Teemu Murtola, Mikko Karttunen, and Ilpo Vattulainen. 2009. Systematic coarse-graining from structure using internal states: application to phospholipid / cholesterol bilayer. Helsinki Institute of Physics preprint HIP–2009–08/TH.

Systematic coarse-graining from structure using internal states: application to phospholipid / cholesterol bilayer

Teemu Murtola,[1] Mikko Karttunen,[2] and Ilpo Vattulainen[1,3,4]

[1] *Department of Applied Physics and Helsinki Institute of Physics, Helsinki University of Technology, P. O. Box 1100, FI–02015 Espoo, Finland*
[2] *Department of Applied Mathematics, The University of Western Ontario, 1151 Richmond St. N., London, Ontario, Canada N6A 5B7*
[3] *Department of Physics, Tampere University of Technology, P. O. Box 692, FI–33101 Tampere, Finland*
[4] *Memphys–Center for Biomembrane Physics, Physics Department, University of Southern Denmark, Campusvej 55, DK–5230 Odense M, Denmark*

(Dated: April 14, 2009)

We present a two-dimensional coarse-grained (CG) model for a lipid membrane composed of phospholipids and cholesterol. The effective CG interactions are determined using radial distribution functions (RDFs) from atom-scale molecular dynamics simulations using the inverse Monte Carlo (IMC) technique, in a spirit of our earlier work [T. Murtola *et al.*, J. Chem. Phys. **121**, 9156 (2004); **126**, 075101 (2007)]. Here, the original model is improved by including an internal discrete degree of freedom for the phospholipid tails to describe chain ordering. We also discuss the problem of RDF inversion in the presence of internal states in general, and present a modified IMC method for their inclusion. The new model agrees with the original models on large-scale structural features such as density fluctuations in pure DPPC and cholesterol domain formation at intermediate concentrations, and also indicates that ordered and disordered domains form at all cholesterol concentrations, even if the global density remains uniform. The inclusion of ordering also improves transferability of the interactions between different concentrations, but does not eliminate transferability problems completely. We also present a general discussion of problems related to RDF inversion.

I. INTRODUCTION

With advances in computer power, atomistic simulations have become a standard tool in studies of complex biomolecular systems such as lipid bilayers. However, even with present-day computers and software, atom-scale simulations can only access length scales up to perhaps 20 nm and time scales of the order of microseconds or less.[1,2] Hence, simplified *coarse-grained* (CG) models are needed for studies of many phenomena occurring at larger scales,[3] examples being phase separation and other structural rearrangements.

A large amount of effort has been devoted to accurate parameterization of atomistic potential energy functions,[4–9] but following similar routes for CG models is less straightforward. Because of the lower level of detail, quantitative reproduction of (some) properties for some specific system can typically only be achieved by a parameterization for that specific system. In such parameterization efforts, it is often advantageous to use information from atom-scale models in a *systematic* fashion. Other approaches are also possible: for example, the popular MARTINI model[10–12] has been parameterized similarly to atomistic potentials, using densities and partition coefficients between different substances. However, such an approach becomes more and more difficult as the level of detail decreases.

Systematic CG approaches calculate a set of target quantities from the atom-scale simulation, and then attempt to parameterize the CG model such that these quantities are reproduced as accurately as possible. Research on these methods has focused on two different approaches, using either structural information[13–15] or forces.[16–19] In the first, the radial distribution functions (RDFs) are calculated for the CG particles from the atomistic simulation, and pairwise interactions are then constructed such that the original RDFs are reproduced. In the second, a similar procedure is carried out for pairwise forces using a set of atomistic configurations: the forces on each CG particle are calculated for each configuration, and a least-squares fit is performed to obtain the CG forces. Although at first sight very dissimilar, the force-matching approach can also be related to the structural properties of the original system.[20]

The approaches above focus on determining CG interactions from atom-scale simulations. Another problem in constructing CG models is choosing good degrees of freedom: it is difficult to determine *a priori* whether some choice is better than another in describing the underlying system, given some constraints on the complexity of the model. A few different approaches have been proposed for this problem, mostly in the field of protein simulations.[21–25] These are either based on analysis of a single structure[21–23] or on preservation of dynamical information.[24,25] Use of clustering algorithms or self-organizing maps in this context has also been discussed.[26,27]

This paper focuses on the problem of determining the pairwise interactions, given the degrees of freedom and the RDFs. This is the so-called *inverse problem in classical statistical mechanics*, and it is known that such interactions are unique and they exist whenever such RDFs can be produced.[28–30] However, there are often practical problems in determining this unique interaction; one aim of this paper is to critically discuss these issues.

As a model system, we use a lipid bilayer composed of dipalmitoylphosphocholine (DPPC) and cholesterol, a system which we have also previously studied both at atomistic level[31] and using coarse-graining.[32,33] Already such a two-component system shows interesting phase behavior, the microscopic details of which are still elusive, in particular at in-

termediate cholesterol concentrations (see, e.g., Ref. 34 and references therein for recent experiments, and Ref. 35 for theoretical work). Computer simulations would, in principle, be ideal to study their detailed behavior. However, both the length and time scales needed to see phase separation in multicomponent lipid bilayers are currently beyond the reach of atomistic simulations. Hence, CG models are needed, and here we focus on a very simple model, with the idea that even such simple models could be able to give insight into the phase behavior if carefully parameterized. Also, a high degree of coarse-graining, i.e., going for as simple models as possible, provides a highly non-trivial test case for development and assessment of CG models and methodology.

In our previous CG studies, we have constructed simple 2D models for the large-scale structure of DPPC/cholesterol mixtures.[32,33] In the first model,[32] each molecule was replaced by a single point-like particle, and the second model[33] extended this description to include three particles for DPPC, one for the headgroup and one for each tail. The particles were thought of as describing the center-of-mass of (the part of) the molecule, and effective interactions were constructed to reproduce RDFs calculated from atomistic simulations. In this paper, we take one step further, and include the ordering of the DPPC tails in the model using a discrete, internal degree of freedom: each chain has a two-state degree of freedom that labels the chain as either ordered or disordered, similarly to the phenomenological models of Nielsen *et al.*.[36,37] Hence, this paper completes the process of systematically constructing a CG model that is similar to these phenomenological models; one of the original motivations was the success of these models in reproducing the phase behavior of lipid/sterol mixtures. As the interactions in the phenomenological models do not depend on the thermodynamic state, it is also interesting to study the effect of the new degrees of freedom on the transferability of the CG interactions: in the earlier models without internal states, transferability was quite poor, in particular across phase boundaries. It should then be expected that including the internal states would improve the transferability properties.

In this paper, we will first discuss the theoretical background of systematic coarse-graining and RDF inversion in Section II. Next, Section III discusses the inverse Monte Carlo (IMC) method,[13] which has been used in this work to achieve RDF inversion. Improvements and changes to the IMC algorithm, necessary for construction of the present model, are also presented. The construction of the new CG model is presented in Section IV, and associated problems are discussed. Monte Carlo (MC) simulations of larger systems using the model are described and discussed in Section V. Finally, Section VI contains a thorough discussion of the results, and also discusses the RDF-based coarse-graining approach in general.

II. COARSE-GRAINING AND CLASSICAL INVERSE PROBLEM

The aim of systematic coarse-graining is to reproduce the properties of the microscopic model as well as possible. If the microscopic degrees of freedom $\vec{R}$ and the CG degrees of freedom $\vec{r}$ are connected by a mapping function $\vec{r} = \vec{M}(\vec{R})$, we can define an *effective Hamiltonian* H_{eff} as

$$e^{-\beta H_{\text{eff}}(\vec{r})} = \int e^{-\beta H(\vec{R})} \delta(M(\vec{R}) - \vec{r}) d\vec{R}, \qquad (1)$$

where $H(\vec{R})$ is the configurational part of the microscopic Hamiltonian and $\delta(x)$ is the Dirac delta function. Although this equation is typically written such that $\vec{r}$ contains only the CG bead positions, it is also more generally valid. In particular, we can include discrete internal degrees of freedom into $\vec{r}$; in this case, the mapping function should also give values for the internal degrees of freedom. For clarity, we do not write these internal degrees of freedom explicitly, but it is to be understood that they are included in $\vec{r}$.

The right-hand side of (1) is the sum of the Boltzmann factors of all microstates consistent with specified values of the CG degrees of freedom $\vec{r}$; the left-hand side is just the Boltzmann factor of a system with Hamiltonian H_{eff}. Hence, if we could sample CG configurations according to H_{eff} defined by (1), all properties that depend only on the CG degrees of freedom $\vec{r}$ could be calculated exactly. Note that H_{eff} is a *free energy* of a particular configuration $\vec{r}$, and hence contains entropic effects in addition to the potential energy. For the same reason, H_{eff} depends on the thermodynamic state.

It is not practical to estimate H_{eff} directly for several reasons: (i) $\vec{r}$ can be high-dimensional, making it impossible to store H_{eff}, e.g., on a grid, (ii) as such, H_{eff} can only be applied to a system identical in size to the microscopic one, and (iii) evaluation of any free energies from microscopic simulations is costly. For problems (i) and (ii), it is possible to *approximate* H_{eff} with

$$\begin{aligned} H_{\text{eff}} &= w_0 + \sum_{i,j} w_2(r_{ij}) + \sum_{i,j,k} w_3(r_{ij}, r_{ik}, r_{jk}) + \dots \\ &\approx V_0 + \sum_{i,j} V_2(r_{ij}), \end{aligned} \qquad (2)$$

where we have assumed that the system is isotropic and homogeneous, and w_n is the direct n-particle interaction than can be fixed by requiring that (2) should hold for all N with the same functions w_n, where N is the number of particles. The sum on the first line can in principle range up to an N-body term. The second line suggests a computationally convenient approximation: the sum is replaced by an *effective pair interaction* $V_2(r)$, which also contains effects from the higher-order terms. In this form, we only need to store the pair interaction, and the form can also be extrapolated to larger systems. It should be noted that the approximation (2) results in an additional state dependence of H_{eff}: the best $V_2(r)$ may have a different state dependence than the exact H_{eff}. This state dependence can then lead to inconsistent thermodynamics if the origins of the pair interaction are not kept in mind.[38] Also, although the constant and the single-particle field [absorbed into V_0 in (2)] are independent of particle positions, they may depend on the thermodynamic state, and should be taken into account to treat thermodynamics accurately.[39]

Approximation (2) does not solve problem (iii); hence, some derived quantity, not H_{eff} itself, needs to be used for obtaining $V_2(r)$. Force matching uses derivatives of (1) with respect to $\vec{r}$;[18] the approach discussed here is based on structural quantities. The structural quantity most often used in this context is the RDF $g(r)$: it gives the probability of observing a particle at distance r from the origin, given that there is one at the origin. The normalization is such that $g(r) = 1$ everywhere for a homogeneous distribution, i.e., in an ideal gas.

It is not immediately clear that taking the RDFs as the target would result in well-defined interactions. A partial answer to this problem was provided by Henderson in 1974, who proved that if two systems with pairwise interactions give the same RDFs, the Hamiltonians can only differ by a constant, i.e., the pairwise interactions are *unique*.[28] Ten years later, a rigorous proof was given for the *existence* of such interactions.[29,30] It has also been shown that of all systems that reproduce the given RDFs, the one with pairwise interactions has the highest entropy.[40] More recently, RDF matching has been shown to be equivalent to maximizing the likelihood that the configurations produced by the CG model produce the atomistic probability distribution.[41]

The uniqueness proof is a straightforward application of the inequality

$$F_2 - F_1 \leq \langle H_2 - H_1 \rangle_1, \tag{3}$$

where $H_{1,2}$ are the Hamiltonians of the two systems, $F_{1,2}$ are their free energies, and the equality holds only if the Hamiltonians differ by a constant.[28] The existence proof is more complex, but the basic idea is to show that the functional $F[V]$ defined by

$$\beta F[V] = \ln \frac{\exp[-\beta \int V(x,x')\rho^{(2)}(x,x')dxdx']}{\int \exp\{-\beta[W(x_1,\ldots,x_N) + \sum_{i,j} V(x_i,x_j)]\}d^N x} \tag{4}$$

attains its maximum with a pair potential V, which is then shown to give the correct RDF g (which is the normalized version of the target two-particle density $\rho^{(2)}$).[29,30] $W(x_1,\ldots,x_N)$ is an *arbitrary fixed N-particle interaction* that acts in the system. The proof by Chayes *et al.*[29,30] then shows that for any such W (with some finiteness criteria) and any $\rho^{(2)}$ that is a two-particle reduction of an admissible N-particle density, there exists a unique pair potential that reproduces $\rho^{(2)}$. Here, admissibility just refers to finiteness of certain thermodynamical functionals.[30] We also note that F differs from the relative entropy defined in Ref. 41 only by a constant that does not depend on V. It should be noted that in the commonly used scenario where the system has a finite volume, the existence proof is strictly valid only in a system with the same volume as in which the RDFs were determined. In a system with a different volume, the pair interactions will also, in general, be different. Hence, when devising methods for the inversion, one should perform fitting for systems with identical sizes, as also noted in the original IMC paper.[13]

As discussed above, the RDFs calculated from any simulation define a unique set of effective pair interactions under quite general conditions. Thus, using them as a target in coarse-graining results, at least theoretically, in a well-defined model. By construction, this model then reproduces the short-range structure, although using the effective interactions in larger systems than in which they were determined can lead to minor changes in the RDFs.[33] In principle, the thermodynamics of the system can also be recovered through the compressibility route,[38,42] because the equation only involves RDFs. Other thermodynamic equations, such as the virial pressure, lead, in general, to different results.[38] Also, three-particle correlation functions will, in general, be different from the atomistic ones.[42,43]

III. INVERSE MONTE CARLO METHOD

IMC is one practical method for constructing the unique pair interactions that lead to given RDFs.[13] For other approaches, see, e.g., Ref 15 and references therein. In this section, we briefly describe the IMC method and our modifications and improvements.

Basic idea The central idea of the inverse Monte Carlo method is to write the Hamiltonian of the system in the form[13]

$$H = \sum_\alpha S_\alpha V_\alpha, \tag{5}$$

where the sum goes over potential bins, V_α is the value of the potential within the bin α, and S_α is the number of particle pairs that fall within this bin. Now, $\langle S_\alpha \rangle_{\text{MD}}$, calculated for the atomistic positions mapped to the CG degrees of freedom, is a grid approximation to the target RDFs that we want the CG model to reproduce. Using Eq. (5) for the Hamiltonian, one can then write a linear approximation for the change in $\langle S_\alpha \rangle$ in terms of changes to V_β as[13]

$$\Delta\langle S_\alpha \rangle = \sum_\beta \frac{\partial S_\alpha}{\partial V_\beta}\Delta V_\beta, \quad \frac{\partial S_\alpha}{\partial V_\beta} = -\frac{\langle S_\alpha S_\beta \rangle - \langle S_\alpha \rangle \langle S_\beta \rangle}{k_{\text{B}}T}. \tag{6}$$

Setting $\Delta\langle S_\alpha \rangle = \langle S_\alpha \rangle_{\text{MD}} - \langle S_\alpha \rangle$, one can then solve for ΔV from a linear set of equations. IMC then proceeds as follows. First, we take an initial guess V^0 for the potentials and evaluate $\langle S_\alpha \rangle$ and the derivatives. We then solve for ΔV and set $V^1 = V^0 + \Delta V$, resulting in a better approximation for the interactions. This procedure is then iterated until the RDFs match sufficiently well. Any interaction (or any other quantity which appears as such a linear term in the Hamiltonian) for which a target value $\langle S_\alpha \rangle_{\text{MD}}$ (or an equivalent quantity) is known can be included in the iteration, although for bonded interactions, special care must be taken to make the solution to the matrix equation unique: in this case, the derivative matrix has a zero eigenvalue for each bonded potential, corresponding to addition of a constant to the potential.[33]

Here, we note an alternative approach for arriving at the same equations: if the functional (4) is maximized with Newton's method and the Hamiltonian (5) is used, the equation for one iteration becomes exactly (6). This also gives some insight into the convergence of IMC: if the derivatives can be calculated exactly and the trial potential is close to the correct one, Newton's method converges quadratically.[44] How-

ever, the effect of noise on the derivatives requires more careful analysis, which is beyond the scope of the present discussion. Also, if the trial potential is not close, there is no guarantee of convergence; the direction of the change is an ascent direction for (4), but the step may be too long.

Regularization and thermodynamic constraints As described previously, it is advantageous to change the linear equation to a quadratic minimization problem:[33] instead of solving a matrix equation of the type $Ax = b$, we minimize $\|Ax - b\|$. In this form, it is possible to apply efficient regularization to better deal with noise in the input data. Further, it is also possible to add a thermodynamic constraint to constrain the surface tension of the system to avoid unphysical interactions.[33] That is, the potentials can be determined under the constraint that the surface tension γ is constrained to a predetermined value. We define γ using the standard virial formula

$$\gamma V = \langle E_{kin} \rangle + \frac{1}{2} \left\langle \sum_{i<j} f_{ij} r_{ij} \right\rangle, \tag{7}$$

where $\langle E_{kin} \rangle = N k_{\mathrm{B}} T$ is the average kinetic energy. Here, 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. The internal degrees of freedom only enter through the force f_{ij}, which depends on the internal states of the involved particles. The constraint can be treated identically to Ref. 33: the constraint that γ should remain at a particular value is added through a Lagrange multiplier. One should note that the value of the surface tension defined by (7) is not the thermodynamic surface tension of the system because of the state dependent interactions (see Section II). Because of this, the value needs to be fixed using some auxiliary quantity. In Ref. 33, the area compressibility was qualitatively fitted to experimental values, and the same surface tensions are also used in this study to facilitate comparison. It is also worth to note that the exact value of the surface tension does not alter the qualitative results obtained from the model.[33]

Internal degrees of freedom For the present model, we also have discrete degrees of freedom that describe the chain ordering. These enter the description through two ways: (i) each internal state is treated as a separate particle species, determining the pair interactions through which the particle interacts, and (ii) additional terms are needed in the Hamiltonian to describe the internal (free) energies of the states. The Hamiltonian that now needs to be parameterized is

$$H = \sum_{\alpha} S_\alpha V_\alpha + \sum_i E_i n_i + \sum_{i,j} E_{ij} \delta n_i \delta n_j, \tag{8}$$

where E_i and E_{ij} are parameters and n_i are the numbers of particles in each state. $\delta n_i = n_i - n_i^{ave}$ measures the fluctuations from an arbitrary *constant* reference number n_i^{ave}. Since the total number of particles of each type remains constant, the state occupancies n_i are not all independent of each other. To make the Hamiltonian unique, Eq. (8) should be interpreted such that the sums only include an independent set of the n_i. In the case of the present model, this means that the total number of tails is a constant. Hence, only the number of disordered tails (denoted as n_d) is required (or equivalently n_o, the number of ordered tails). The Hamiltonian then becomes

$$H = \sum_{\alpha} S_\alpha V_\alpha + \Delta E n_d + E_{\text{fluct}} \delta {n_d}^2, \tag{9}$$

where we have used ΔE to denote the internal energy difference between the ordered and disordered states, and E_{fluct} for the second-order term.

Full justification for Eq. (8) is given in Appendix A. Here it suffices to note that the second-order term in n_i, which distinguishes the system from a standard semi-grand canonical ensemble, is needed for precise control of the number of particle pairs in the system: $n_i n_j$ terms appear in the number of pairs in the system, and a certain number of pairs is required for exact reproduction of the two-particle densities.

The non-linearity in n_i makes it non-trivial to generalize the model to a larger system as the Hamiltonian as such is no longer an extensive quantity. If the particle number increases by a factor M, n_i^{ave} should be transformed to $M n_i^{ave}$ to preserve $\langle n_i \rangle / N$, where N is the total particle number. However, the E_{ij} terms can be treated in two different ways: (i) E_{ij} are kept constant, or (ii) E_{ij} are scaled to E_{ij}/M. However, neither works well if $E_{ij} < 0$, as (ii) does not preserve the energetics of flipping a single state, while (i) can make the cases $n_i = 0$ and $n_i = N$ exceedingly favorable. As accurate matching of the RDFs seems to lead to such problems in the present case, we have only used interactions with $E_{ij} \geq 0$ for larger systems, and the other cases are only studied in a system identical in size to the underlying MD.

It is also possible to apply the IMC such that E_{ij} is set to zero and not included in the iteration, also neglecting the value of $\langle \delta n_i \delta n_j \rangle$ during the inversion. This corresponds to the standard semi-grand canonical ensemble. This can be combined with any of the implementations discussed below, and is applied in some of the calculations discussed in Sections IV C and V.

Straightforward implementation IMC can be generalized straightforwardly to Hamiltonians of the form (8) since all the parameters appear linearly in the Hamiltonian. The calculation of the derivatives can be carried out as before, and the linearization (6) still holds with additional terms for the derivatives with respect to E_i and E_{ij}. A similar linearization can also be written for the $\Delta\langle n_i \rangle$ and $\Delta\langle \delta n_i \delta n_j \rangle$, which results in an extended set of linear equations that can be solved for changes in the potentials and the internal energies. Subsequently, this approach is referred to as method A. In Section IV C, we also briefly employ a variant where E_i are fixed and the linearization (6) for each $\Delta\langle n_i \rangle$ are used as a constraint to keep $\langle n_i \rangle$ at its target values. This variant is referred to as A′.

Alternative implementations There are issues with method A that may result in poor generalizability of the determined interactions to larger systems; these are discussed in more detail in Sections IV C and VI. Hence, we have implemented several alternative approaches, collectively referred to as method B. The basic principle in this approach is to optimize the RDFs and the energy parameters separately.

This is achieved by calculating the target pair counts for each iteration as $S_\alpha^t = \langle S_\alpha \rangle_{\mathrm{MD}} \langle N_\alpha \rangle / \langle N_\alpha \rangle_{\mathrm{MD}}$, where N_α is the number of pairs interacting through the interaction associated with S_α. This effectively separates the changes in the structure from changes in the number of particle pairs. This is in contrast with method A, where the target pair counts are calculated directly from the target two-particle density ($S_\alpha^t = \langle S_\alpha \rangle_{\mathrm{MD}}$), and remain the same for every iteration. For consistency, we also need to use (6) for $\langle S_\alpha \rangle$ without the terms coming from E_i or E_{ij} in method B. Essentially, method B then solves ΔV as if there were no internal states, only scaling the target pair counts appropriately.

Several different subimplementations of method B can be imagined, differing in the way the derivative matrix $\partial \langle S_\alpha \rangle / \partial V_\beta$ is evaluated:

(i) Use (6) directly (without terms from E_i or E_{ij}).

(ii) Do an *ad hoc* modification to factor out the effects of state occupancy changes on the pair counts, using a replacement $S_\gamma \to S_\gamma \langle N_\gamma \rangle / N_\gamma$ inside all the expectation values in (6).

(iii) Calculate the derivative for $\langle S_\alpha \rangle / \langle N_\alpha \rangle$ instead of $\langle S_\alpha \rangle$.

Of these, (i) is the simplest, but it does not necessarily converge to the correct solution, because the derivative matrix still contains effects from changes in the number of pairs. The concrete effect is that E_i remain close to their initial values, because any change in the interactions that would change the pair counts is disfavored. This limits the space of interactions that the method searches, but within this space, it still finds a good match for the RDFs. Approach (ii) cannot be easily justified theoretically, but solves the above problem. (iii) leads to a non-symmetric matrix and has stability problems, and was only studied briefly in this work. The different approaches are referred to as B(i), B(ii) and B(iii) later in the text.

The implementation so far discussed does consider how E_i and E_{ij} should change. For all the subimplementations, we have used the same approach to adjust these parameters. First, after the potential change has been evaluated, E_i are adjusted using the linearizations (6) for $\Delta \langle n_i \rangle$ to keep $\langle n_i \rangle$ constant. An additional adjustment is then made to obtain the desired $\Delta \langle n_i \rangle$ and $\Delta \langle \delta n_i \delta n_j \rangle$, using the iteration formulae derived in Appendix B. Here, let us only quote the final formulae:

$$\Delta E^{(2)} = \frac{1}{2\beta} \Delta(\sigma^{-1}), \tag{10}$$

$$\beta\sigma(\Delta E^{(1)} + 2\Delta E^{(2)} \langle n \rangle) = -\Delta \langle n \rangle + \Delta\sigma(\sigma^{-1} \langle n \rangle). \tag{11}$$

Here, $E^{(1)}$ and $E^{(2)}$ are the vector and matrix formed from E_i and E_{ij}, respectively, $\langle n \rangle$ is a vector with the average state occupancies from the CG simulation, and σ is a matrix with the fluctuations in the state occupancies from the CG simulation. In our implementation, we have neglected the second term in the latter equation, because $\Delta\sigma$ is typically small.

IV. MODEL CONSTRUCTION

A. General

As before, we construct a simple model with the following assumptions.[32,33] First, the interaction between the two monolayers is assumed to be weak, allowing us to construct a model for one monolayer only. The effect of undulations is also assumed weak. Together, these assumptions allow a 2D model to be constructed. Also, we assume that the system can be adequately described with pairwise and isotropic effective potentials.

Our new model uses three particles per each DPPC molecule, similarly to Ref. 33. One particle is used for the headgroup and the glycerol parts and one for each of the tails. One particle is used for each cholesterol molecule. These particles are thought to describe the center-of-mass (CM) positions of the corresponding groups of atoms. Each tail particle has a two-state internal degree of freedom that describes ordering. This state can change during the simulation, and it selects the non-bonded interaction tables that the particle uses. Although the tails of a DPPC molecule are not completely equivalent, we assume that the non-bonded interactions of these particles can be treated identically.

For bonded interactions, there are two separate head–tail bond interactions (one for each tail) and one tail–tail interaction, but no angle potentials. The derivation for Eq. (8) is strictly valid only when exclusions are not used, i.e., bonded particles should also interact through the non-bonded interactions. Hence, when considering models with the full Hamiltonian ($E_{\text{fluct}} \neq 0$), non-bonded interactions are also present between bonded pairs, in contrast to Ref. 33. If E_{fluct} is zero, the treatment of exclusions is theoretically not as important, and in this case we did use exclusions. For simplicity, we assume that the bonded interactions do not depend on the ordering state of the particles.

With the above assumptions, the model consists of four types of particles (headgroup, ordered and disordered tails, and cholesterol) that interact through ten different non-bonded interactions (one for each pair) and three different bonded interactions. There are also two internal energy parameters [see Eq. (9)]. We set the reference point for the fluctuations as $n_d^{ave} = \langle n_d \rangle_{\mathrm{MD}}$.

As with the previous models,[32,33] the IMC process requires the RDFs between all pairs of coarse-grained particles and the areas per molecule from the atomistic simulations. In addition, $\langle n_d \rangle_{\mathrm{MD}}$ and $\langle \delta n_d \delta n_d \rangle_{\mathrm{MD}}$ are needed. For method A, we need the two-particle density $\langle S_\alpha \rangle_{\mathrm{MD}}$ and not the RDFs directly; hence, we also need the expected number of particle pairs for each RDF (or we can calculate the two-particle density directly).

In order to calculate the above quantities from atomistic simulations, we need a definition that allows us to characterize any conformation of a DPPC tail as either ordered or disordered. For this purpose, we have decided to use the chain order parameter that is commonly used in characterizing the ordering of the tails, even when considering instantaneous order.[35,45,46] Here, we have averaged the order parameter over

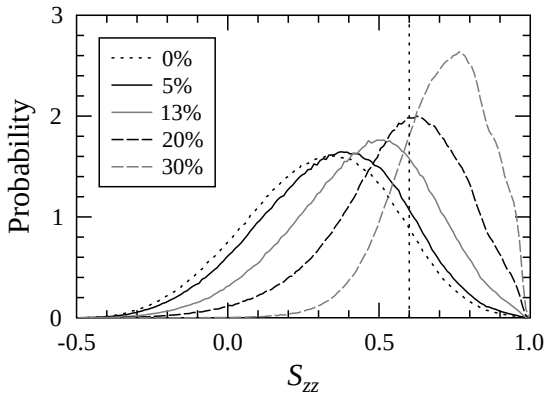

FIG. 1: Order parameter distributions for different cholesterol concentrations calculated from MD data. The vertical dotted line shows the cutoff employed for classifying a chain as ordered (see text for details).

all carbon atoms in the chain:

$$S_{zz} = \frac{1}{N_c - 2} \sum_{i=2}^{N_c - 1} \frac{1}{2}(3\cos^2\theta_i - 1), \qquad (12)$$

where $N_c = 16$ is the number of carbon atoms in a chain and θ_i is the angle between the bilayer normal and the orientation of the chain at the i'th carbon. We have taken the z axis as the bilayer normal, and the orientation θ_i is calculated using the vector connecting the carbons $i - 1$ and $i + 1$. A chain was identified as being ordered if $S_{zz} > C$, where C is a cutoff value whose selection will be discussed below.

B. MD Analysis

The MD simulations used as the basis for the model construction were the same as with the previous models in Refs. 32 and 33, and details of the simulations can be found in Ref. 31. Briefly, the simulated systems consisted of 128 fully hydrated lipid molecules, i.e., DPPCs and cholesterols. The Berger set of parameters was used for DPPC,[47] and cholesterol parameters were taken from Höltje *et al.*[48]. The simulations were performed with 0, 3, 8, 13, and 19 cholesterols per monolayer, corresponding to cholesterol molar concentrations 0%, 5%, 13%, 20%, and 30%. Each simulation was run for 100 ns in the NpT ensemble at $T = 323$ K using the GROMACS simulation package,[49] and first 20 ns was treated as equilibration.

Figure 1 shows the distribution of instantaneous chain S_{zz} values for the different cholesterol concentrations. The distribution is smooth with a single maximum at all concentrations, showing that there is no clear distinction between ordered and disordered chains. Hence, the selection of the S_{zz} cutoff for our two-state model is somewhat arbitrary, and is only limited by the condition that at all concentrations, there should be a significant fraction of both ordered and disordered chains to give reasonable statistics for all the RDFs. Also, *a priori* it is reasonable to take the cutoff at the same value for each concentration. Based on these considerations, a value of 0.6 was selected. However, it should be noted that other cutoffs could also be used without any clear advantage or disadvantage. More than two states could also be justified, but would lead to a significant increase in the number of interactions and would also increase the statistical noise in the RDFs. This makes it a substantial amount of work to construct such models. As there are no clear *a priori* advantages over two-state models, such models were not studied in this work in any detail.

Figure 2 shows the RDFs calculated with an ordered chain defined by $S_{zz} > 0.6$. The fraction of ordered chains as a function of cholesterol concentration is also shown, together with the fluctuation around this average value. Bonded RDFs are not shown, as they are identical to those published previously,[33] and are not important for the present discussion. The RDFs were calculated from the 2D projections of the CM positions corresponding to the CG particles, considering each monolayer separately. Note that bonded pairs were not excluded in the calculation, i.e., the results cannot be directly compared to those in Ref. 33, where exclusions were used. This treatment is in line with the selected form of the Hamiltonian, i.e., exactly those pairs that would interact through a certain CG pair potential were included in the calculation of the corresponding RDF.

Qualitatively, the RDFs behave as in Ref. 33, i.e., order generally increases with increasing cholesterol concentration, while still maintaining a liquid-like structure. This effect is most clearly seen in the RDFs in the tail region, i.e., the two bottom rows. However, the division of the tails into ordered and disordered shows that the heights of the RDF peaks increase the most for the RDFs involving the disordered tails. In contrast, the nearest-neighbor peak of ordered tails actually decreases when cholesterol concentration increases. Nevertheless, the minima become deeper for all the RDFs. The fact that the ordering of disordered tails also increases indicates that the presence of cholesterol has an effect that is not only mediated by tail ordering, at least when ordering is measured using S_{zz} only (the RDFs behave similarly for other S_{zz} cutoffs as well). Also, the ordered–disordered RDF shows much smaller changes than the others, the peaks remaining small. This indicates that tails prefer to be next to other tails which are in the same state, i.e., that ordered and disordered tails tend to occupy different regions. The ordered fraction behaves as expected: with increasing cholesterol concentration, more and more tails become ordered. Fluctuations around the average are quite small, with a maximum at 20% concentration. This maximum is easily explained by the fact that at this concentration, the ordered fraction is closest to 0.5, i.e., entropic effects from the different number of microstates for different fractions limit the fluctuations least at this concentration.

Areas per molecule were also obtained for each cholesterol concentration by dividing the average total box area by the number of molecules in a monolayer (i.e., by 64). The values are identical to those used for the previous models,[31–33] and are omitted here. It is sufficient to note that the area decreases monotonically with increasing cholesterol concentra-

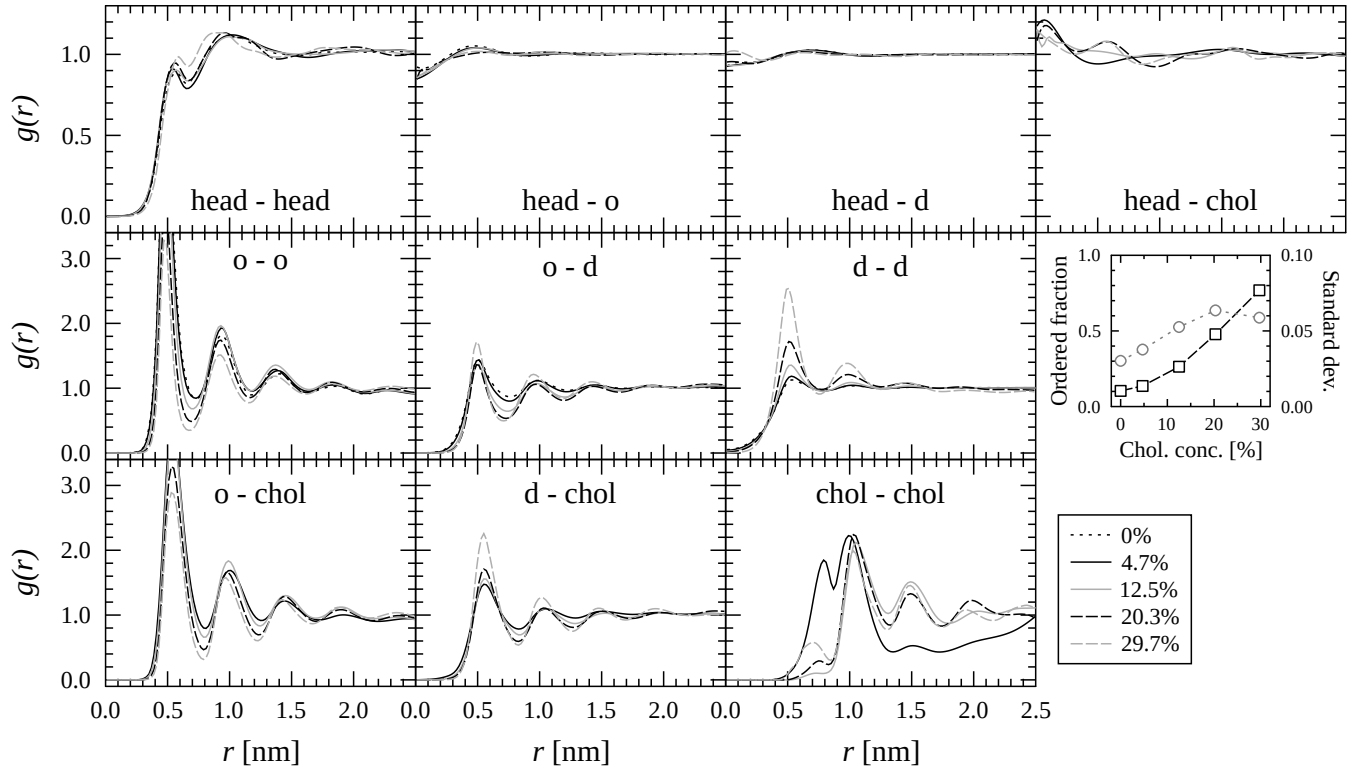

FIG. 2: Radial distribution functions for pairs of CG beads calculated from atomistic MD simulations calculated without exclusions. The fraction of ordered chains (squares) and the fluctuations around this average (circles) are shown in the small figure. A chain was identified as ordered if $S_{zz} > 0.6$. The bonded RDFs are identical to those published previously,[33] and are not shown here. RDFs not involving tail particles are identical to those in Ref. 33.

tion, starting from $65\,\text{Å}^2$ for the 0% concentration and ending at $42\,\text{Å}^2$ at the 30% concentration.[31] The linear sizes of the MD systems hence decreased from 6.5 nm to 5.2 nm.

C. Determination of Interactions

Several sets of effective interactions for the CG particles were constructed using the different IMC approaches described in Sec. III, using the RDFs in Fig. 2 (or ones with exclusions). The details of the calculations are given in Table I, and discussed in more detail below. The names in the first column of the table are used to refer to the calculations throughout the text.

For all calculations, a separate interaction was used for each pair used for RDF calculation (Fig. 2 plus three bonded interactions[33]). The cutoff for all interactions was 2.5 nm, a limit imposed by the fact that the RDFs can only be calculated up to half the linear size of the system. The input RDFs and the effective potentials were pre- and postprocessed identically to the previous model (Ref. 33), i.e., a spline-fitting algorithm[50] was used to smooth the input RDFs and the final potentials, and power-law forms were used for the potentials in regions where the RDFs were zero.

The sampling for the IMC runs, as well as the simulations reported in Section V, were done using standard Metropolis Monte Carlo. Two types of moves were used: (i) a randomly selected particle was displaced, or (ii) the internal state of a randomly selected chain was flipped.

Using the potential of mean forces ($V(r) = -k_\text{B}T \ln g(r)$) as the initial potentials resulted in very unstable iteration. Hence, for each concentration, we first performed an IMC run where the internal states were treated as identical. The final potentials from this run were then used as the initial potentials for the full IMC inversion.

As with our previous model,[33] we observed that the surface tension needs to be constrained during the IMC process to obtain positive virial pressures. The target values for the surface tension constraints were the same as with the previous model to facilitate comparison, and the obtained area compressibilities were within 10% of those obtained within the previous model. The constraint was applied for all the reported calculations: for pure DPPC systems, the surface tension was constrained to $\gamma \approx 90\,\text{dyn/cm}$, while for cholesterol-containing systems a value of $\gamma \approx 220\,\text{dyn/cm}$ was used.[33]

Figure 3 shows the set of potentials for calculation B2, which is also the set used in Section V to study large-scale behavior of the system (the potentials used for transferability studies are shown in supplementary material, Fig. S1[51]). The internal energy terms are also shown as a function of cholesterol concentration. The B2 potentials were determined using IMC approach B(i) from the RDFs in Fig. 2. The po-

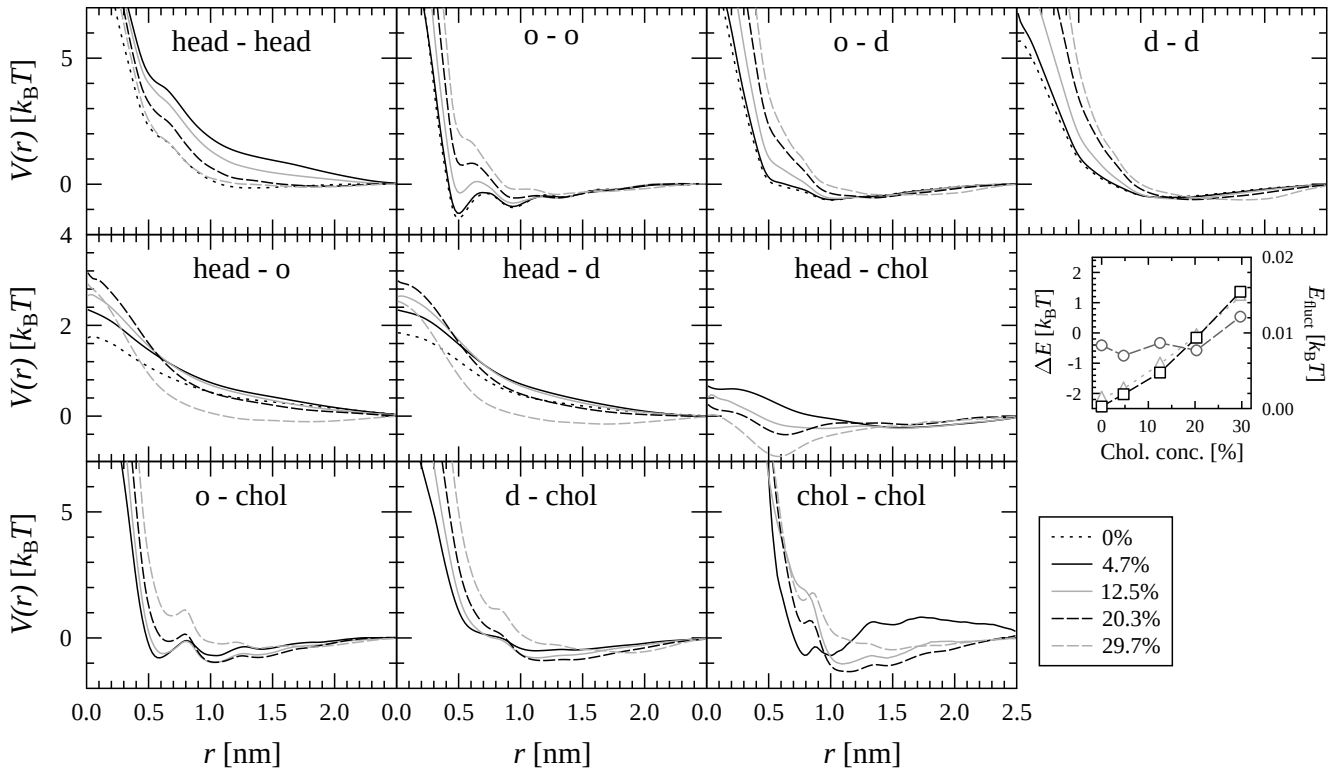

FIG. 3: Effective pairwise interactions determined from RDFs in Fig. 2 using IMC approach B(i) (calculation B2). The small figure shows the internal energy difference between the disordered and ordered tails (squares), as well as the magnitude of the fluctuation energy term (circles). The initial ΔE values in the iteration are shown as triangles.

TABLE I: List of IMC calculations performed. All the calculations were performed only for the 30% cholesterol concentration; only calculations A3, B1, B2, and B3 were performed for all concentrations.

Name	Method	ΔE $[k_BT]^d$	E_{fluct} $[10^{-2}k_BT]^d$	Fig. 4 style
A1	A	-22	-2.5	grey solid
A2	A′	1.4	0.7	grey dashed
A3	A′^{c}	1.48	0	grey dash-dotted
B1	B(ii)	-21	-2.1	black solid
B2^a	B(i)	1.4	1.2	black dashed
B3^b	B(i)^c	1.7	0	black dash-dotted
B4	B(i)^{c,m}	-3.9	0	black dotted

^{a} Used for large-scale studies and shown in Fig. 3.. ^{b} Used for transferability studies and shown in Fig. S1.[51] ^{c} In standard semi-grand canonical ensemble with non-bonded interactions excluded between bonded particles. ^{d} Energy values given for 30% cholesterol concentration. ^{m} With a manual intervention to change ΔE.

tentials are, in general, very soft due to the high degree of coarse-graining.[32,33] Note also that the internal energies ΔE are very close to the potential of mean force result $\Delta E = -kT\ln(\langle n_d\rangle_{\text{MD}}/\langle n_o\rangle_{\text{MD}})$ used as the initial guess for ΔE (shown in triangles in the figure), as expected for approach B(i).

The behavior of the effective potentials as a function of cholesterol concentration is very similar to the earlier three-particle model in Ref. 33. In the tail region, higher cholesterol concentrations imply stronger repulsion at short ranges, while at intermediate ranges the 20% potentials are the most attractive, with monotonic reduction in attractiveness when the concentration is changed either way. The main difference between the potentials involving ordered and disordered tails seems to be the increase in detail at short length scales for the potentials involving ordered tails. Interestingly, the inclusion of the ordering in the model has very little effect in the strength of the concentration dependence, in contrast to our *a priori* assumption that the new degree of freedom would improve transferability.[33] However, despite the concentration dependence, the transferability properties actually improve; this is discussed in more detail in Section V.

The B2 potentials in Fig. 3 do not result in perfect RDF agreement. This is because the method B(i) restricts ΔE to an incorrect value. However, one should note that the root-mean-square difference between the CG and the MD RDFs is of the order of 0.02 or less, i.e., the differences are mostly of the order of the line width in Fig. 2 (the differences are shown in Figs. S2 and S3 in the supplementary material[51]). Nevertheless, these minor differences prompted us to construct several different sets of interactions to understand the RDF inversion in more detail. Largest differences between different approaches were observed in the highest cholesterol concentrations, and hence we focus here on only the 30% case.

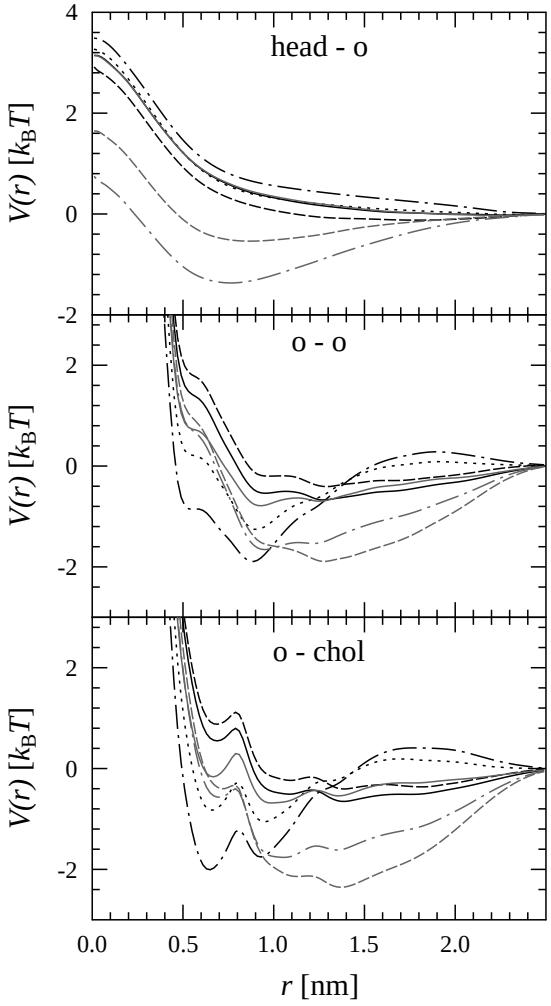

FIG. 4: Effective pairwise interactions for 30% cholesterol concentration from different calculations in Table I. All calculations give essentially the same RDFs and virial pressure for a small system. Each calculation is identified by a different line style, as described in Table I. The dashed black line corresponds to Fig. 3. A figure of the omitted potentials is provided as supplementary material (Fig. S4).[51]

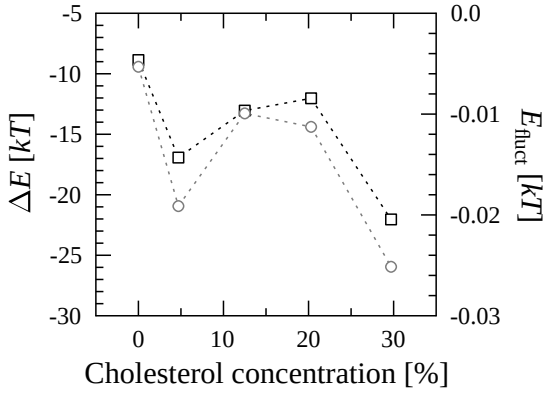

FIG. 5: Internal energy terms for interactions determined with IMC approach A without any constraints (calculation A1). ΔE is shown in squares, E_{fluct} in circles. Full set of potentials is provided as supplementary material (Fig. S5).[51]

It turned out that not all of the potentials generalized well to larger systems, as will be discussed below in more detail. For the pure DPPC, all potentials worked, but the problems became larger and larger as cholesterol concentration increased.

Figure 4 shows all the different potential sets from Table I for the 30% cholesterol concentration. Only head–ordered, ordered–ordered, and ordered–cholesterol potentials are shown for brevity; the full figure is provided as supplementary material (Fig. S4).[51] Each set of potentials is identified with a different line style as described in Table I. The potentials from Fig. 3 are shown in dashed black. The color of the line determines the IMC approach: black lines used approach B, grey lines approach A. The solid lines (A1 and B1) imposed no restrictions on the internal energy terms, and used either A or B(ii). Dashed lines (A2 and B2) used either A′ or B(i), and had $\Delta E \approx -kT\ln(\langle n_d\rangle_{\text{MD}}/\langle n_o\rangle_{\text{MD}})$. The dash-dotted lines (A3 and B3) are otherwise similar, but have, in addition, $E_{\text{fluct}} = 0$. The dotted black line (B4) had $\Delta E \approx -4kT$ and $E_{\text{fluct}} = 0$, i.e., ΔE was in between those of B1 and B2 (approach B(i) was used with some manual intervention to modify the internal energies).

The best RDF agreement is achieved, as expected, with the full Newton-type method in A1, i.e., the solid grey line in Fig. 4 (see Figs. S2 and S3 in the supplementary material[51]). Methods B(ii) and B(iii) also lead to similar potentials and internal energy values (calculation B1), although B(iii) seems to have stability problems and was only briefly studied. The behavior of the internal energies for the A1 potentials is shown in Fig. 5; the full set of potentials can be found in the supplementary material (Fig. S5).[51] The difference to Fig. 3 is quite large, as ΔE is now of the order of -10 to $-20kT$. The lower end is of the same order of magnitude as the free energy difference between the ordered and disordered states in the phenomenological models of Nielsen *et al.*[36,37] The energy terms ΔE and E_{fluct} also show non-monotonic behavior as a function of cholesterol concentration. Their magnitude seems to be closely coupled, with a larger $-\Delta E$ resulting in a larger $-E_{\text{fluct}}$. It is not straightforward to give a physical meaning to these parameters; they are mainly mathematical constructs that are used in combination with the effective pair interactions to estimate the effective Hamiltonian, which includes both energetic and entropic contributions. Nevertheless, the qualitative match between the phenomenological models and the present model at most concentrations is encouraging.

By construction, each of the interaction sets should reproduce the RDFs when a system identical in size to the original MD system is simulated. This is indeed the case, with the root-mean-square difference between the CG and MD RDFs being of the order of 0.02 or less in all cases. All of the sets seem to have minor systematic deviations from the target RDFs (see Figs. S2 and S3 in the supplementary material[51]), but it is very difficult to judge whether they are significant or not, in particular *a priori* without comparison to other potentials. However, we performed a test that shows that the differences, although almost invisible to the eye, actually are

important: if we take the A2 potentials for the 30% concentration and use them to produce the target RDFs for an IMC run that is otherwise identical to the B2 run, the A2 potentials are recovered. The same holds true if the roles of the two potential sets are reversed. Similar observations have also been reported by Bolhuis and Louis in the context of semi-dilute polymer systems.[52]

Although the differences between the potentials in Fig. 4 were minor in a system identical in size to the atomistic simulation, major differences were found when linear system size was increased eight-fold (we still focus on the 30% case, but similar effects could also be seen in lower concentrations, although not as clearly). This is in contrast to our earlier model, where the value of the virial pressure could be constrained to different values without any qualitative changes in the behavior.[33]

In the present case, potentials A2 and A3 result in clearly unphysical situations, having dense clusters of particles separated by empty space. For A1 and B1, the failure is more subtle and can be contested: these potentials result in very strong density variations, caused by phase separation of cholesterol and ordered tails in one phase and disordered tails in another phase. In principle, such a behavior could occur also for the 30% case, but the number $\langle n_d \rangle / N$ actually deviates significantly from the value used in the parameterization, casting some doubt on the validity. Also, this does not happen for other interactions, nor for our earlier models, so we have opted to use interactions for which this does not happen. Representative configurations from both these cases are included as supplementary material (Fig. S6).[51]

The potentials determined with IMC approach B(i), i.e., B2, B3, and B4, did not have the above problems (a representative configuration of this case is also shown in the supplementary material[51]). All of these gave similar results. For B2 (Fig. 3), where E_{fluct} was not zero, the interactions worked only when the E_{fluct} were not scaled with the number of particles; if scaling was used, similar problems as above were observed. The reason for these differences is returned to in Section VI, after discussing the large-scale behavior of the model with the potentials B2 and B3. For these interactions, the RDFs change only little when the system size is increased (see Figs. S7, S8, and S9 in the supplementary material[51]). However, there is an effect that comes from segregation of the ordered and disordered tails into separate regions (see next section).

V. BEHAVIOR AT LARGE LENGTH SCALES

Let us now briefly analyze the behavior of the new CG model at larger scales. The results presented here were obtained using the potentials B2 from Fig. 3 (transferability studies were done with B3 that had $E_{\text{fluct}} = 0$, shown in Fig. S1 in the supplementary material[51]). The Monte Carlo simulations were mostly conducted for systems with 16 times the linear size of the original MD simulation (transferability studies were done with half this size). That is, the linear system size was in the range 80–110 nm, depending on the cholesterol concentration. Reasonable statistics, good for qualitative conclusions, could be obtained within a few days on a standard desktop computer.

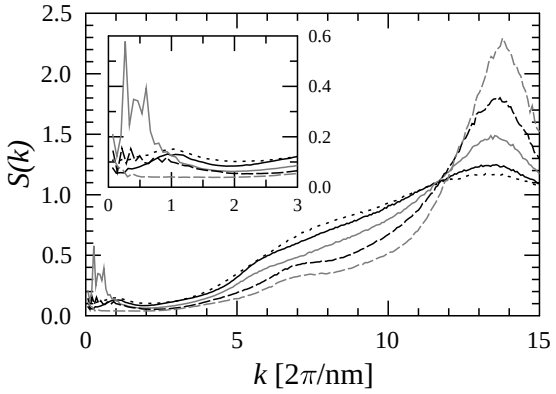

FIG. 6: Total static structure factors for different cholesterol concentrations. The inset shows the small k region magnified. Different line styles correspond to different cholesterol concentrations as in Fig. 3. The structure factors were obtained from systems with 16 times the linear size of the original MD simulations, with the interactions from Fig. 3.

We use the static structure factor $S(\vec{k})$ to characterize the long-range order in the system. $S(\vec{k})$ is defined by

$$S(\vec{k}) = \frac{1}{N} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \exp[-i\vec{k} \cdot (\vec{r}_j - \vec{r}_j)] \right\rangle , \qquad (13)$$

where N is the number of particles, $\vec{r}_i$ are the positions of the particles, and $\vec{k}$ is a reciprocal space vector. It is also possible to calculate the structure factor for a subset of the particles by restricting the summations to these particles. Although in principle, $S(\vec{k})$ and the corresponding RDF carry the same information (they are related by a Fourier transform), the static structure factor is more convenient to characterize the long-range structure (small $\vec{k}$ values).

Figure 6 shows the total static structure factor for each cholesterol concentration. The behavior is very similar to the earlier models: at 13% and 20% concentrations, there is some large-scale order, as shown by an increasing $S(k)$ when k is reduced. However, this effect is weaker than for the three-particle model without ordering[33] and comparable to the one-particle model.[32] For 20% concentration, it is even slightly weaker than for either of the previous models. Nevertheless, decomposition of the structure factors into partial structure factors shows that the cholesterol–cholesterol structure factor has a similar peak at small k. In the present model, the ordered and disordered tails give an additional contribution to the peak (see below). We also note that the 0% and 5% systems have a peak around $k \approx 1\,\text{nm}^{-1}$ similarly to the earlier three-particle model; this peak was interpreted as indicating strong density fluctuations in the fluid phase of pure DPPC systems, and was later confirmed through large-scale atomistic simulations.[46]

The new feature of the present model is chain ordering, and it is interesting to see how the different chain states take part in the behavior discussed above. This is shown in Fig. 7: it

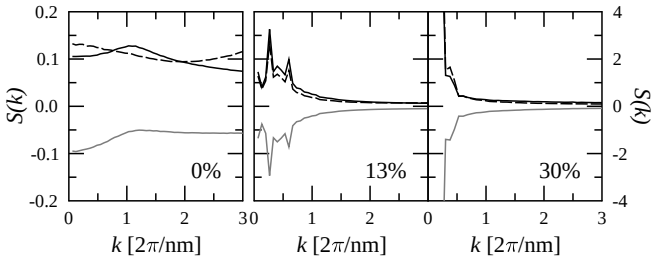

FIG. 7: Partial structure factors at three cholesterol concentration for o–o (solid black), d–d (dashed), and o–d (grey) pairs. The structure factors were obtained from systems with 16 times the linear size of the original MD simulations, with the interactions from Fig. 3. Note that the large opposite peaks in 30% cancel more or less exactly to give the total structure factor in Fig. 6 (there is also a small contribution from the other partial structure factors).

shows the partial static structure factors for the tail–tail pairs for the 0%, 13%, and 30% concentrations. Both the o–o and d–d structure factors increase as $k \to 0$ at all concentrations, while the cross-$S(k)$ is negative, and becomes more so at the $k \to 0$ limit. This shows that at all concentrations, the ordered and disordered tails tend to segregate and form larger domains; however, at the 30% concentration the total density remains homogeneous, as seen in the total structure factor (Fig. 6). At 0% and 5% concentrations, it is seen that exactly this segregation is the reason for the peak at $k \approx 1\,\mathrm{nm}^{-1}$. This is in agreement with the atomic-scale simulations, which showed that the chains within the denser domains are generally more ordered.[46] The RDF comparison shown in the supplementary material (Figs. S7, S8, and S9)[51] also show that such segregation occurs when system size is increased.

One of the main motivations for including the chain ordering was an *a priori* assumption that it would improve the transferability of the effective interactions between the different concentrations.[33] As before, we evaluate the transferability by using the interactions determined at a certain concentration to study the neighboring concentrations, and compare the results to those obtained with the correct effective potential. This addresses the consistency of the effective interactions evaluated from different atomistic simulations, and assesses whether it could be possible to evaluate, e.g., phase boundaries with the current model. However, it does not provide any direct measure for the quality of the CG model.

In the general case, the form (8) for the Hamiltonian makes transferability studies more difficult. Namely, it is not clear how the final term should be treated, as it includes information from the state in the values n_i^{ave}. In the present case, we only studied the transferability of potentials B3, which had $E_{\mathrm{fluct}} = 0$, and hence these problems were avoided. A figure of these potentials is included as supplementary material (Fig. S3).[51]

The transferability results are summarized in Fig. 8 (similar figures, but with RDFs instead of structure factors, can be found in the supplementary material, Figs. S10 and S11[51]). Between all concentrations, the results are qualitatively correct for the nearest-neighbor peak ($k \approx 13\,\mathrm{nm}^{-1}$). This is demonstrated in the middle figure, which shows the transferability between 5% and 13% concentrations: although the change in the nearest-neighbor peak is smaller than the difference between the correct structure factors, it is in the correct direction. Between 20% and 30%, as well as from 13% to 20%, the transferability is even better in this respect: also the height of the nearest-neighbor peak is reproduced (the bottom figure). For the small k region, the results are not as good: the interactions determined at 5% or 30% do not give the marked increase at 13% or 20% concentrations. However, the opposite direction works better: very little remains of the small k peak when the interactions from 13% and 20% are used in 5% and 30%, respectively. Between 13% and 20% the transferability is reasonable in both directions, although from 13% to 20% the small k peak increases slightly instead of decreasing.

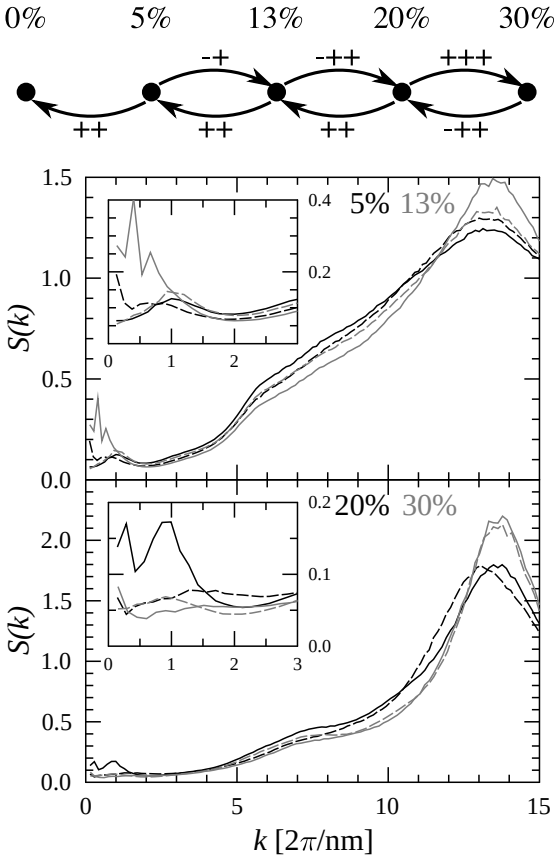

FIG. 8: Potential transferability properties between different concentrations using interactions B3 (potentials shown in the supplementary material,[51] Fig. S1). The diagram on top summarizes the transferability properties between the concentrations: the first +/- stands for (qualitative) reproduction of the small k behavior of $S(k)$, the second + for qualitative reproduction of the nearest neighbor peak in $S(k)$, and the possible third + for quantitatively nearly correct $S(k)$ away from the small k region. The figures on bottom show total static structure factors for two cases: in both figures, the correct $S(k)$ is shown as a solid line, and the $S(k)$ given by the transfered interactions as a dashed line. Each figure shows the transferability in both directions between adjacent concentrations, and the color of the lines shows the simulation concentration.

The transferability is mostly similar to the earlier three-

particle model without the internal degrees of freedom:[33] also the earlier model showed qualitatively correct behavior of the nearest-neighbor peak. However, there are two improvements: (i) the behavior of the nearest-neighbor peak is also quantitatively correct in some cases, and (ii) the small k region shows a degree of transferability also across the phase boundaries, i.e., from the inhomogeneous concentrations 13% and 20% to the homogeneous ones 5% and 30%. However, the latter does not work in the opposite direction, i.e., from the homogeneous to the inhomogeneous. Hence, the transferability did improve, but only slightly.

It could be possible to improve the transferability by improving the mapping of the atomistic simulations to the CG degrees of freedom, in particular for the ordering states. This could be done either by having more than two internal states or with an different definition of the ordered–disordered division. However, the S_{zz} order parameter distributions in Fig. 1 seem to indicate that the order changes quite continuously, so there might not be any good way of dividing the tails into two (or more) well-defined, separate groups. Increasing the number of states is also problematic from the technical point of view, as discussed in Section IV B. It is also possible that the system is simply too complex for good transferability. The fact that all the different models we have constructed actually have similar cholesterol–cholesterol interactions indicates that the concentration dependence might be an intrinsic feature of the chosen particle description. The transferability properties may also be adversely affected by sampling issues in the underlying MD; these are briefly discussed in the next section.

VI. DISCUSSION AND CONCLUSIONS

Although theoretically the solution to the inverse problem is unique, we have demonstrated that widely different potentials can give rise to essentially the same RDFs (see Fig. 4), in line with observations by Bolhuis and Louis.[52] Further, in our earlier study the virial pressure of the system could be constrained to widely different values with only minimal changes in the RDFs.[33] However, not all of the interactions perform well in larger systems. In Ref. 33, the virial pressure was used as an additional constrained quantity to obtain workable interactions, but in the present case also that fails to distinguish "good" interactions from "bad" ones: all the interactions discussed in this paper give essentially the same value for the virial pressure in the small system where they were determined. This demonstrates that it is non-trivial to evaluate the quality of effective interactions even if the target RDFs are well reproduced.

The above problems also reflect in the sensitivity of the potentials to the RDFs. However, although the differences between the different interactions may not even be visible to the eye without magnification, they are typically enough to distinguish the underlying interactions in the inversion: if a pairwise interaction is used to create the RDFs, exactly this pairwise interaction is recovered with algorithms such as IMC if no additional constraints are imposed. Because of the above issues, one needs to be careful when using RDFs determined from complex systems to determine just a single set of interactions, because there may be subtle effects from incomplete sampling that may result in large changes in the interactions.

Why do the interactions from some IMC approaches then perform worse than those from others? One reason seems to be that the "natural" internal energy values for the system have $E_{\text{fluct}} < 0$, which makes it impossible to generalize the interactions to a larger system (see Section III). The interactions that worked had $E_{\text{fluct}} \geq 0$, and did not suffer from this problem. However, not all of the interactions with $E_{\text{fluct}} \geq 0$ worked, and most notably so those from IMC approach A. A clue to this difference is provided by the observation that the problems become larger as cholesterol concentration increases. With increasing cholesterol concentration, the size of the simulation box becomes closer and closer to twice the cutoff. This means that the number of particle pairs that, on average, are beyond the cutoff is much smaller for the 30% system. This makes it possible to better and better compensate for incorrect values of ΔE and E_{fluct} with changes in the potential without other major changes (if all the particles were within the cutoff, ΔE and E_{fluct} parameters would not be needed, because their effect could be included in the potentials [see Appendix A]). Hence, one possible reason for the poor performance of approach A is its tendency to compensate any inaccuracies in the energy parameters with changes in the potential, which only works for the original system size. In approach B, the target counts are rescaled such that these inaccuracies are more or less invisible in the potential equation. However, this separation is not perfect, and it is possible that some of the variants suffer from similar problems as approach A. In particular, constraining E_{fluct} to zero in approach B(ii) does not solve the problem, although E_{fluct} no longer is negative.

It is possible that the underlying reason for the problems lies in incomplete sampling of the underlying MD simulations. As our model is more complex than the previous ones, the statistics on the RDFs are worse. The spline fitting used to smooth the RDFs can then induce small errors. Also, the one-particle densities calculated from the underlying MD simulations are not homogeneous, with the largest differences seen at higher cholesterol concentrations in the ordered/disordered tail densities. Further, autocorrelation functions of the chain states (calculated such that the two states correspond to 0 and 1) show that the timescale for converting a chain from one state to the other also depends strongly on the cholesterol concentration: for pure DPPC, the chains lose the memory of their previous state within 200 ps, while for the highest cholesterol concentrations this reaches 10 ns. In the highest concentrations, different chains also sample the different states quite differently, i.e., the fraction of time the chain is ordered depends on the chain much more than in pure DPPC. Together, these observations suggest that the complete phase space of the system may not be sampled for the highest concentrations. This can then alter the RDFs even to the point that the pair interactions that exactly reproduce them no longer exist. However, with the present data, it is difficult to ascertain if this is the case. Studies on simpler model systems where the sampling is not an issue should provide additional insight into

this problem and the suitability of RDF inversion for models with internal states in general, but is beyond the scope of the present article.

It is also interesting to note that the problems seen in Ref. 33 with the virial pressure follow the same pattern as those here: the "good" virial pressure and the virial pressure obtained without any constraint in the IMC differ more and more as the cholesterol concentration increases. Also in that case, the pure DPPC system worked quite well without the pressure constraint. It is thus possible that the origins of these problems are linked, although no concrete evidence can be provided.

Let us now briefly turn to the different algorithms used in inverting the RDFs. The most common of these are iterative Boltzmann inversion (IB)[14,53] and IMC;[13] for a review of other possible approaches, see Ref. 15. In this paper, we noted that IMC corresponds to Newton's method for maximizing the functional (4), while Soper has provided an argument for the global convergence of IB.[14] In our experience, IMC requires careful regularization during the initial steps of the iteration to remain stable. Hence, IB might be more suitable for use during the first steps in the optimization, while IMC could be used for the final optimization to take advantage of the quadratic convergence of Newton's method (assuming exact sampling of the required quantities). This is also supported by the observation that the convergence of the long-range part of the interaction is slow in IB, as measured by the virial pressure.[54] In contrast, our experience indicates that IMC converges to the "native" virial pressure within a few iterations if the RDFs are close to the target ones. As a final note on the convergence of the different approximations, we observe that the second derivative of the functional (4) is exactly the cross-correlation matrix that also appears in IMC. As such, it is positive definite, and hence the functional is convex. This rules out the possibility of local maxima, and ensures that an iterative method to refine the RDFs cannot get stuck in a local optimum.

RDF inversion is only one of the approaches for constructing CG models from detailed simulations. It is possible to devise other approaches based on the matching of forces[17,18] or potential energy.[42,55] All these different approaches lead to different effective interactions, each of the interactions performing better for some quantities than for others. In general, there is no single effective interaction that can reproduce all quantities of interest.[38,42] To some extent, this problem of *representability* is related to the fact that the same quantities can have a different meaning in the different models: for example, if we coarse-grain the water away from a solution of ions, the virial pressure in the CG model is actually the osmotic pressure in the detailed picture.[40,56] However, it may be difficult to find such relationships for all quantities, and also the approximation of H_{eff} by an effective pairwise potential results in correction terms in such equations that may be difficult to evaluate. Hence, care is needed when applying models with effective interactions, and the origins of the interactions should be kept in mind.[38]

In conclusion, we have extended the IMC method to handle models with internal degrees of freedom, and applied it to extend our earlier CG model for a DPPC/cholesterol bilayer to include the tail ordering in a discrete fashion. Our series of CG models of increasing complexity contributes to understanding of RDF inversion and its limitations as a CG technique in general, and this issue has been discussed in this paper at length. Despite the limitations, interesting results can be obtained already with the simple models we have constructed. Examples include the cholesterol-rich and cholesterol-poor domains at intermediate concentrations, tail density fluctuations at low cholesterol concentrations, and the effect of the internal degrees of freedom on the transferability properties. However, detailed analysis of the behavior of the models, e.g., in terms of domain sizes, still does not seem reasonable as there are uncertainties in the model construction process (such as the selection of the cutoff,[33] possibility of insufficient sampling, and definition of the ordered/disordered boundary) that can have a significant effect on the quantitative results.

It has proven to require a substantial amount of work to improve our original one-particle CG model constructed in Ref. 32, but each improvement has yielded some additional insight. The first three-particle model uncovered the tail density fluctuations at low cholesterol concentrations. The present work provided some additional information on ordering of the tails, but the main contribution in our view is the generalization of RDF inversion to models with internal states and demonstration of the problems that may result.

If the uncertainties in the model construction can be solved, systematic construction of the interactions can give substantial added value to the phenomenological models by Nielsen *et al.*. In particular, the interactions could then be linked to a particular lipid system, and different mixtures, e.g., with different sterols could be studied to understand the effects of the microscopic features on the general behavior. However, a substantial amount of work may be needed. Further, easy-to-interpret results may prove difficult to obtain, because transferability problems are expected, even when chain ordering is included as in the present work. Understanding simpler model systems might help alleviate the problems, and detailed studies in this respect are an interesting direction for future research.

It would also be interesting to study the extent of the problems seen in the present study in semi-atomistic models for lipid bilayers,[57,58] where the degree of coarse-graining is not as high. Again, systematic approach to the interactions could provide a link to atom-scale chemical detail. Comparison of the resulting interactions with the simple Lennard-Jones form used in, e.g., the MARTINI force field,[10,11] could also yield insight into the strengths and weaknesses of general (e.g., MARTINI) vs. specific (RDF and force matching) approaches to coarse-graining. Finally, comparing different systematic CG approaches, both on simple model systems and more realistic cases as studied here, could provide better understanding of their differences, as well as the source of the problems seen in the present study.

Acknowledgments

This work has been supported by the Academy of Finland through its Center of Excellence Program (T. M., I. V.), the Academy of Finland (I. V.), the National Graduate School in

Nanoscience (T. M.), and National Sciences and Engineering Research Council of Canada (M. K.). We would also like to thank the Finnish IT Center for Science and the HorseShoe (DCSC) supercluster computing facility at the University of Southern Denmark for computer resources.

Appendix A: Most General Hamiltonian with Internal States

In this appendix, we justify Eq. (8), i.e., the form of the Hamiltonian used for our CG model. In our model, the configurational space for the tails contains a discrete degree of freedom in addition to the particle positions, i.e., the tail is now described by $x_i = (\vec{r}_i, s_i)$, where $s_1 \in \{0, 1\}$. The uniqueness and existence proofs make no explicit reference to the form of the space,[29,30] so they apply equally well for the present case. Now, the pair potential(s) $V(x, x')$ can be thought of as consisting of several distinct pair potentials, each of which is a function of the positions only. There is one such pair potential for each combination of s values, and hence an equivalent description can be achieved for having a separate potential for each combination of internal states. However, the number of particles is no longer constant, and hence the two-particle density should be used instead of RDF as the target quantity: RDFs do not contain the information necessary to determine the occurrence probabilities of different s values. From now on, we will use the second picture unless explicitly noted, i.e., the pair potentials will be functions of positions only. Also, for clarity, we refer to *particle types* when all internal states are considered as a single type, and *particle kinds* when each internal state is separated.

The pairwise Hamiltonian is unique up to a constant. Without internal states, this statement is equivalent to each pair interaction being unique up to a constant. With internal states, this holds only for each $V(x, x')$, i.e., not for each individual pair potential if they are considered as functions of positions only. Hence, in the present case, we cannot require that each potential approach zero when $r \to \infty$.

When we introduce an approximation that each pair potential is zero beyond some cutoff, we limit the class of Hamiltonians available. In order to recover the full class of Hamiltonians on which the existence proof holds, we need to be able to add additive constants to each individual pair potential also beyond the cutoff. To calculate the minimum Hamiltonian that allows this, we calculate how the Hamiltonian changes when a different constant is added to each pair potential.

Let us consider a system with N_k particle kinds and $\frac{1}{2}N_k(N_k+1)$ pairwise potentials $V_{ab}(r)$. If we add constants C_{ab} to respective potentials, the Hamiltonian becomes

$$\begin{aligned} H - H_0 &= \frac{1}{2}\sum_{i\neq j} C_{k_i k_j} = \frac{1}{2}\sum_{k=1}^{N_k}\sum_{l=1}^{N_k}(n_k n_l - \delta_{kl} n_k)C_{kl} \\ &= -\frac{1}{2}\sum_{k=1}^{N_k} C_{kk} n_k + \frac{1}{2}\sum_{k=1}^{N_k}\sum_{l=1}^{N_k} C_{kl} n_k n_l, \end{aligned} \tag{A1}$$

where k_i is the kind of particle i, H_0 is the Hamiltonian before the addition, and n_k is the number of particles of kind k. Since the total number of particles of each type is constant (in a canonical ensemble), not all n_k are independent and we can choose $N_s = N_k - N_t$ independent n_k, where N_t is the number of particle types. In terms of these independent n_k numbers, the Hamiltonian becomes

$$H = H_0 + \sum_{i=1}^{N_s}\Delta E_i n_i + \sum_{i=1}^{N_s}\sum_{j=i}^{N_s}\Delta E_{ij}\delta n_i \delta n_j + C_0, \tag{A2}$$

where ΔE_i, ΔE_{ij}, and C_0 are linear combinations of the C_{ab}. Hence, if we take the original Hamiltonian as (8), we can now add ΔE_i and ΔE_{ij} to the corresponding E_i and E_{ij}, and the new Hamiltonian has the same functional form as the old one.

The last term in the Hamiltonian (8) can be written either as $\delta n_i \delta n_j$ or $n_i n_j$, and these forms can be interchanged using a suitable modification of E_i (also, any value of n_i^{ave} can be selected). However, the one with $\delta n_i \delta n_j$ [Eq. (8)] and $n_i^{ave} = \langle n_i \rangle_{\mathrm{MD}}$ is more suitable for the IMC iteration because the magnitude of $\delta n_i \delta n_j$ is much smaller than $n_i n_j$. The form selected for (8) also perhaps shows more clearly the physical significance of the parameters E_{ij} in controlling the magnitude of fluctuations.

Appendix B: Iteration Formulae for Internal Energies

In this appendix, we derive the iteration formulae (10) and (11) used for the internal energies in the IMC approach B.

We start by writing the partition function of the system and integrating out the positional degrees of freedom to arrive at

$$Z = \mathrm{Tr}\exp[-\beta(\sum_i E_i n_i + \sum_{i,j} E_{ij}\delta n_i \delta n_j + f(n))], \tag{B1}$$

where f is some unknown function of the vector n formed from the n_i, and the trace is over all possible combinations of n_i. $f(n)$ is essentially the free energy of such a combination of n_i (of which the two internal energy sums have been factored out). Switching to matrix notation and approximating f with a quadratic form around $\langle n \rangle$ as $f(n) = C_0 + (D^{(1)})^{\mathrm{T}} n + n^{\mathrm{T}} D^{(2)} n$, the partition function becomes

$$\begin{aligned} Z = \mathrm{Tr}\exp\{-\beta[\tilde{C}_0 + (D^{(1)} + E^{(1)} + 2E^{(2)} n^{ave})^{\mathrm{T}} n \\ + n^{\mathrm{T}}(D^{(2)} + E^{(2)})n]\}, \end{aligned} \tag{B2}$$

where $E^{(1)}$ and $E^{(2)}$ are a vector and a matrix formed from E_i and E_{ij}, respectively. Setting $v = D^{(1)} + E^{(1)} + 2E^{(2)} n^{ave}$ and $A = D^{(2)} + E^{(2)}$ and replacing the trace with an integral over the full real axis for each n_i, the partition function can be evaluated exactly as a Gaussian integral:

$$\ln Z = \frac{1}{4}\beta v^{\mathrm{T}} A^{-1} v + C, \tag{B3}$$

where C does not depend on $E^{(1)}$. Differentiating both sides with respect to E_i and canceling a common factor $-\beta$ gives

$$\langle n \rangle = -\frac{1}{2} A^{-1} v, \tag{B4}$$

and a second differentiation gives

$$\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle = \frac{1}{2\beta} [A^{-1}]_{ij}. \quad \text{(B5)}$$

Denoting $\sigma_{ij} = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$ and inverting both sides of (B5), we can then solve the change required in A to bring the fluctuations to the correct value as

$$\Delta(\sigma^{-1}) = 2\beta \Delta A. \quad \text{(B6)}$$

Substituting (B5) into (B4) gives

$$\langle n \rangle = -\beta \sigma v, \quad \text{(B7)}$$

which can be linearized to yield

$$\Delta \langle n \rangle = -\beta \sigma \Delta v - \beta (\Delta \sigma) v. \quad \text{(B8)}$$

Solving βv from (B7), substituting this into the above equation and rearranging gives

$$\beta \sigma \Delta v = -\Delta \langle n \rangle + \Delta \sigma (\sigma^{-1} \langle n \rangle). \quad \text{(B9)}$$

Assuming that $D^{(1)}$ and $D^{(2)}$ are constants, Eqs. (B6) and (B9) immediately give (10) and (11).

[1] P. S. Niemelä, M. T. Hyvönen, and I. Vattulainen, Biochim. Biophys. Acta **1788**, 122 (2008).
[2] S. J. Marrink, A. H. de Vries, and D. P. Tieleman, Biochim. Biophys. Acta **1788**, 149 (2008).
[3] G. A. Voth, ed., *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, Boca Raton, FL, 2008).
[4] J. W. Ponder and D. A. Case, in *Protein Simulations*, edited by V. Daggett (Academic Press, 2003), vol. 66 of *Advances in Protein Chemistry*, pp. 27–85.
[5] A. D. MacKerell, Jr., M. Feig, and C. L. Brooks, III, J. Comp. Chem. **25**, 1400 (2004).
[6] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. V. Gunsteren, J. Comp. Chem. **25**, 1656 (2004).
[7] D. P. Tieleman, J. L. MacCallum, W. L. Ash, C. Kandt, Z. Xu, and L. Monticelli, J. Phys.: Condens. Matter **18**, S1221 (2006).
[8] J. Sonne, M. O. Jensen, F. Y. Hansen, L. Hemmingsen, and G. H. Peters, Biophys. J. **92**, 4157 (2007).
[9] S. W. I. Siu, R. Vácha, P. Jungwirth, and R. A. Böckmann, J. Chem. Phys. **128**, 125103 (2008).
[10] S. J. Marrink, A. H. de Vries, and A. E. Mark, J. Phys. Chem. B **108**, 750 (2004).
[11] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, J. Phys. Chem. B **111**, 7812 (2007).
[12] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, J. Chem. Theory Comput. **4**, 819 (2008).
[13] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).
[14] A. K. Soper, Comp. Phys. **202**, 295 (1996).
[15] G. Tóth, J. Phys.: Condens. Matter **19**, 335220 (2007).
[16] F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[17] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, J. Chem. Phys. **120**, 10896 (2004).
[18] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).
[19] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, J. Chem. Phys. **128**, 244115 (2008).
[20] W. Noid, J.-W. Chu, G. Ayton, and G. Voth, J. Phys. Chem. B **111**, 4116 (2007).
[21] H. Gohlke and M. F. Thorpe, Biophys. J. **91**, 2115 (2006).
[22] A. Arkhipov, P. L. Freddolino, and K. Schulten, Structure **14**, 1767 (2006).
[23] A. Arkhipov, P. L. Freddolino, K. Imada, K. Namba, and K. Schulten, Biophys. J. **91**, 4589 (2006).
[24] D. Gfeller and P. D. L. Rios, Phys. Rev. Lett. **100**, 174104 (2008).
[25] Z. Zhang, L. Lu, W. G. Noid, V. Krishna, J. Pfaendtner, and G. A. Voth, Biophys. J. **95**, 5073 (2008).
[26] T. Murtola, M. Kupiainen, E. Falck, and I. Vattulainen, J. Chem. Phys. **126**, 054707 (2007).
[27] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, Phys. Chem. Chem. Phys. **11**, 1869 (2009).
[28] R. L. Henderson, Phys. Lett. A **49**, 197 (1974).
[29] J. T. Chayes, L. Chayes, and E. H. Lieb, Commun. Math. Phys. **93**, 57 (1984).
[30] J. T. Chayes and L. Chayes, J. Stat. Phys. **36**, 471 (1984).
[31] E. Falck, M. Patra, M. Karttunen, M. T. Hyvönen, and I. Vattulainen, Biophys. J. **87**, 1076 (2004).
[32] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainen, J. Chem. Phys. **121**, 9156 (2004).
[33] T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, J. Chem. Phys. **126**, 075101 (2007).
[34] Y.-W. Chiang, A. J. Costa-Filho, and J. H. Freed, J. Phys. Chem. B **111**, 11260 (2007).
[35] S. A. Pandit and H. L. Scott, Biochim. Biophys. Acta **1788**, 136 (2009).
[36] M. Nielsen, L. Miao, J. H. Ipsen, O. G. Mouritsen, and M. J. Zuckermann, Phys. Rev. E **54**, 6889 (1996).
[37] M. Nielsen, L. Miao, J. H. Ipsen, M. J. Zuckermann, and O. G. Mouritsen, Phys. Rev. E **59**, 5790 (1999).
[38] A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[39] F. H. Stillinger, H. Sakai, and S. Torquato, J. Chem. Phys. **117**, 288 (2002).
[40] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **55**, 5689 (1997).
[41] M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[42] M. E. Johnson, T. Head-Gordon, and A. A. Louis, J. Chem. Phys. **126**, 144509 (2007).
[43] S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site, J. Chem. Phys. **128**, 024503 (2008).
[44] B. Polyak, J. Math. Sci. **133**, 1513 (2006).
[45] Q. Shi and G. A. Voth, Biophys. J. **89**, 2385 (2005).
[46] T. Murtola, T. Róg, E. Falck, M. Karttunen, and I. Vattulainen, Phys. Rev. Lett. **97**, 238102 (2006).
[47] O. Berger, O. Edholm, and F. Jahnig, Biophys. J. **72**, 2002 (1997).
[48] M. Höltje, T. Förster, B. Brandt, T. Engels, W. von Rybinski, and H.-D. Höltje, Biochim. Biophys. Acta **1511**, 156 (2001).
[49] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Mod. **7**, 306 (2001).
[50] B. J. Thijsse, M. A. Hollanders, and J. Hendrikse, Computers In Physics **12**, 393 (1998).

[51] See EPAPS Document No. ... for additional figures. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.
[52] P. G. Bolhuis and A. A. Louis, Macromolecules **35**, 1860 (2002).
[53] W. Schommers, Phys. Lett. A **43**, 157 (1973).
[54] S. Jain, S. Garde, and S. K. Kumar, Ind. Eng. Chem. Res. **45**, 5614 (2006).
[55] G. Tóth, J. Phys.: Condens. Matter **19**, 335222 (2007).
[56] R. L. C. Akkermans and W. J. Briels, J. Chem. Phys. **114**, 1020 (2001).
[57] A. P. Lyubartsev, Eur. Biophys. J. **35**, 53 (2005).
[58] J. Elezgaray and M. Laguerre, Comp. Phys. Comm. **175**, 264 (2006).