# Coarse-grained model for phospholipid/cholesterol bilayer

Teemu Murtola and Emma Falck
*Laboratory of Physics and Helsinki Institute of Physics, Helsinki University of Technology,*
*P.O. Box 1100, FIN-02015 HUT, Finland*

Michael Patra and Mikko Karttunen
*Biophysics and Statistical Mechanics Group, Laboratory of Computational Engineering,*
*Helsinki University of Technology, P.O. Box 9203, FIN-02015 HUT, Finland*

Ilpo Vattulainen[a]
*Laboratory of Physics and Helsinki Institute of Physics, Helsinki University of Technology,*
*P.O. Box 1100, FIN-02015 HUT, Finland*

We construct a coarse-grained (CG) model for dipalmitoylphosphatidylcholine (DPPC)/cholesterol bilayers and apply it to large-scale simulation studies of lipid membranes. Our CG model is a two-dimensional representation of the membrane, where the individual lipid and sterol molecules are described by pointlike particles. The effective intermolecular interactions used in the model are systematically derived from detailed atomic-scale molecular dynamics simulations using the Inverse Monte Carlo technique, which guarantees that the radial distribution properties of the CG model are consistent with those given by the corresponding atomistic system. We find that the coarse-grained model for the DPPC/cholesterol bilayer is substantially more efficient than atomistic models, providing a speedup of approximately eight orders of magnitude. The results are in favor of formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the experimental phase diagram of the system. We also explore the limits of the coarse-grained model, and discuss the general validity and applicability of the present approach. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1803537]

## I. INTRODUCTION

Cell membranes are remarkably flexible and durable structures enclosing and protecting the contents of cells or organelles.[1–6] They are, however, by no means mere casings, but bustling hubs, where signaling, recognition, and transport take place. The fact that a huge variety of cellular processes are governed by membranes makes them a fascinating and biologically relevant example of soft and distinctly thin interfaces.

The complexity and biological relevance of membranes is largely due to the variety of proteins and lipids that are their main building blocks. It is intriguing that there are typically more than a hundred different lipid species in a given type of biological membrane, all assumed to have some particular purpose.[1,3] Instead of being static, membranes are highly dynamic and characterized by distinct phases and internal fluctuating structures, thus allowing proteins to function under nonequilibrium conditions.[2] Understanding the overall properties of membranes is therefore a major challenge involving studies over large scales in time and space: starting from the atomistic and molecular regimes where small-scale processes such as ion flow through channel proteins takes place all the way to mesoscopic and macroscopic regimes where large-scale processes such as phase separation and membrane fusion are important.

The present understanding of membrane systems is largely based on experimental studies, where techniques such as fluorescence spectroscopy, nuclear magnetic resonance, and various scattering methods have been used.[1–4,6] At the same time, experiments have been complemented by theoretical and computational modeling.[2,3,7–13] Thanks to the interplay between the two fields, a more detailed understanding of membranes and their biological relevance is emerging.

As for computational modeling of membrane systems in the *atomistic regime*, classical molecular dynamics (MD) is the method of choice.[8–13] It provides detailed information on the structure and dynamics of individual lipid molecules, as well as insight into processes such as the complexation and hydrogen bonding properties of different lipid species or the effect of enzymes on membranes. The main limitation of the atomistic approach is the computational load. With currently available computer power, the standard size for systems is 128 lipid molecules, corresponding to a linear system size of about 5–7 nm in the bilayer plane. The duration of such a simulation is typically limited to about 100 ns. A few more ambitious MD simulations on systems with a larger number of molecules have been reported,[14–16] but the sizes are still rather modest: the largest MD studies of lipid bilayers contain of the order of $10^3$ lipid molecules. The time scales reached in such simulations are currently only tens of nanoseconds.

The above limitations are problematic since many interesting phenomena in lipid membrane systems occur at much longer time and length scales. Examples of such phenomena

---

[a] Author to whom correspondence may be addressed. Electronic mail: ilp@fyslab.hut.fi

are domain formation, bilayer fusion, and cooperative motions associated with phase changes. Domain formation is a particularly interesting issue, since there is plenty of experimental evidence pointing to the formation of lateral domains in many-component bilayers.[2,3] The most topical issue are *lipid rafts*,[17–20] which are thought to be dynamic, ordered lateral domains comprised mainly of phosphatidylcholine, cholesterol, and sphingomyelin molecules. Rafts have been suggested to be involved in a wide range of cellular processes including membrane trafficking and sorting of proteins,[17–20] which emphasizes the need to understand large-scale properties of membrane domains. The dimensions of these dynamic domains are believed to range from tens to hundreds of nanometers, still beyond the limits of atomistic simulations. To reach the necessary length scales, i.e., hundreds of nanometers, we therefore need to resort to *coarse-grained models* that employ effective interaction potentials for simplified molecular descriptions.[2,21–23] To gain insight into large-scale properties of lipid membrane systems, the main objective is hence to develop and employ coarse-grained membrane models incorporating only the essential properties of the underlying system.

Previous studies in this field have been few, although recent progress is very promising. As for *semiatomistic models* of bilayers, a number of interesting approaches have been suggested.[24–33] The guiding principle is that small groups of atoms are represented as single interaction sites, thus reducing the computational complexity of the model. In the model by Marrink, de Vries, and Mark[24] there are Lennard-Jones, harmonic, and electrostatic interactions between the coarse-grained particles, and the interaction parameters have been tuned to match experimental quantities such as heats of vaporization or densities. The systems may be simulated using classical MD. The approach by Lipowsky and co-workers is somewhat more phenomenological but similar in nature.[25–27] Shelley *et al.*, in turn, have employed a large variety of different interactions between the coarse-grained particles,[31–33] and the interaction parameters have been adjusted to match results from both experiments and atomic simulations. Groot and Rabone have employed the dissipative particle dynamics (DPD) technique and chosen soft potentials between the coarse-grained particles.[28] The repulsive interaction parameters have been derived from compressibility and solubilities. The DPD studies by Groot and Rabone have been complemented by the simulations of Smit and co-workers,[29,30] and Ayton *et al.* have also employed the DPD technique. They have used material properties from atomistic simulations to parametrize mesoscale and macroscale models of lipid bilayers and unilamellar vesicles.[34–37]

Alternatively, one can design *phenomenological models* where the number of degrees of freedom is as small as possible. One of the best examples of this approach is the work by Mouritsen and co-workers[7,38–42] They have developed and used an off-lattice model where lipid and sterol molecules are described as hard-core particles with internal (spin-type) degrees of freedom. This approach has allowed them to design models whose phase diagrams are in qualitative agreement with experimental ones for phosphatidylcholine (PC)/cholesterol and PC/lanosterol systems.[7,38–41] Addi-

tionally, the models have been successful in describing lateral diffusion in PC/sterol bilayer mixtures.[42] The work by Mouritsen *et al.* demonstrates that purely phenomenological models can be very useful in accessing scales larger than those within reach of atomistic simulation techniques. On the other hand, due to their phenomenological nature, the scope of these models may be limited.

In general, there is no single method of constructing models for mesoscopic or macroscopic phenomena, and each case has to be considered separately. Hence, systematic and general approaches that simplify the construction of coarse-grained models and reduce the number of phenomenological and tunable parameters would be of great interest. A generally useful approach would be easily extendable or modifiable to describe different kinds of systems.

A promising candidate is the Inverse Monte Carlo technique (IMC).[43,44] It allows one to derive *all* effective interaction potentials systematically from atomic-level information such that the most relevant structural properties of the atomic-level system are reproduced by the coarse-grained model. This approach has been used in other soft matter systems. For example Lyubartsev *et al.*[43–45] have used IMC for constructing a coarse-grained model for sodium and chloride ions in water, and for describing the binding of different alkali ions to DNA.[46] The approach employed by Shelley *et al.* is also, in part, based on the central ideas of the IMC method.[31] Notably, IMC is not only systematic, but also highly adjustable, as the level of coarse-graining and the number of degrees of freedom can be tuned. It is thus possible to choose how large scales are to be studied as well as how much detail is to be included.

In the present study, we apply the IMC approach[43,44] to construct a coarse-grained (CG) model for a lipid bilayer containing dipalmitoylphosphatidylcholine (DPPC) and cholesterol. This system was chosen because DPPC is one of the most studied phospholipids, and cholesterol is the most important sterol molecule found in plasma membranes of eukaryotic cells. Further, the system has a rich and interesting phase behavior[47] characterized by three main phases (see Fig. 1). It has been suggested that at certain temperatures and cholesterol concentrations the system might display cholesterol-rich domains[48] or superlattice domains.[49]

We first performed extensive atomic-level MD simulations of the DPPC/cholesterol system at six different cholesterol concentrations.[50] The results of these simulations agree well with experiments and other simulations. Based on these atomistic considerations, we construct a coarse-grained model in which we describe each molecule by a single point-like particle moving in a two-dimensional plane. We use the IMC technique to derive effective interaction potentials for the coarse-grained particles. These interactions are constructed such that the CG model reproduces the radial distribution functions (RDFs) calculated from the atomic-level MD simulations. Because the RDFs can be used to characterize the phase behavior, the model should also qualitatively reproduce the phase behavior of the microscopic model.

Using the coarse-grained model, we study DPPC/cholesterol bilayers with cholesterol concentrations varying from 0% to 50%. We first validate the CG model by com-
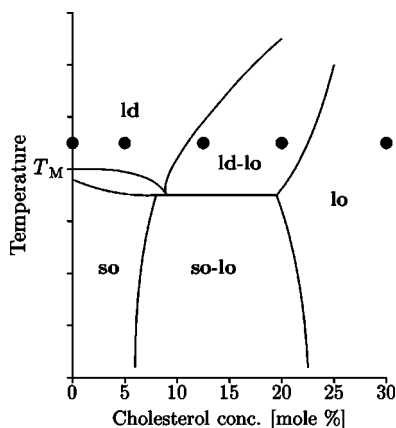
FIG. 1. Sketch of experimental phase diagram for the DPPC/cholesterol system (Ref. 47). At high temperatures and low cholesterol concentrations, there is a liquid-disordered (ld) phase, which is a fluid-phase characterized by lipid acyl chains with a high degree of conformational disorder. When the temperature is lowered, the system goes through the main phase transition at $T_M \approx 311$ K to a solid-ordered (so) phase. The so phase is essentially a solid phase in which acyl chains are conformationally ordered and the positions of the molecules are characterized by translational order in the bilayer plane. 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 found in the ld phase. 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. An additional MD simulation was performed at 50% cholesterol.

paring its behavior to that of the atomic-scale model. As the degree of coarse graining is very high, the model allows us to study the properties of the bilayer on length scales of the order of 100 nm along the plane of the membrane. The computational gain can be approximated to be around eight orders of magnitude compared to the atomistic MD case.

We find that the coarse-graining approach can provide plenty of insight into large-scale properties of many-component membrane systems. In this case it allows us to observe formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the experimental phase diagram of the system.[47] We also explore the limitations of the model, and discuss its general validity as well as possible future applications.

## II. MOLECULAR DYNAMICS SIMULATION DETAILS

The underlying MD simulations have been described in detail elsewhere,[50] and only a brief summary is given here. We simulated fully hydrated lipid bilayer systems consisting of 128 macromolecules, i.e., DPPCs and cholesterols, and 3655 water molecules. The simulations were performed at six cholesterol molar fractions: 0%, 4.7%, 12.5%, 20.3%, 29.7%, and 50%. These concentrations are indicated in the phase diagram of Fig. 1. The duration of each simulation was 100 ns and the linear sizes of the systems in the plane of the bilayer were between 5 and 7 nm.

The starting point for the simulations was a united atom model for a fully hydrated pure DPPC bilayer that has been validated previously.[50–52] The parameters for bonded and nonbonded interactions for DPPC molecules were taken

from a study of a pure DPPC bilayer,[54] and partial charges from the underlying model description.[53] The cholesterol force field was taken from an earlier study.[55]

As an initial configuration for the pure DPPC bilayer we used the final structure of run *E* discussed in Ref. 53. For systems containing cholesterol, the initial configurations were constructed by replacing randomly selected DPPC molecules with cholesterols. The same number of DPPC molecules was replaced in each of the two monolayers. To fill the small voids left by replacing DPPC molecules by somewhat smaller cholesterol molecules, the system was equilibrated in several stages.[50]

The MD simulations were performed at a temperature $T = 323$ K using the GROMACS molecular simulation package.[56] The main phase transition temperature for a pure DPPC bilayer is $T_M \approx 314$ K,[57] indicating that the MD simulations have been conducted above $T_M$. The time step for the simulations was chosen to be 2.0 fs. Long-range electrostatic interactions were handled using the particle-mesh Ewald method.[58] After the initial equilibration we performed 100 ns of MD in the $NpT$ ensemble with a Berendsen thermostat and barostat[59] for each cholesterol concentration. For all systems up to and including the cholesterol concentration of 29.7%, a simulation lasting 100 ns guarantees good sampling of the phase space. The results for 50% cholesterol should be regarded with some caution as the diffusion of the DPPC and cholesterol molecules is already quite slow.[50]

## III. COARSE-GRAINED MODEL

Using the MD simulations as a basis, we have constructed a coarse-grained model for a DPPC/cholesterol bilayer. Since the main goal of the present project is to study the *large-scale structural properties of the bilayer*, the degree of coarse graining must be high. A way to achieve this goal is to describe each DPPC and cholesterol molecule by its center-of-mass (CM) position. The macromolecules are taken to be single pointlike particles that move and interact in two dimensions with continuous coordinates. At the same time, the solvent degrees of freedom have been integrated out altogether, i.e., the model contains no explicit water.

Let us briefly list the assumptions we have made in constructing the CG model. To start with, we consider a lipid bilayer as a purely two-dimensional sheet comprised of two weakly interacting leaflets. For this reason, we focus on one monolayer only. This assumption is well justified since interdigitation in DPPC/cholesterol bilayers is minor.[50] Consequently, the friction between the two leaflets is weak and they can be regarded as largely independent from each other. Furthermore, we neglect the out-of-plane fluctuations of the bilayer and assume it to be strictly planar. Such fluctuations decrease when the cholesterol concentration is increased,[15] making this a reasonable assumption especially at higher cholesterol concentrations.

Due to its coarse-grained nature, our model is dissipative. This stems mainly from the fact that the water molecules are not included in the CG model. Further, in constructing the model the conformational degrees of freedom of the macromolecules have been integrated out. Yet another reason is that we consider one of the leaflets rather than the

whole membrane. If we were to study dynamical phenomena, the dynamics should be chosen such that there are both stochastic and dissipative force components describing those degrees of freedom that have been excluded from the CG model. As we will focus on structural quantities of the membrane system, we will not have to worry about the choice of dynamics. Instead, we use the Metropolis Monte Carlo (MC) technique.[60] The question of implementing realistic dynamics to the model is considered in more detail at the end of Sect. VI.

As for the interactions between the pointlike DPPC and cholesterol particles, we assume they can be adequately described using pairwise, radially symmetric effective potentials. The effective interactions are computed as follows. From the atomistic MD simulations we calculate radial distribution functions for the CM positions of the molecules. To link our coarse-grained model to the atomic-level system, we require that the coarse-grained model accurately reproduces these RDFs. This is accomplished by constructing the effective interaction potentials using the Inverse Monte Carlo method.[43,44] In principle, also other canonical averages than the RDFs could be used as an input. However, the RDFs calculated from the atomic-scale MD simulations are easy to compute, and more importantly, give a detailed structural description of the system in the plane of the bilayer at short length scales. Because the RDFs can be used for characterizing the phase behavior of the system, the coarse-grained model should at least qualitatively reproduce the phase behavior of the original atomic-scale system.

In addition to the above, we fix the area per molecule in the CG model to be the same as the average area per molecule calculated from the atomistic MD simulations. The MC simulations will therefore be conducted in the canonical ensemble.

## IV. MODEL CONSTRUCTION AND VALIDATION

### A. Radial distribution functions and areas per molecule from atomistic MD simulations

For the construction of the coarse-grained model, we need to obtain the radial distribution functions for the center-of-mass positions of the molecules. Figure 2 shows the RDFs calculated from the atomic-scale MD simulations. Before calculating the RDFs, the CM positions have been projected to the plane of the bilayer. The two monolayers have been considered separately and the resulting RDFs are averages of the two. The first 20 ns of the MD data have been discarded to allow the system to equilibrate fully.[50] After 20 ns the area per molecule has converged for all cholesterol concentrations,[50] and the radial distribution functions show no systematic changes. The radial distribution functions were calculated up to 2.5 nm for concentrations lower than 50%. At the highest concentration of 50%, the linear size of the system in the bilayer plane is occasionally below 5 nm, and therefore in this case the RDFs were cut off at 2.4 nm. 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. To mini-
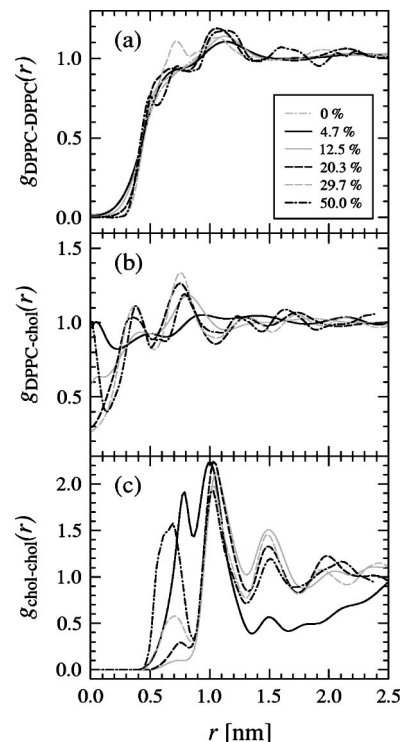


FIG. 2. Radial distribution functions calculated from MD simulations for (a) DPPC-DPPC, (b) DPPC-cholesterol and (c) cholesterol-cholesterol pairs. The RDFs are calculated from the center-of-mass positions of the molecules, which have been projected to the plane of the bilayer.

mize the effect of random errors on the Inverse Monte Carlo procedure, we used a spline-fitting procedure designed for noisy data[61] to smooth the RDFs.

For all concentrations the RDFs indicate liquidlike behavior. At short length scales there are broad peaks and at larger $r$ the functions approach unity. In other words, although there is short-range order, there are no signs of long-range order characteristic to solidlike phases. This confirms that we are probing the region of the phase diagram where the system is in the ld, lo, or coexisting ld and lo phases (see the points marked in Fig. 1). As the cholesterol concentration is increased, the radial distribution functions change more or less systematically. The peaks in the DPPC-DPPC distribution become sharper, manifesting an increase in the lateral short-range ordering. Further, more peaks appear at larger $r$, which means that the range of the ordering increases slightly. Similar effects are observed in the case of the cholesterol-cholesterol RDFs, although the 4.7% concentration deviates somewhat from the general trend. For the DPPC-cholesterol distribution the changes are not quite as systematic. Nonetheless, when the cholesterol concentration grows, the range of ordering seems to increase slightly and the peaks generally become sharper.

A notable feature of the RDFs, especially for the DPPC-cholesterol pairs, is the fact that some RDFs do not approach zero at the origin. This is because the RDFs have been calculated for the CM positions of the molecules, which have been projected to the plane of the bilayer. It is not too difficult to imagine a situation where the projected CM positions of a rigid, short cholesterol molecule and a DPPC molecule
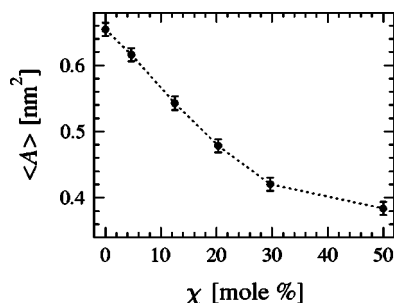
FIG. 3. Average area per molecule as a function of cholesterol concentration, calculated from MD simulations (Ref. 50).

with long, flexible tails are essentially on top of each other.

In addition to the RDFs, we need the average area per molecule from the MD simulations to fix the area per molecule in the coarse-grained model. The average areas per molecule at different cholesterol concentrations have been published previously,[50] and are shown in Fig. 3 for reference. The area per molecule for a given configuration has been computed by dividing the size of the simulation box in the bilayer plane by the number of molecules in one monolayer. The area per molecule decreases monotonically with an increasing cholesterol concentration, in agreement with other simulations[15,62] and related experiments.[63]

## B. Constructing effective potentials

Based on the RDFs, we have constructed effective interaction potentials for our coarse-grained particles using the IMC method. For each cholesterol concentration, the RDFs calculated from the MD simulations were given as an input to the IMC procedure to obtain the effective interactions.

Due to finite size effects, in some cases the RDFs calculated from the MD simulations deviate from unity at the cutoff. This results in discontinuities in the effective potentials at the cutoff. To handle these, we applied a simple shifting scheme from 2.0 nm to the cutoff distance to adjust the potentials such that they approach zero smoothly at the cutoff. The approach used is essentially similar to that presented in Ref. 64. The main difference is that in the present case shifting is applied to the potential rather than to the force.

The IMC does not give reasonable estimates for the effective potentials at short interparticle distances where the RDFs vanish. In these regions, we replaced the potential given by the IMC method by polynomials such that the potential and its first derivative are continuous at the edge of the region. Finally, the effective potentials were smoothed using the same spline-fitting procedure[61] as was used for the RDFs to reduce statistical noise. We have verified that the potentials are not sensitive to the details of the process of obtaining them. Thus the above changes can be made without seriously altering the resulting RDFs.

Figure 4 shows the computed effective interaction potentials. Due to the high level of coarse graining, the potentials are soft. The DPPC-DPPC and DPPC-cholesterol potentials become systematically more repulsive with an increasing cholesterol concentration, and the very small attractive component present at low concentrations is lost at higher concen-
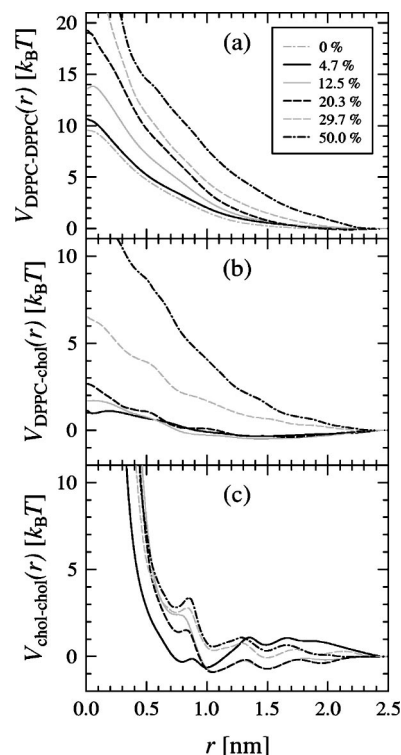


FIG. 4. Effective potentials for different pairs of coarse-grained particles: (a) DPPC-DPPC, (b) DPPC-cholesterol, and (c) cholesterol-cholesterol.

trations. For the cholesterol-cholesterol potentials, the behavior is more complex. For 29.7% and 50.0% concentrations, the interaction is mostly repulsive, but for 12.5% and 20.3% there is weak attraction for $r \gtrsim 0.9$ nm up to the cutoff. For 4.7%, the interaction is again repulsive for $r \gtrsim 1.1$ nm, but now there is a weak attraction for small separations. In addition, the cholesterol-cholesterol potentials have multiple minima, whereas the other potentials are much simpler.

## C. Validation

As any model, the coarse-grained model should be validated. This can be done by comparing the results it generates to the results from the MD simulations. By construction, the CG model should reproduce the short-range structural properties of the atomic-scale model.

Figure 5 shows a comparison between RDFs calculated from the MD simulations and those obtained from the CG model using canonical MC. The CG simulations contained the same number of particles as the MD simulations. We show only a few, selected cholesterol concentrations, but for other concentrations the results are similar: the agreement between the results from MD and CG is excellent at all concentrations, as it should be. The minor differences near the cutoff arise from the use of a shifting function for the potentials. Without the shift, the lines coincide up to the cutoff, but in some cases there is a small discontinuity in the RDFs at the cutoff.

We have also calculated the static structure factors computed over all pairs of particles on a two-dimensional grid. To this end, consider a set of particles $i = 1, \ldots, N$ whose positions are $\vec{r}_i$. Then the static structure factor defined as
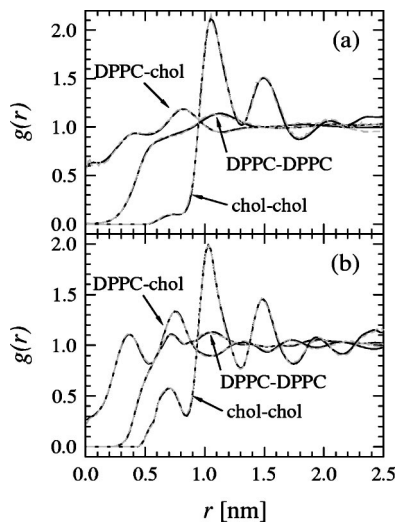
FIG. 5. Comparison between radial distribution functions calculated from MD simulations (solid black lines) and from CG model (dashed gray lines). Two different cholesterol concentrations are shown: (a) 12.5% and (b) 29.7%. The results for the other concentrations are similar.

$$S(\vec{k}) = \frac{1}{N^2} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \exp[-i\vec{k} \cdot (\vec{r}_j - \vec{r}_i)] \right\rangle \quad (1)$$

is given in terms of the reciprocal vector $\vec{k}$. The $S(\vec{k})$ have been calculated for a system whose linear size varies between 120–160 nm. This is 24 times the size of the original system studied by MD. In all cases the structure factors were found to be radially symmetric. The circularly averaged structure factors $S(k)$ for different cholesterol concentrations are shown in Fig. 6. These curves are discussed in detail in Sect. V. At this point it is sufficient to note that these calculations confirm that the system is isotropic at all concentrations and that there is no long-range solidlike order. The system is in a fluidlike state as it should be.

## V. BEHAVIOR AT LARGE LENGTH SCALES

When the system size is increased from that of the MD simulations, several new phenomena are observed. The simulations described below were mostly performed on systems containing 36 864 particles, corresponding to linear sizes 24 times those of the original MD simulations. Hence,
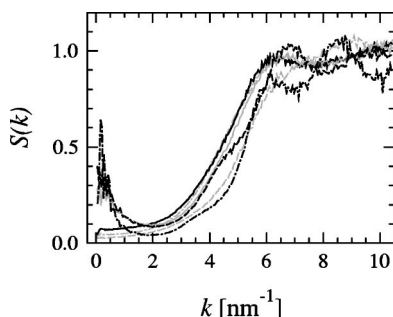


FIG. 6. Total circularly averaged static structure factors computed from CG simulations at different cholesterol concentrations. The curves are labeled as in Fig. 2.
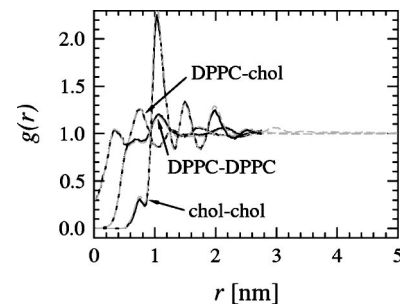


FIG. 7. Changes in radial distribution functions when the size of the simulation box is increased at 20.3% cholesterol. Black solid lines correspond to a small system with 64 particles and dashed gray lines to a system with a linear size 24 times larger.

the linear sizes of the systems were 120–160 nm, depending on the concentration. A typical simulation required 50–100 CPU hours on a desktop computer.

When the system size is increased, there are some minor changes in the radial distribution functions. These are illustrated in Fig. 7 for the 20.3% cholesterol concentration. For other concentrations the results are similar. These changes are most probably a finite-size effect caused by the small sizes of the original atomic-scale systems. It is very likely that if the MD simulations could be performed on larger systems, similar changes in the RDFs should take place. The figure also shows that the RDFs rapidly approach unity at large distances.

To study possible large-scale organization, we have looked at the static structure factors calculated at different cholesterol concentrations. They are shown in Fig. 8 together with snapshots of the system at each concentration. The structure factors have been calculated over all pairs of molecules, and also separately for DPPC-DPPC, cholesterol-cholesterol, and DPPC-cholesterol pairs. To ensure that the static structure factors do not depend on the initial configuration, we have rerun the calculations using several different initial states. We find that different initial configurations lead to results that are consistent with each other.

At 12.5% and 20.3% cholesterol, the snapshots in Fig. 8 suggest that there are domains where the local concentration of cholesterol is higher than in other regions. At these concentrations the cholesterol-cholesterol structure factor shows a rather wide peak at small $k$, supporting our interpretation of the presence of cholesterol-rich and cholesterol-poor domains. The maximum of the peak corresponds to length scales of the order of 20 nm or more. A more precise analysis is difficult, since even for the largest systems we have studied, with linear sizes of $\approx 280$ nm, the peak is rather broad, and the small $k$ side of the peak is not fully clear due to fluctuations at large length scales.

We have, however, studied how the peak behaves if the system size is varied by calculating the static structure factors for four systems with linear sizes 6, 12, 24, and 48 times that of the original MD simulations. In this finite size scaling analysis, we found no clear change in either the position or shape of the peak.

The formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations is in
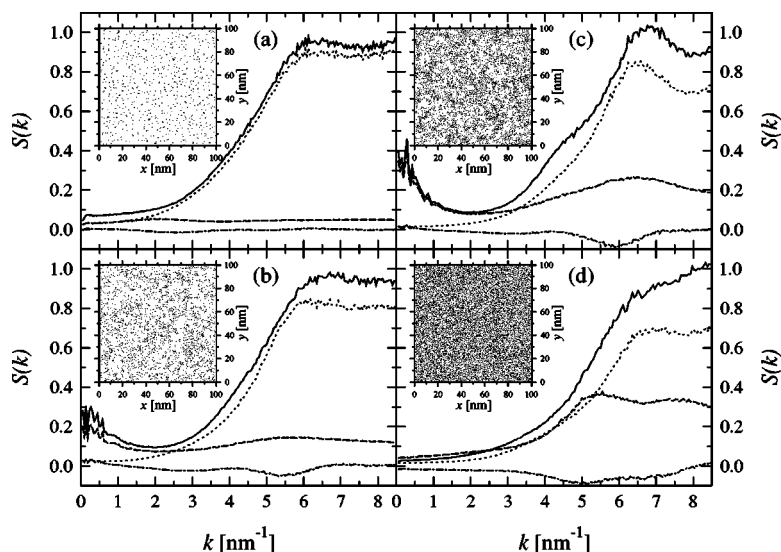
FIG. 8. Static structure factors calculated for different cholesterol concentrations: (a) 4.7%, (b) 12.5%, (c) 20.3%, and (d) 29.7%. In addition to the total structure factor computed over all pairs of particles (solid line), structure factors calculated over DPPC-DPPC (dotted), cholesterol-cholesterol (dashed) and DPPC-cholesterol (dot-dashed) pairs are shown. Also a snapshot of the system is shown for each concentration. In the snapshots, each cholesterol molecule is represented as a single dot, while DPPC molecules are not shown.

agreement with the phase diagram, see Fig. 1. According to the phase diagram we should, at $T = 323$ K, expect both the MD and the CG model to display coexisting ld and lo phases. In the present case, we cannot directly distinguish between the ld and lo phases, since we have not included the ordering of the DPPC tails in the model. However, we might argue as follows to establish that the cholesterol-rich phases should be lo, while the cholesterol-poor must be ld. The ordering effect of cholesterol on the phospholipid tails has been clearly demonstrated: the higher the cholesterol concentration, the more ordered the tails.[15,50] It is thus plausible that the tails should be more ordered in the cholesterol-rich regions.

To study domain formation in more detail, we have computed probability distribution functions for finding square domains with a linear size $\ell$. Several different system sizes $\ell$ were considered. If there were domains of clearly different compositions, the distribution should display two peaks. These studies do not provide direct support for formation of domains (data not shown). At each concentration, and for each $\ell$, the distribution is very close to a Gaussian. However, with 12.5% and 20.3% cholesterol the variance of the distribution is clearly larger than for 4.7% or 29.7% cholesterol. Thus, it is possible that the observed domains are caused by large-scale fluctuations in the cholesterol concentration. Furthermore, at 12.5% the distribution is not quite symmetric for $\ell \lesssim 20$ nm. This may be due to two peaks that are located very close to each other.

We have also investigated the local concentrations of different types of molecules in the vicinity of a given type of molecule as in a study by de Vries, Mark, and Marrink.[65] More specifically, we have studied the respective numbers of different types of molecules among the $n$ nearest neighbors of DPPC and cholesterol molecules, for different values of $n$. The average number of molecules of type $A$ among the $n$ nearest neighbors of a molecule of type $B$ was compared to the value in a purely random configuration. In all cases the average number of cholesterol molecules in the vicinity of another cholesterol molecule was somewhat smaller than it should be in a random configuration. For $n = 6$, there were no

significant differences between different cholesterol concentrations. For larger $n$, say $n = 15$ or $n = 30$, the difference between the value from an actual simulation and the value from a random configuration was clearly smaller at the concentrations where we observe cholesterol-rich and cholesterol-poor domains than at other concentrations. This observation may be interpreted as support to the existence of domains. The reason why the difference is not visible in the case of $n = 6$ is probably that such a small neighborhood represents the area from which other cholesterol molecules are largely excluded by the repulsive interactions between the cholesterol molecules.

Based mainly on fluorescence measurements, it has been proposed that at certain cholesterol concentrations the cholesterol molecules could adopt a regular arrangement in parts of the bilayer.[66] We have studied whether or not our coarse-grained model shows such organization for a few of the proposed ''magical'' concentrations. The nearest-neighbor analysis described above suggests that finding a cholesterol molecule in the immediate vicinity of another cholesterol molecule is lower compared to a random configuration, whereas the probability of finding a cholesterol molecule next to a DPPC molecule is higher. Such a situation would occur if there were superlattices in the system. Nevertheless, we have seen no evidence for any regular latticelike ordering of cholesterols. Despite this conclusion, we are looking forward to further studies using other models. In our model the DPPC particles are radially symmetric, while in reality the two hydrocarbon tails of the molecules may be important for the occurrence of superlattices, if they do exist.

In the case of 50% cholesterol the situation is more complex. Visual inspection of snapshots clearly indicates that the DPPC and cholesterol molecules phase separate. All other studies we have performed support this view. There are at least two alternative explanations for this phenomenon. First, it is possible that the atomistic MD simulations do not, at this high cholesterol concentration characterized by a very small lateral diffusion coefficient, adequately sample the phase space. If this is the case, it may lead to errors in the RDFs extracted from the atomic-level simulations. As a conse-

quence, the effective potentials given by the IMC method would not describe the true behavior. On the other hand, there is experimental evidence for the formation of crystalline cholesterol domains at very high cholesterol concentrations.[67] Also this could be the reason for the observed phase separation. Based on the current simulations, it is difficult to say which of these, if indeed any, is the case.

## VI. DISCUSSION

The coarse-graining approach described here has several advantages. It is a systematic method to generate a mesoscale model that is linked to atomic-scale information. It is adjustable, as it allows the user to control the level of coarse-graining and the number of degrees of freedom to be included in the model. The approach is general in the sense that it can be applied to a wide range of systems. The only prerequisite is that we should somehow acquire radial distribution functions for pairs of the coarse-grained particles. These are available from atomic-scale simulations, but in certain cases also experiments could be used to derive the RDFs.

The most important advantage, however, is speed. Monte Carlo simulations using the coarse-grained model are several orders of magnitude faster than MD simulations of the original atomic-scale model. The speedup can be estimated by considering the decay time of the fluctuations in the area of a single molecule, where the area is calculated, through Voronoi analysis.[50,52] For the atomic-scale model, this decay time is $\approx 0.7$ ns,[52] whereas in the case of the CG model similar decay is observed after one MC step. The CPU time required for generating a 0.7 ns trajectory for the atomic-scale system using MD is around 70 h, whereas for a CG model of the same linear size as the atomic-scale system, $\approx 25\,000$ Monte Carlo steps can be taken during a minute. Thus the speedup can be estimated to be eight orders of magnitude.

Despite all advantages, there are limitations. An important point is that any problems in the atomic-scale molecular dynamics simulations are transferred to the coarse-grained model. In particular, if there is any artificial ordering in the atomic-scale simulations, either due to a small system size, poor sampling of the phase space, or sloppy treatment of electrostatic interactions,[51,52] the radial distribution functions will be erroneous. As a consequence, the effective interaction potentials will be affected.

Another limitation may be the concentration dependence of the effective potentials, see Fig. 4. It is not trivial in what concentration ranges the effective potentials are valid. The worst case scenario is that when the cholesterol concentration is altered ever so slightly, the interaction potentials must be rederived starting from time-consuming atomic-scale molecular dynamics simulations. We have investigated whether the effective potentials can be used at concentrations other than those at which they were determined. The results suggest that they may be used at nearby concentrations, but if phase boundaries are crossed, problems will arise. For instance, when using the effective potentials determined for the system with 29.7% cholesterol, a system in the lo phase, to simulate a system at 20.3% cholesterol, which should be in

the coexistence region, no long-range structure appears. When, on the other hand, potentials determined at 20.3% cholesterol are used with 29.7% cholesterol, the long-range structure is still present at the higher concentration, although the peak is more shallow. The situation is similar when we compare the 4.7% and 12.5% concentrations. When the potentials determined at 12.5% are used for 20.3% cholesterol or vice versa, i.e., both concentrations are in the coexistence region, we still observe domains. The detailed form of the static structure factor will, however, be altered. We may therefore conclude that the effective potentials cannot be used for mapping the precise phase boundaries of a given system. Similar conclusions can be drawn from a study of the temperature dependence of the effective interaction potentials.

An additional problem is the implementation of dynamics. In this study, we considered only structural quantities, and needed not to worry about the choice of dynamics. To study lateral diffusion it is necessary to incorporate realistic dynamics into the system. It is well known that this can be very difficult in the case of coarse-grained models. In addition to the MC studies reported here, we attempted to study lateral diffusion and include dynamics into our model. In choosing the dynamics, the following criteria should be met. First, as explained in Sect. III, the dissipative nature of the system should be taken into account. In addition, the simulations should be run in the canonical ensemble. We tested several thermostats that comply to the above requirements: standard and generalized Brownian dynamics,[68] and an approach similar to Andersen's thermostat for temperature coupling.[68] None of these could be tuned to give realistic dynamics for the system with rattling-in-the-cage movement and separate jump events. Both the standard Brownian dynamics approach and the Andersen scheme require very large friction or coupling parameters to give diffusion coefficients of the same order of magnitude as those obtained from the molecular dynamics simulations, or alternatively, to match the decay times of velocity autocorrelation functions. Such high values of the parameters completely determine the dynamics at short time scales, and the interactions between the particles only give rise to small corrections at long time scales. The generalized Brownian dynamics can be used to force the short time dynamics to match that of molecular dynamics simulations, but this does not mend the problem of unrealistic dynamics at longer time scales. We thus found no simple way of implementing realistic dynamics into the coarse-grained model. Reducing the degree of coarse graining and introducing more detail into the model, e.g., by including the tails of DPPC and cholesterol molecules, might help to solve the problem. The presence of tails would greatly increase the friction between the molecules, and possibly allow for entanglements. Both effects would help in slowing down the unrealistically fast dynamics.

The results motivate further studies of the coarse-graining approach. There are several possible directions in which the model could be developed. Several of these could be realized without additional MD simulations. A possible line of development is the inclusion of the conformational degrees of freedom of the DPPC molecules in the model.

This could be done in the spirit of the model by Nielsen *et al.*,[39] i.e., by giving each DPPC molecule two possible states: an ordered and a disordered state. We would have three kinds of particles and a total of six pairwise potentials to determine. Another possible modification would be to include the two tails of the lipid molecules as separate particles and possibly model the head group as a third particle. Results from simulations of such models could be compared to the present study to assess the possible benefits of such modifications, and to gain further insight into the coarse-graining process. Additionally, to better understand the underlying reasons for domain formation, it is natural to ask what the relative roles of entropic and energetic contributions in this process are. Since the close-packed areas of the CG particles are not well defined, a study of this broad and nontrivial issue is beyond the scope of this work and will be discussed elsewhere.

The coarse-graining approach presented here could also be applied to other lipid bilayer systems to compare their behavior to the DPPC/cholesterol bilayer. A particularly interesting system is the sphingomyelin (SM)/cholesterol bilayer, and ultimately the PC/SM/cholesterol ternary mixture. The study of these systems at large length scales would be particularly interesting, because experimental results suggest the existence of specific interactions between SM and cholesterol molecules.[20] These interactions are believed to forward the formation of domains and lipid rafts at high concentrations of SM and cholesterol. This line of development is limited by the computational cost of the underlying MD simulations. Currently this confines us to study quite simple bilayer systems. With increasing computer power and improvements in existing simulation methods, studies of ternary mixtures of lipids, as well as studies of systems containing membrane proteins, are becoming feasible.

## VII. SUMMARY AND CONCLUSIONS

We have used a systematic approach for constructing coarse-grained models for DPPC/cholesterol bilayers. The central ingredient is the application of the Inverse Monte Carlo method, which can be used for finding effective interactions such that the coarse-grained model reproduces given radial distribution functions. The approach allows easy tuning of the level of coarse graining, and it can be applied to a wide range of systems.

The radial distribution functions given as input to the IMC method have been extracted from detailed atomic-level molecular dynamics simulations. The effective interactions found using the IMC method can then be used to simulate the system on much longer length scales. We have found that the coarse-grained model thus constructed is in favor of the formation of cholesterol-rich and cholesterol-poor domains at intermediate cholesterol concentrations, in agreement with the phase diagram of the system. We have also explored the limitations of the constructed coarse-grained model.

As for further studies, it would be interesting to see how modifications such as the inclusion of the conformational degrees of freedom of the DPPC tails would influence the behavior of the model. Similar models could also be constructed for other many-component lipid bilayer systems,

and their behavior compared to the present study. This could yield valuable information on both the systems under study and the suitability of the IMC method for coarse graining in general.

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

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

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

[4] *Biological Membranes: A Molecular Perspective from Computation and Experiment*, edited by J. K. M. Merz and B. Roux (Birkhäuser, Boston, 1996).

[5] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell*, 3rd. ed. (Garland Publishing, New York, 1994).

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

[7] M. J. Zuckermann, M. Bloom, J. H. Ipsen *et al.*, Methods Enzymol. **383**, 198 (2004).

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

[9] S. E. Feller, Curr. Opin. Colloid Interface Sci. **5**, 217 (2000).

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

[11] L. Saiz, S. Bandyopadhyay, and M. L. Klein, Biosci Rep. **22**, 151 (2002).

[12] L. Saiz and M. L. Klein, Acc. Chem. Res. **35**, 482 (2002).

[13] I. Vattulainen and M. Karttunen, in *Computational Nanotechnology*, edited by M. Rieth and W. Schommers (Americal Scientific Press, in press).

[14] S. W. Chiu, S. Vasudevan, E. Jakobsson, R. J. Mashl, and H. L. Scott, Biophys. J. **85**, 3624 (2003).

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

[16] A. H. de Vries, A. E. Mark, and S. J. Marrink, J. Am. Chem. Soc. **126**, 4488 (2004).

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

[18] S. Mayor and M. Rao, Traffic Q. **5**, 231 (2004).

[19] M. Edidin, Annu. Rev. Biophys. Biomol. Struct. **32**, 257 (2003).

[20] R. F. M. de Almeida, A. Fedorov, and M. Prieto, Biophys. J. **85**, 2406 (2003).

[21] O. G. Mouritsen, B. Dammann, H. C. Fogedby *et al.*, Biophys. Chem. **55**, 55 (1995).

[22] *Bridging Time Scales: Molecular Simulations for the Next Decade*, edited by P. Nielaba, M. Mareschal, and G. Ciccotti (Springer, Berlin, 2002).

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

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

[25] R. Goetz and R. Lipowsky, J. Chem. Phys. **108**, 7397 (1998).

[26] R. Goetz, G. Gompper, and R. Lipowsky, Phys. Rev. Lett. **82**, 221 (1999).

[27] A. Imparato, J. C. Shillcock, and R. Lipowsky, Eur. Phys. J. E **11**, 21 (2003).

[28] R. D. Groot and K. L. Rabone, Biophys. J. **81**, 725 (2001).

[29] M. Kranenburg, M. Venturoli, and B. Smit, Phys. Rev. E **67**, 06090(R) (2003).

[30] M. Kranenburg, M. Venturoli, and B. Smit, J. Phys. Chem. B **107**, 11491 (2003).

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

[32] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, P. B. Moore, and M. L. Klein, J. Phys. Chem. B **105**, 9785 (2001).

[33] S. O. Nielsen, C. F. Lopez, P. B. Moore, J. C. Shelley, and M. L. Klein, J. Phys. Chem. B **107**, 13911 (2003).

[34] G. Ayton, S. G. Bardenhagen, P. McMurtry, D. Sulsky, and G. A. Voth, J. Chem. Phys. **114**, 6913 (2001).

[35] G. Ayton, A. M. Smondyrev, S. G. Bardenhagen, P. McMurtry, and G. A. Voth, Biophys. J. **82**, 1226 (2002).

[36] G. Ayton and G. A. Voth, Biophys. J. **83**, 3357 (2002).

[37] G. Ayton, A. M. Smondyrev, S. G. Bardenhagen, P. McMurtry, and G. A. Voth, Biophys. J. **83**, 1026 (2002).

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

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

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

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

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

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

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

[45] A. Lyubartsev and A. Laaksonen, J. Phys. Chem. **100**, 16410 (1996).

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

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

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

[49] B. Cannon, G. Heath, J. Hyang, P. Somerharju, J. A. Virtanen, and K. H. Cheng, Biophys. J. **84**, 3777 (2003).

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

[51] M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, P. Lindqvist, and I. Vattulainen, Biophys. J. **84**, 3636 (2003).

[52] M. Patra, M. Karttunen, M. T. Hyvönen, E. Falck, and I. Vattulainen, J. Phys. Chem. B **108**, 4485 (2004).

[53] D. P. Tieleman and H. J. C. Berendsen, J. Chem. Phys. **105**, 4871 (1996).

[54] O. Berger, O. Edholm, and F. Jahnig, Biophys. J. **72**, 2002 (1997).

[55] M. Höltje, T. Förster, B. Brandt, T. Engels, W. von Rybinski, and H.-D. Höltje, Biochim. Biophys. Acta **1511**, 156 (2001).

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

[57] U. Essman, L. Perera, M. L. Berkowitz, H. L. T. Darden, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[58] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[59] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, San Diego, 2002).

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

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

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

[63] GROMACS manual; see www.gromacs.org (2004).

[64] A. H. de Vries, A. E. Mark, and S. J. Marrink, J. Phys. Chem. B **108**, 2454 (2004).

[65] P. Somerharju, J. A. Virtanen, and K. H. Cheng, Biochim. Biophys. Acta **1440**, 32 (1999).

[66] D. Bach and E. Wachtel, Biochim. Biophys. Acta **1610**, 187 (2003).

[67] D. L. Ermak and H. Buckholtz, J. Comput. Phys. **35**, 169 (1980).

[68] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).