure DPPC, the equilibration is rather fast since only the packing of the molecules needs to be altered to form the domains. This allows the domains to be seen even in atomistic simulations.[57] For the cholesterol domains, the equilibration is much slower since the cholesterol molecules need to diffuse for domains to form or change. Due to the slow equilibration of the largest length scales, one possibility is that the snapshots in Fig. 7 only show an intermediate state in a process of complete phase separation. It is also possible that the domains are similar to the pure DPPC domains, i.e., thermal, critical-like fluctuations amplified by competition between thermal disorder and the tendency of cholesterol to order the nearest lipids. Also, the interplay between cholesterol and the small ordered domains seen in pure DPPC may be relevant for the formation of the cholesterol-rich domains. In the latter cases, thermal disorder can be a limiting factor for the sizes of the domains.[57]

To assess the role of the underlying MD simulations on the coarse-grained model and its behavior, we have also con-

structed an alternative set of interactions for the 20% concentration. As the starting point, we used the atomistic simulation reported in Ref. 15. The main difference between that simulation and those used in the present study is in the cholesterol force field: the cholesterol is slightly more rigid. There are some minor differences in the behavior of the models, e.g., the height of the structure factor peak at small k is slightly lower with the alternative interactions, but no qualitative differences are observed. Hence, the conclusions presented are robust concerning small differences in the atomistic simulations.

One of the major concerns in the construction of coarse-grained models is the dependence of the effective interactions on the thermodynamic state of the system. Typically, the interactions derived at some temperature, density, and concentration can only be used to describe a limited vicinity of that state point. It is generally difficult to determine the limits for such transferability of the effective interactions *a priori*, i.e., if the desired behavior is not known. In principle, different effective interactions can be constructed for all different state points, but this can be a time-consuming process. Hence, understanding the limitations of the transferability is crucial for understanding the applicability of coarse-grained models.

To this end, we have studied the limits of the transferability of the interactions in the current model. This was done by using the potentials derived for a certain concentration, say, 13%, to simulate a system at the neighboring concentrations, i.e., 5% and 20% for the 13% case. These studies were carried out for systems with linear sizes in the range of 40–55 nm, i.e., eight times the linear size of the original system. The results are generally the same as in Ref. 39: qualitative behavior is mainly determined by the interactions and not the concentration.

For example, the potentials for the 20% concentrations, i.e., for a system in the coexistence region, also give coexistence for the 30% concentration where the system should be homogeneous. Nevertheless, the proportion of the cholesterol-poor regions decreases significantly, which is qualitatively correct. Also, the changes in the short-range structure are in the correct direction, but the magnitude of the changes is too small. Similarly, the potentials from the 30% concentration give a homogeneous system also at the 20% concentration, but the changes in the short-range structure are qualitatively correct. Identical conclusions hold for transferability between the 5% and 13% concentrations. Hence, the effective interactions cannot be used, e.g., to map the phase boundaries of the system.

Transferability between the 13% and 20% concentrations, i.e., within the coexistence region, is better. Here, the domains persist, and the qualitative change in the height of the $S(k)$ peak at small k is correct. As in the earlier cases, changes in the short-range structure are also qualitatively correct, and the proportion of the system in the cholesterol-poor phase decreases with increasing cholesterol concentration. Finally, transferability from the 5% concentration to the pure DPPC system is similar to that within the coexistence region. However, in this case the comparison is complicated by the differences in the surface tensions.

Poor transferability of the potentials between different regions of the phase diagram can be easily understood. Due to the different average ordering of the tails in the ld and lo phases, the typical conformations of the atoms forming the tail particles are very different in the different phases. Similar although smaller differences may also be present for the other particles. Due to the way the model is constructed, all these differences need to be incorporated in the effective interactions because only they can change as the concentration is changed. Hence, for example, the effective interactions derived in the lo phase are tuned to describe the interactions of ordered chains, and are thus unable to adequately represent systems where the chains are less ordered. This conclusion is supported by the fact that the transferred interactions give at least a qualitatively correct behavior within similar regions of the phase diagram. The small differences can again be attributed, at least partially, to the changes in the average order in the system due to changes in the cholesterol concentration.

The above deduction implies that including the ordering of the tails to the model could help with the transferability issues. This is particularly so because one of the main effects of cholesterol is to increase the order of the chains. The simplest way would be to include the conformational degrees of freedom in the spirit of the phenomenological model of Nielsen *et al.*,[32] i.e., by giving each chain two possible states, describing an ordered and a disordered state. The results of Nielsen *et al.* suggest that already such a simple description of the conformational states can give the correct form for the phase diagram. Hence, a systematic construction of such a model could give additional insight into the system.

A major drawback of the IMC method is the requirement that the four-particle correlation functions should be sampled adequately during each iteration. This limits the applicability of the method to systems with relatively few distinct interactions as the required memory and computational time grow very fast with an increasing number of interactions. If we want to keep the requirement that the RDFs are reproduced, then there is relatively little we can do. An iterative refinement procedure is needed already for relatively simple systems containing just ethanol and water to accurately reproduce the RDFs at high ethanol concentrations.[47] The Boltzmann inversion scheme used in that study, and also frequently in polymer physics,[43,44] is computationally much lighter than IMC since only the RDFs are used in the iteration. It neglects the correlations between the different RDFs, and can thus converge poorly for more complex systems. However, such an *a priori* comparison is complicated by the fact that the Boltzmann inversion does not rely on a linearization of the changes in RDFs, in contrast to IMC. Hence, Boltzmann inversion could be a better alternative to IMC for more complex systems.

It could be possible to make technical improvements to the IMC procedure to improve its speed, but these could have an effect on the convergence, and hence a good compromise would require experimentation with each particular system under study. In particular, physical insight into the problem could be used to assume that the four-particle correlation functions vanish for certain pairs of bins. Avoiding

calculation of large parts of matrix A in Eq. (1) could improve the speed substantially. The matrices occurring in the present problem indeed have regions composed mostly of noise, and hence this approach could help. We could also keep some of the potentials fixed while the others are being iterated, and then use this to determine groups of potentials sequentially as in Ref. 43. However, both of these approaches are expected to work best when we are close to the correct interactions; i.e., they do not solve the problem of poor initial potentials (see below). It could also be possible to optimize the potentials first with a coarser grid (possibly using simple splines within the bins) since the total number of bins is the critical quantity in determining the speed of the method. The obtained approximate potentials could then be used to construct an initial approximation for a second iteration with the desired accuracy for the potentials.

The problem with the speed of the IMC is partially associated with the use of the potentials of mean force as the initial interactions. For simple systems, they work relatively well, but for more complex systems they may produce a structure that is too wide off the target to allow for any sensible refinement (e.g., the system may be in the wrong phase). Another problem stems from the linear approximation used in the IMC: it may not work that well if the desired change in the structure is too large. Already, in the present case we had to resort to very strong regularization during the initial phases of the iteration to obtain convergence for the above reasons. Hence, it could be worth the effort to construct better initial interactions for more complex systems. One possibility would be to use the Ornstein-Zernike integral equation together with a closure relation[71] to obtain the initial approximations for the potentials, and then fine-tune these with the IMC method. This could help in achieving convergence without the need to extensively tune the regularization parameters. Another interesting starting point would be to produce the initial set of interactions using the force matching procedure of Voth and co-workers.[41,42] Such studies could also yield valuable insight into the similarities and differences of the different approaches. An interesting alternative for constructing the initial interactions has also been suggested by Elezgaray and Laguerre: one could first optimize the effective interactions for simpler systems where the interaction centers are solvated in water, and use the resulting interactions as initial guesses for the next model.[51]

A separate question that may be asked of any coarse-grained model is the representability of the effective interactions. Namely, for a particular many-body system, there may not be unique effective pairwise interactions that could reproduce all quantities of interest.[72] In the present case, we have constructed the effective interactions to reproduce the RDFs, i.e., the structure. We have then used the model to study other structural properties, and the method used for the calculation of area compressibilities is based on the RDFs,[65,66] so major problems are not expected. However, if we would like to calculate other quantities, these issues would have to be considered. In particular, we found in this study that by tuning the surface tension we can produce a rather wide range of interactions that give essentially the same RDFs. In general, it is not trivial to say which of them is the "best" for a given purpose, especially due to minor fluctuations in the data analyzed. The qualitative results for the structural quantities do not seem to depend on such details of the interactions, but for other quantities more studies are needed. A systematic comparison of the different coarse-graining approaches could also be interesting from this point of view.

VI. SUMMARY AND CONCLUSIONS

In this article we have refined our earlier coarse-grained model[39] for DPPC/cholesterol bilayers to include a more detailed description of the DPPC molecules. We have used the inverse Monte Carlo approach to construct the effective interactions for the coarse-grained particles such that the radial distribution functions of atomistic simulations are reproduced. We have also improved the IMC method by including a thermodynamic constraint into the procedure and by improving the convergence properties to speed up the method. The IMC method is an appealing alternative for constructing relatively simple coarse-grained models because it can construct models that reproduce the RDFs of an underlying model. As the RDFs can be used to characterize the phase behavior, the CG model should then also reproduce at least the qualitative phase behavior of the original model.

The constructed coarse-grained model allows us to study the structure of the bilayer at much larger length scales than accessible through atomistic simulations. We have found that the refined model is able to predict the formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the original model. However, the organization is much more pronounced with the new model. In addition, the new model is able to predict the formation of denser domains in pure DPPC bilayers, which has also been confirmed in a separate publication using large-scale atomistic simulations. We have also discussed different aspects of the present approach.

The present as well as earlier results demonstrate that the coarse-graining approach based on the IMC is very promising. The current study could be carried further by including the conformational degrees of freedom to the model. This could further reduce the concentration dependence of the effective interactions. It could also give further insight into the behavior of the system. Another possible direction would be to look at bilayer systems with different compositions and compare the properties of different systems. As demonstrated here, already quite small systems can be used as a starting point to obtain relevant information on the large-scale behavior of the system. Finally a comparison of different approaches to coarse graining could yield valuable insight into the coarse-graining process itself. Such information could help in constructing better models and better methods in general.

ACKNOWLEDGMENTS

[1] *Structure and Dynamics of Membranes: From Cells to Vesicles*, edited by R. Lipowsky and E. Sackmann (Elsevier, Amsterdam, 1995).

[2] *Lipid Bilayers: Structure and Interactions*, edited by J. Katsaras and T. Gutberlet (Springer-Verlag, Berlin, 2001).

[3] O. G. Mouritsen, *Life: As a Matter of Fat*, The Frontiers Collection (Springer, Berlin, 2005).

[4] D. P. Tieleman, S. J. Marrink, and H. J. C. Berendsen, Biochim. Biophys. Acta **1331**, 235 (1997).

[5] I. Vattulainen and M. Karttunen, in *Computational Nanotechnology*, edited by M. Rieth and W. Schommers (American Scientific, Stevenson Ranch, CA, 2005).

[6] M. M. Sperotto, S. May, and A. Baumgaertner, Chem. Phys. Lipids **141**, 2 (2006).

[7] H. L. Scott, Curr. Opin. Struct. Biol. **12**, 495 (2002).

[8] S. A. Pandit, E. Jakobsson, and H. L. Scott, Biophys. J. **87**, 3312 (2004).

[9] P. Niemelä, S. Ollila, M. T. Hyvönen, M. Karttunen, and I. Vattulainen, PLOS Comput. Biol. (in press).

[10] J. Aittoniemi, P. Niemelä, M. T. Hyvönen, M. Karttunen, and I. Vattulainen, Biophys. J. **92**, 1125 (2007).

[11] M. Bloom, E. Evans, and O. G. Mouritsen, Q. Rev. Biophys. **24**, 293 (1991).

[12] K. Simons and E. Ikonen, Nature (London) **387**, 569 (1997).

[13] K. Simons and W. L. C. Vaz, Annu. Rev. Biophys. Biomol. Struct. **33**, 269 (2004).

[14] E. Falck, M. Patra, M. Karttunen, M. T. Hyvönen, and I. Vattulainen, Biophys. J. **87**, 1076 (2004).

[15] S. Vainio, M. Jansen, M. Koivusalo, T. Rog, M. Karttunen, I. Vattulainen, and E. Ikonen, J. Biol. Chem. **281**, 348 (2005).

[16] J. Aittoniemi, T. Róg, P. Niemelä, M. Pasenkiewicz-Gierula, M. Karttunen, and I. Vattulainen, J. Phys. Chem. B **110**, 25562 (2006).

[17] O. G. Mouritsen, B. Dammann, H. C. Fogedby, J. H. Ipsen, C. Jeppesen, K. Jørgensen, J. Risbo, M. C. Sabra, M. M. Sperotto, and M. J. Zuckermann, Biophys. Chem. **55**, 55 (1995).

[18] *Novel Methods in Soft Matter Simulations*, edited by M. Karttunen, I. Vattulainen, and A. Lukkarinen (Springer-Verlag, Berlin, 2004).

[19] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, J. Phys.: Condens. Matter **16**, R481 (2004).

[20] T. Murtola, M. Kupiainen, E. Falck, and I. Vattulainen, J. Chem. Phys. **126**, 054707 (2007).

[21] M. Venturoli and B. Smit, PhysChemComm **10**, 1 (1999).

[22] S. J. Marrink, A. H. de Vries, and A. E. Mark, J. Phys. Chem. B **108**, 750 (2004).

[23] S. Izvekov and G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005).

[24] S. Izvekov and G. A. Voth, J. Chem. Theory Comput. **2**, 637 (2006).

[25] Q. Shi, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **110**, 1545 (2006).

[26] S. J. Marrink and A. E. Mark, J. Am. Chem. Soc. **125**, 11144 (2003).

[27] S. J. Marrink and A. E. Mark, J. Am. Chem. Soc. **125**, 15233 (2003).

[28] S. J. Marrink and A. E. Mark, Biophys. J. **87**, 3894 (2004).

[29] R. Faller and S. J. Marrink, Langmuir **20**, 7686 (2004).

[30] S. J. Marrink, J. Risselada, and A. E. Mark, Chem. Phys. Lipids **135**, 223 (2005).

[31] M. Nielsen, L. Miao, J. H. Ipsen, O. G. Mouritsen, and M. J. Zuckermann, Phys. Rev. E **54**, 6889 (1996).

[32] M. Nielsen, L. Miao, J. H. Ipsen, M. J. Zuckermann, and O. G. Mouritsen, Phys. Rev. E **59**, 5790 (1999).

[33] M. Nielsen, J. Thewalt, L. Miao, J. H. Ipsen, M. Bloom, M. J. Zuckermann, and O. G. Mouritsen, Europhys. Lett. **52**, 368 (2000).

[34] L. Miao, M. Nielsen, J. Thewalt, J. H. Ipsen, M. Bloom, M. J. Zuckermann, and O. G. Mouritsen, Biophys. J. **82**, 1429 (2002).

[35] J. M. Polson, I. Vattulainen, H. Zhu, and M. J. Zuckermann, Eur. Phys. J. E **5**, 485 (2001).

[36] G. A. Khelashvili, S. A. Pandit, and H. L. Scott, J. Chem. Phys. **123**, 034910 (2005).

[37] Q. Shi and G. A. Voth, Biophys. J. **89**, 2385 (2005).

[38] R. Elliott, I. Szleifer, and M. Schick, Phys. Rev. Lett. **96**, 098101 (2006).

[39] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainen, J. Chem. Phys. **121**, 9156 (2004).

[40] F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).

[41] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, J. Chem. Phys. **120**, 10896 (2004).

[42] S. Izvekov and G. A. Voth, J. Chem. Phys. **123**, 134105 (2005).

[43] D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).

[44] R. Faller, Polymer **45**, 3869 (2004).

[45] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).

[46] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, Soft Mater. **1**, 121 (2003).

[47] J. R. Silbermann, S. H. L. Klapp, M. Schoen, N. Chennamsetty, H. Bock, and K. E. Gubbins, J. Chem. Phys. **124**, 074105 (2006).

[48] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **55**, 5689 (1997).

[49] A. P. Lyubartsev and A. Laaksonen, J. Chem. Phys. **111**, 11207 (1999).

[50] A. P. Lyubartsev, Eur. Biophys. J. **35**, 53 (2005).

[51] J. Elezgaray and M. Laguerre, Comput. Phys. Commun. **175**, 264 (2006).

[52] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, J. Phys. Chem. B **105**, 4464 (2001).

[53] M. R. Vist and J. H. Davis, Biochemistry **29**, 451 (1990).

[54] J. P. Slotte, Biochim. Biophys. Acta **1235**, 419 (1995).

[55] H. M. McConnell and A. Radhakrishnan, Biochim. Biophys. Acta **1610**, 159 (2003).

[56] C. Ege, M. K. Ratajczak, J. Majewski, K. Kjaer, and K. Y. C. Lee, Biophys. J. **91**, L01 (2006).

[57] T. Murtola, T. Róg, E. Falck, M. Karttunen, and I. Vattulainen, Phys. Rev. Lett. **97**, 238102 (2006).

[58] R. L. Henderson, Phys. Lett. **49A**, 197 (1974).

[59] F. Jähnig, Biophys. J. **71**, 1348 (1996).

[60] S. E. Feller and R. W. Pastor, J. Chem. Phys. **111**, 1281 (1999).

[61] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).

[62] S. W. Chiu, E. Jakobsson, R. J. Mashl, and H. L. Scott, Biophys. J. **83**, 1842 (2002).

[63] C. Hofsäß, E. Lindahl, and O. Edholm, Biophys. J. **84**, 2192 (2003).

[64] D. Needham, T. J. McIntosh, and E. Evans, Biochemistry **27**, 4668 (1988).

[65] F. L. Román, J. A. White, and S. Velasco, J. Chem. Phys. **107**, 4635 (1997).

[66] F. L. Román, J. A. White, and S. Velasco, Europhys. Lett. **42**, 371 (1998).

[67] B. J. Thijsse, M. A. Hollanders, and J. Hendrikse, Comput. Phys. **12**, 393 (1998).

[68] T. Róg and M. Pasenkiewicz-Gierula, Biophys. J. **81**, 2190 (2001).

[69] P. Jedlovszky and M. Mezei, J. Phys. Chem. B **107**, 5311 (2003).

[70] H. I. K. Petrache, K. Tu, and J. F. Nagle, Biophys. J. **76**, 2479 (1999).

[71] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd. ed. (Academic, London, 1986).

[72] A. A. Louis, J. Phys.: Condens. Matter **13**, 9187 (2002).