Systematic coarse-graining of molecular models by the Newton inversion method.
— Source link ⧉

Alexander P. Lyubartsev, Alexander Mirzoev, Li Jun Chen, Aatto Laaksonen

Institutions: Stockholm University

Related papers:

- Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach.

- A multiscale coarse-graining method for biomolecular systems.

- Deriving effective mesoscale potentials from atomistic simulations.

- The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models.

- The MARTINI force field : Coarse grained model for biomolecular simulations

Faraday Discussions

This paper is published as part of Faraday Discussions volume 144: Multiscale Modelling of Soft Matter

Papers

Coarse-grained simulations of charge, current and flow in heterogeneous media
Benjamin Rotenberg, Ignacio Pagonabarraga and Daan Frenkel, *Faraday Discuss.*, 2010
DOI: 10.1039/b901553a

Multi-particle collision dynamics simulations of sedimenting colloidal dispersions in confinement
Adam Wysocki, C. Patrick Royall, Roland G. Winkler, Gerhard Gompper, Hajime Tanaka, Alfons van Blaaderen and Hartmut Löwen, *Faraday Discuss.*, 2010
DOI: 10.1039/b901640f

Can the isotropic-smectic transition of colloidal hard rods occur *via* nucleation and growth?
Alejandro Cuetos, Eduardo Sanz and Marjolein Dijkstra, *Faraday Discuss.*, 2010
DOI: 10.1039/b901594a

Multi-scale simulation of asphaltene aggregation and deposition in capillary flow
Edo S. Boek, Thomas F. Headen and Johan T. Padding, *Faraday Discuss.*, 2010
DOI: 10.1039/b902305b

The crossover from single file to Fickian diffusion
Jimaan Sané, Johan T. Padding and Ard A. Louis, *Faraday Discuss.*, 2010
DOI: 10.1039/b905378f

Mori–Zwanzig formalism as a practical computational tool
Carmen Hijón, Pep Español, Eric Vanden-Eijnden and Rafael Delgado-Buscalioni, *Faraday Discuss.*, 2010
DOI: 10.1039/b902479b

Discussion

General discussion
Faraday Discuss., 2010
DOI: 10.1039/b917709b

Papers

Hierarchical coarse-graining strategy for protein-membrane systems to access mesoscopic scales
Gary S. Ayton, Edward Lyman and Gregory A. Voth, *Faraday Discuss.*, 2010
DOI: 10.1039/b901996k

Towards an understanding of membrane-mediated protein–protein interactions
Marianna Yiannourakou, Luca Marsella, Frédérick de Meyer and Berend Smit, *Faraday Discuss.*, 2010
DOI: 10.1039/b902190f

Measuring excess free energies of self-assembled membrane structures
Yuki Norizoe, Kostas Ch. Daoulas and Marcus Müller, *Faraday Discuss.*, 2010
DOI: 10.1039/b901657k

Lateral pressure profiles in lipid monolayers
Svetlana Baoukina, Siewert J. Marrink and D. Peter Tieleman, *Faraday Discuss.*, 2010
DOI: 10.1039/b905647e

Concerted diffusion of lipids in raft-like membranes
Touko Apajalahti, Perttu Niemelä, Praveen Nedumpully Govindan, Markus S. Miettinen, Emppu Salonen, Siewert-Jan Marrink and Ilpo Vattulainen, *Faraday Discuss.*, 2010
DOI: 10.1039/b901487j

Membrane poration by antimicrobial peptides combining atomistic and coarse-grained descriptions
Andrzej J. Rzepiela, Durba Sengupta, Nicolae Goga and Siewert J. Marrink, *Faraday Discuss.*, 2010
DOI: 10.1039/b901615e

Discussion

General discussion
Faraday Discuss., 2010
DOI: 10.1039/b917710h

Concluding remarks

Concluding remarks
Herman J. C. Berendsen, *Faraday Discuss.*, 2010
DOI: 10.1039/b917077b

Systematic coarse-graining of molecular models by the Newton inversion method

Alexander Lyubartsev,* Alexander Mirzoev, LiJun Chen and Aatto Laaksonen*

Systematic construction of coarse-grained molecular models from detailed atomistic simulations, and even from *ab initio* simulations is discussed. Atomistic simulations are first performed to extract structural information about the system, which is then used to determine effective potentials for a coarse-grained model of the same system. The statistical-mechanical equations expressing the canonical properties in terms of potential parameters can be inverted and solved numerically according to the iterative Newton scheme. In our previous applications, known as the Inverse Monte Carlo, radial distribution functions were inverted to reconstruct pair potential, while in a more general approach the targets can be other canonical averages. We have considered several examples of coarse-graining; for the united atom water model we suggest an easy way to overcome the known problem of high pressure. Further, we have developed coarse-grained models for L- and D-prolines, dissolved here in an organic solvent (dimethylsulfoxide), keeping their enantiomeric properties from the corresponding all-atom proline model. Finally, we have revisited the previously developed coarse-grained lipid model based on an updated all-atomic force field. We use this model in large-scale meso-scale simulations demonstrating spontaneous formation of different structures, such as vesicles, micelles, and multi-lamellar structures, depending on thermodynamical conditions.

1   Simulating the real world

Three rather practical issues are always crucial in planning to perform computer experiments. All are equally important and therefore assigned collectively based on the available computing resources at the moment. We could consider them spanning an operational space with three "orthogonal" axes as bases:

1. System *size*
2. Motional *time* scale†
3. The *accuracy* of the model

Within this space we wish to perform the simulations in an optimal region to keep the computations feasible in one hand, while expecting to obtain reliable results in the other hand. In other words a reasonable compromise is always sought.

Possibly because of the incredibly fast development in hardware technology, very little effort was put in to improve the methods and the models during the first decades of computer simulations. Only during the last two decades have new simulation methods been introduced to stretch the *time* and *length* scales much further.

Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: sasha@physc.su.se; aatto@physc.su.se

† The number of moves in Monte Carlo (MC) simulations.

New schemes are also proposed to increase the *accuracy* by bringing MD and MC to the domains of quantum mechanics (QM) or to the first principles of physics. This allows simulations free of any empirical parameters (such as potential functions or molecular mechanical (MM) force fields) as input. The most widespread technique of this kind of method is the Car–Parrinello molecular dynamics.[1] Hybrid methods mixing QM and MM based schemes[2,3] are common tools today. The most interesting development may still be the schemes beyond atomistic resolution to model meso- and nano-scale systems and soft matters.[4,5] There are now reliable simulation methods available to treat a system at three levels of physical description (QM, MM and mesoscopic soft matter), where the accuracy is successively decreasing while allowing the system length and time scales to be increased. Examples of these are Car–Parrinello molecular dynamics, classical atomistic MD based on MM force fields, and dissipative particle dynamics (DPD). In common terminology, models beyond the atomistic resolution are a result of coarse-graining (CG). There is no unique way to do coarse-graining within an off-lattice framework. For heterogeneous systems like biological molecules normally fairly *ad hoc* approaches have been used to parameterize CG potentials,[6–8] while for homogeneous systems, like in materials design, finite element and grid-based models are commonly employed. In the case of biological systems a coarse-grained description of water molecules surrounding biomolecules represents a great challenge. Such simplifications may include implicit description of the solvent with the help of solvent-mediated potentials[9] or coarse-grained representation of solvent molecules.[8] The problem is, however, how we can accurately enough specify the interaction potentials for such coarse-grained models.

We will discuss here a straight-forward hierarchical multiscale modeling approach enabling us to link together different levels (physical models) of simulations by consecutively removing non-interesting degrees of freedom. The approach is based on a previously developed Inverse Monte Carlo scheme for reconstruction of pair potentials from known RDFs which can be computed from radial distribution functions. Previously this approach was illustrated for parameterization of effective potentials for water molecules and water-ion interactions starting from RDF obtained in *ab initio* CPMD simulations,[10,11] effective solvent mediated ion-ion[12] and ion-DNA[13] potentials as well as parameterization of a coarse-grained lipid model.[14] This paper reports further development in the field. It the next section, we describe a general approach enabling us to reconstruct parameters of the interaction potential (or the potential as a whole) from canonical averages computed from detailed simulations (or known from the experiment). As a simple example, we consider removing hydrogen atoms from a water molecule and derive effective potential for a single-site model, which provides realistic phase behavior of such a model. In the next section, we derive effective solvent-mediated potentials for united atom proline molecules in implicit DMSO solvent. Finally, we describe new updates on effective potentials for implicit solvent lipid models, and demonstrate applications of this in a description of behavior of lipid assemblies.

2   Hierarchical multiscale modelling approach

We begin from a quantum-mechanical first-principles simulation with the system size which we can afford. Here a Car–Parrinello molecular dynamics can be employed or any other *ab initio* molecular dynamics with quantum-mechanical computation of forces. Results of such *ab initio* simulations are used to parameterize an atomistic molecular mechanical force field which allows an increase of the size of the system and the simulation time by several orders of magnitude. On the next level, molecular dynamics simulation with *ab initio* derived force field will provide data for parameterization of the next level coarse-grained model for mesoscopic simulations. This idea of hierarchical, from the first principle modeling of matter can in principle be continued to further levels, but evidently the main problem to be solved in order

to make this idea feasible, is to find a way on how to use the results obtained from a more detailed, more fundamental model, to derive interaction parameters for simulation on the next, coarse-grained level. A general formalism on how to do this can be depicted as follows.

Let us consider transition from atomistic to coarse-grained level of molecular modeling. We divide all the degrees of freedom of the detailed system into important ones $\{R_I, I = 1, ..., N\}$, which we want to keep on the coarse-grained level, and unimportant degrees of freedom $\{r_i, i = 1, ..., n\}$ (which can be solvent coordinates, local details on macromolecular structure *etc.*). The "important" degrees of freedom may be real atom coordinates of the detailed system, or some combinations of them, for example centers of masses of some atomic groups. We can then define the effective N-body coarse-grained Hamiltonian (potential energy) which is also the N-body mean force potential and which is equal to the free energy of the removed degrees of freedom:[9,15]

$$\beta H^{\mathrm{PMF}}(R_1, ..., R_N) = -\int dr_1, ...dr_n \exp(-\beta H(R_1, ..., R_N, r_1, ..., r_n)) \tag{1}$$

where $\beta = 1/kT$ and $H(R_1, ..., R_N, r_1, ..., r_n)$ is Hamiltonian of the atomistic system. Simulation with Hamiltonian eqn (1) provides the same structural and thermodynamic properties of the coarse-grained system as those observed in atomistic simulation.‡[15] Practical simulations employing N-body potential are however unfeasible, that is why we need to approximate it by a more convenient expression, for example as a sum of effective pair potentials:

$$H^{\mathrm{PMF}}(R_1, ..., R_N) \approx \sum_{I<J} V^{\mathrm{eff}}_{IJ}(R_{IJ}) \tag{2}$$

with $R_{IJ} = |R_I - R_J|$. It is however not necessary to be limited by pair potentials, any practically usable expression can be given in eqn (2), for example angle or torsion potential terms for macromolecular CG models. Even orientational and three- and higher-order terms can be considered. However, from the computational point of view, we are interested just in pair-wise solutions: the very aim of coarse-graining is the computational speed-up, and use of many-body potentials would greatly hamper this goal.

The task of building a coarse-grained force field can be thus reformulated to find "as best as possible" approximation according to eqn (2). What is "the best approximation" can be determined however in different ways. One can determine a set of target properties which one may wish the CG model to keep. These properties can be either of microscopic character, as forces or instantaneous energies, or canonical averages as radial distribution functions, average energies, pressure. For example, minimizing the force difference coming from both sides of eqn (2) (weighted with the Boltzmann factor) is equivalent to the force matching method.[16–18] Similarly, comparing the energy, one can formulate a potential matching method.[19] If the target is a radial distribution function, the pair potential in eqn (2) can be reconstructed by the Inverse Monte Carlo method.[9] In principle, other properties of interest or any combination of them can be used for parameterization of effective potentials.

Assume our effective potentials are determined by a set of parameters $\{\lambda_i\}$, and the set of target properties (which we know from the atomistic simulations) is $\{A_j\}$. If we know the set of $\{\lambda_i\}$, we can always, simulating the system directly, compute average properties $\{\langle A_j \rangle\}$. The inverse problem, finding parameters $\{\lambda_i\}$ from averages $\{\langle A_j \rangle\}$, is less trivial. We can consider the relationship between $\{\lambda_i\}$ and $\{\langle A_j \rangle\}$ as a nonlinear multidimensional equation, and use the Newton inversion method (known also as Newton–Raphson method) to solve it iteratively.

---

‡ For thermodynamic properties, implicit dependence of H^{PMF} on temperature and density should be taken into account.

At each iteration of the Newton inversion, we need to compute the matrix of derivatives (Jacobian), showing how different potential parameters affect different averages. It is not difficult, using expression for the averages in the canonical ensemble, to arrive at the following formula:

$$\frac{\partial \langle A_j \rangle}{\partial \lambda_i} = -\beta \left(\left\langle \frac{\partial H}{\partial \lambda_i} A_j \right\rangle - \left\langle \frac{\partial H}{\partial \lambda_i} \right\rangle \langle A_j \rangle \right) \tag{3}$$

The Jacobian eqn (3) can be used to find parameters $\{\lambda_i\}$ corresponding to target values of $\{\langle A_j \rangle\}$ solving the system of linear equations:

$$\Delta \langle A_j \rangle = \sum_i \frac{\partial \langle A_j \rangle}{\partial \lambda_i} \Delta \lambda_i + O(\Delta \lambda^2) \tag{4}$$

where the second order corrections are neglected. One starts from some initial potential determined by a trial set of parameters $\{\lambda_i^{(0)}\}$, runs a simulation, and computes deviation of computed values $\{\langle A_j^{(0)} \rangle\}$ from the target values A^*:

$$\Delta \langle A_j \rangle = A^* - \{\langle A_j^{(0)} \rangle\} \tag{5}$$

as well as the Jacobian eqn (3). Then, from the system of linear eqn (4), one finds corrections to the potential parameters $\Delta \lambda_i$. The procedure is repeated with the new parameter set $\lambda_i^{(1)} = \lambda_i^{(0)} + \Delta \lambda_i$ until convergence is reached. If initial approximation $\{\lambda_i^{(0)}\}$ is poor, some regularization of the iterative procedure might be necessary, in which the difference in eqn (5) is multiplied by some factor between 0 and 1.

The procedure described above can be readily implemented if the number of potential parameters is equal to the number of target properties. If the number of properties exceeds the number of potential parameters, the problem can be solved in the variational sense, by finding the set of $\{\lambda_i\}$ which provides the least possible deviation of the computed properties from the target values. Note, that to get the solution of the variational problem one also needs computation of derivatives according to eqn (3).

An important example of the described above approach is the case when parameters $\{\lambda_i\}$ are the values of the pair potential in a number of points covering the whole range of distances, and the target properties are the values of RDF in the same set of points. Then the inverse problem is reformulated as finding the pair interaction potential which reconstructs the given radial distribution functions. It is known that the solution of such an inverse problem is unique,[20] a statement which holds even in the multi-component case for the relationship between a set of RDFs and the corresponding set of pair potentials. The equations given above become equivalent in this case to the inverse Monte Carlo approach described by us in previous publications.[9,21]

Another example where the Newton inversion for finding potential parameters can be applied is a problem of simultaneous fitting Lennard–Jones parameters σ and ε to two known thermodynamic observables, for example average energy and pressure. Such a problem was considered in ref. 22 using the weak coupling approach.

While the Newton inversion procedure was depicted above for the transition from the atomistic to CG level, it can work in the same way for the connection between *ab initio* and atomistic levels, with the only difference that the quantum-mechanical energy surface is used instead on N-body potential of mean force in eqn (2).

In the following, we shall concentrate on reconstruction of pair-wise potentials from RDFs which are computed in simulations on a more detailed level. In case of having several different CG sites in the system, indexes i, j in eqn (3–5) run both over all pairs of sites and distance points, and corresponding $\langle A_i \rangle$ represent a complete set of pair distribution functions. We include into the described scheme some other potential terms which are important for the description of

macromolecular structures: covalent bonds, angles, and torsion potentials. Then indexes i, j run even over possible (discretized) values of the bond lengths, angles and torsions, with λ_i values representing the bond, angle or torsion potentials and $\langle A_i \rangle$ representing the corresponding distributions over bond lengths, angles or torsions. In all cases, average values of $\langle A_i \rangle$ can be acquired from a detailed simulation of a small system.

There exist a few other approaches which can be used to invert RDFs as well as bond or angle distributions. Soper[23] introduced an empirical potential structure refinement method (also known as the "Iterative Boltzmann Inversion") in which the pair potential is corrected after each iteration according to the mean field approximation:

$$V^{(i+1)} = V^{(i)} + kT \ln \frac{g^{(i)}(r)}{g_{\text{ref}}(r)} \qquad (6)$$

Correction of potential according to eqn (6) is straightforward to implement, and such an approach was used in a number of studies.[24–27] In cases when several different types of coarse-grained sites, and correspondingly several different potentials are involved, cross-correlations between RDFs according to eqn (3–5) need to be taken into account in order to provide convergence.

In some cases it is possible to solve the inverse problem using numerical solutions of the liquid theory equations, for example the Hypernetted-Chain (HNC) approximation.[28] In the case of solvent-mediated potentials between ions, HNC solution was found to provide very accurate solutions of the inverse problem, coinciding with the results obtained by the inverse MC simulations.[29] We should also mention a few other works devoted to the inverse problem.[30–34]

3 Applications

3.1 A united atom water model

Water is the most common solvent. It is also the main substance in all biological organisms. Still, water is a difficult liquid in computer modeling and simulations. On the atomistic level, there exist a large number of models, most of them parameterized empirically, which more or less reasonably describe properties of water. Often explicit water can also be replaced by continuum models, from the generalized Born theory to effective solvent-mediated potentials. There is also a current interest to develop intermediate models in between atomistic and continuum models, presenting water as a liquid of particles, but without atomistic details. Example of such "coarse-grained" models may be the one-site model used in DPD simulations of water,[35] or the one used in the Martini force field.[8] These models are purely empirical. Here we present an alternative coarse-grained water model where the parameters are derived from atomistic simulations.

We consider here the simplest possible way to coarse-grain liquid water, presenting each molecule as a spherically symmetrical particle interacting with others by a distance-dependent pair potential. Such a potential can readily be obtained by inversion of the oxygen-oxygen RDF obtained in atomistic simulations of water (see Fig. 1). Corresponding RDF between molecular center-of-mass points can also be used. Similar water potentials have been presented in a number of publications recently.[36–38] Our water potential is displayed in Fig. 2. This potential reproduces perfectly the oxygen-oxygen water RDF from atomistic simulations (Fig. 1) if simulations are run exactly at the same density. The problem is, however, that the pressure in the CG system at normal density (1 g/cm³) is typically at the level 8000–9000 atm.[38] One can clearly see from the potential shape in Fig. 2, that such a potential is essentially repulsive, with only a very weak attractive minimum of about −0.5 kJ/mol at distances corresponding to second neighbors. One would
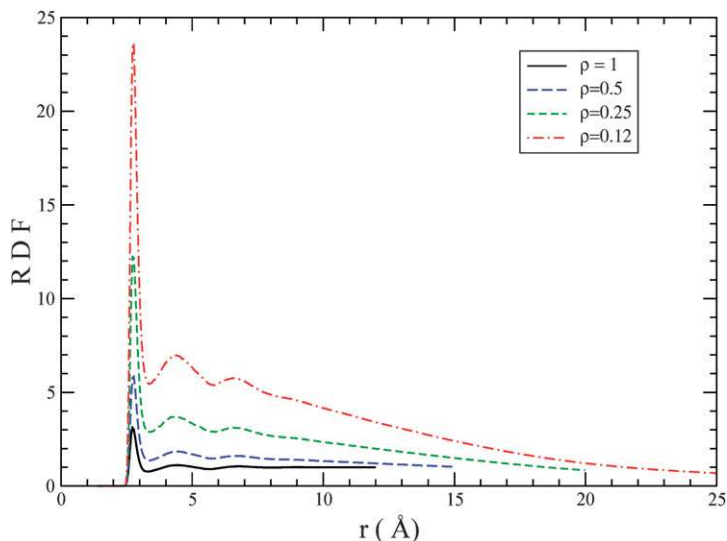
Fig. 1   Oxygen-oxygen RDF of water simulated at constant volume corresponding to densities 1, 0.5, 0.25 and 0.12 g/cm³.
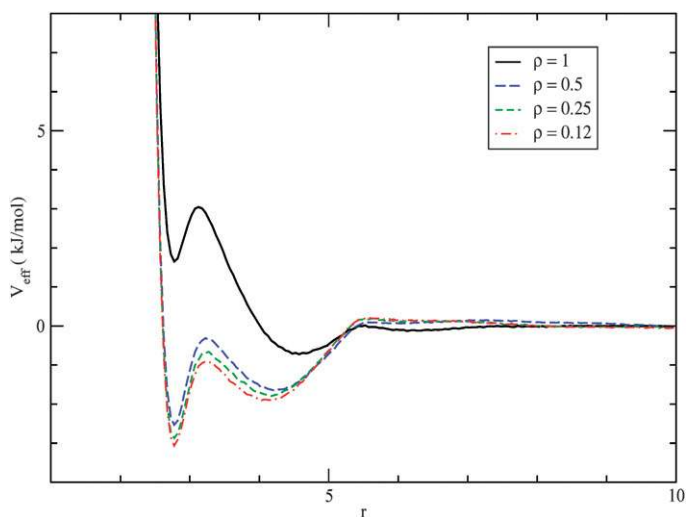
Fig. 2   Effective potentials for coarse-grained water obtained by inversion of RDFs shown in Fig. 1.

need to apply a high pressure to force the particles to approach each other to obtain the distances corresponding to those at normal liquid density. An undesirable consequence of this is that the system will expand immediately if there is space available, making it behave as a strongly compressed gas rather than a liquid. Simulations of various other properties within this model are questionable because of the very high internal pressure.

In a recent paper,[38] a method was suggested to build a "pressure consistent" one-site water model where the potential obtained by inversion of RDF from atomistic simulation was complemented by a slowly decreasing function, varying which one

can get correct pressure with only a minor effect on RDF. Here we describe an alternative approach to obtain a pressure-consistent effective potential avoiding the use of a fitting function. This example also illustrates problems related to concentration dependence of structure-derived effective potentials.

Assume that we run atomistic simulations of water using an SPC-like potential, in a volume several times larger than that corresponding to normal density. One can then expect two phases of water where molecules will condense in one place forming a liquid of nearly correct density, while the rest of the available volume will be mainly empty, representing the vapor phase. If we then use the RDF obtained from the phase-separated system to build the effective potential for the CG model, we can expect that the CG model would behave similarly (in order to maintain atomistic RDF), it will condense forming the same structure as the atomistic model. The important point is that if the system is phase separated into liquid and gas phases, the total pressure must be low. Thus we have reached two goals: we get a single-site model which correctly reproduces liquid structure and also provides a realistic pressure.

The idea described above was tested for the flexible SPC water model.[39] Simulations were run for 500 molecules at densities 1, 0.5, 0.25 and 0.12 g/cm³, and the corresponding oxygen-oxygen RDFs are displayed in Fig. 1. In all cases except in the first, a phase separation was observed. This can be seen from the snapshots, and also concluded from the fact that first and second RDF maxima increase nearly inversely proportional to the density. These RDFs were used as input to the inverse procedure, with the same number of molecules and the volume as the corresponding atomistic simulations. The results are shown in Fig. 2. One can clearly see that, for all phase separated systems, the effective potentials are very similar to each other, and they drop down substantially lower than the potential generated at bulk water density. The shape of these potentials is also similar to that obtained in paper[38] after addition of the term correcting the pressure.

The potentials generated from the RDFs simulated in low density systems, were also tested for the normal density 1 g/cm³. In these simulations, the potentials were cut after 10 Å (note that deviations of these potentials from zero are less than 0.15 kJ/mol already after 6 Å). The result is shown in Fig. 3. One can see
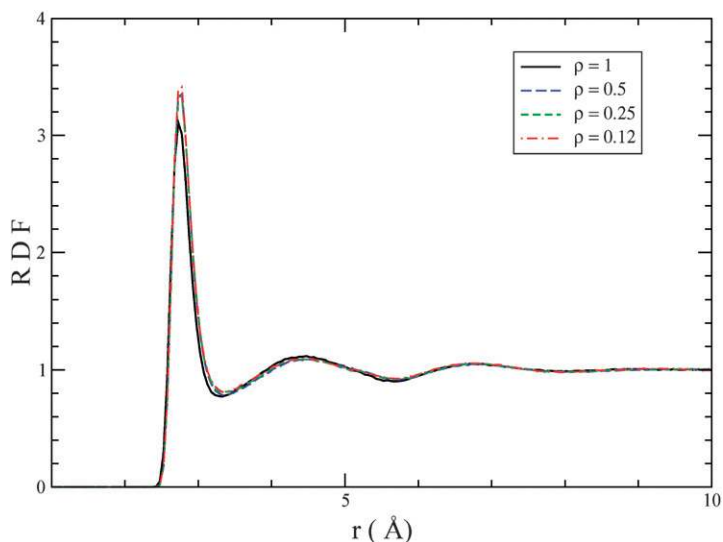

Fig. 3 RDFs of coarse-grained water presented by potentials given in Fig. 2 obtained at density 1 g/cm³.

Table 1 Computed pressure at density 1 g/cm³, and density corresponding to pressure 1 atm, for coarse-grained water potentials obtained from atomistic simulations of water at density ρ_0

ρ_0 (Atomistic simulations)	Pressure (bar)	Density (g/cm³) at pressure 1 atm
1	9050	<0.01
0.5	2370	0.71
0.25	60	0.99
0.12	−2070	1.17

that all these potentials give RDF practically overlapping with the result of atomistic simulation. One can note only a slight increase of the first RDF maximum. This means that changes in the effective potential, because of the initial lower densities, do not noticeably affect results after the correct density is restored.

A more important feature of the potentials obtained from lower density simulations is that they now provide more realistic pressures at the normal density. The values for pressure at density 1 g/cm³, for each of the four considered potentials, are given in Table 1. Although it is still different from the normal pressure 1 atm, one can see a very substantial improvement from the potential obtained by the direct inversion of bulk RDF. The system generated from atomistic simulation at 0.25 g/cm³ density provides practically correct pressure. The too low negative pressure in the case of potential generated in 0.12 g/cm³ simulation can be a result of the effect of the surface tension in the drop of water, which causes a higher pressure and thus higher density inside the drop simulated at the atomistic level, and which manifests in lowering pressure when the surface is no longer present at density 1 g/cm³. Observe that all the three potentials generated in lower density simulations, describe a liquid state. By a change of the density, the pressure can be brought to 1 atm. This feature makes such potentials suitable for constant-pressure simulations, as well as in the simulation of macromolecular assemblies in contact with water.

3.2 Coarse-grained simulations of an equimolar mixture of L- and D-proline in DMSO

A molecular point of view, based on simulations, can help to better understand the crystallization behavior and the relative stability of the racemic and enantiopure compounds. In this study we investigated L- and D-prolines (see Fig. 4) dissolved in dimethyl sulfoxide (DMSO) to study nucleation and crystallization behavior of the racemates in supersaturated solution. In this investigation we have focused on 0% enantiomeric excess solution of the two stereo isomers. This study is linked with asymmetric proline-mediated catalysis taking place in DMSO. In crystal and in solution, proline exists in a zwitterionic form becoming engaged in a network of strong hydrogen bonds due to the charged amine and carboxyl groups. Because hydrogen bonds in computer simulations are produced entirely by Coulombic interactions the simulation results are very sensitive to the used partial charges of the atoms. The electrostatic potentials of proline were calculated using the RESP method. The new RESP fitted charges, combined with the AMBER force field parameters, were imported into M.DynaMix[40] package to do the molecular dynamics simulations. Molecular dynamics simulations of proline and DMSO mixtures (a box containing 1000 DMSO molecules and 30 proline molecules, giving proline concentration about 0.4 M) were performed. Nose-Hoover thermostat and barostat were employed. A double time-step algorithm (0.2 and 2.0 fs for the short and long step, respectively) was implemented and the temperature is set at $T = 278\text{K}$.

We have removed all the hydrogen atoms and substituted the COO− group by a single site in our coarse-grained proline model. Totally, the coarse-grained proline was described by 6 sites, of which three carbon united atoms of the ring were
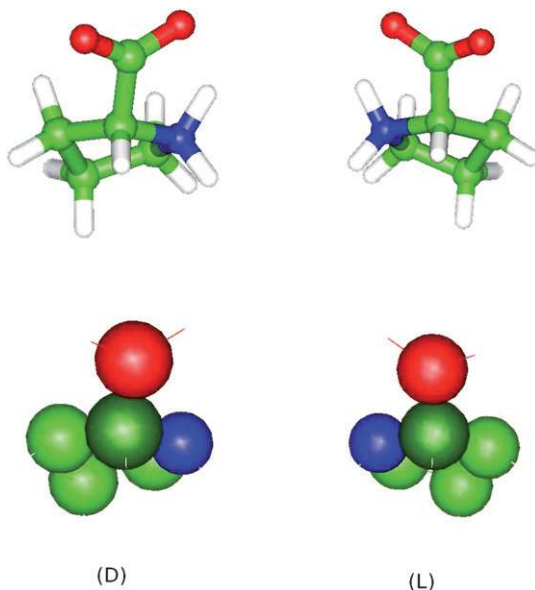
Fig. 4 Atomistic and coarse-grained representations of D- and L-prolines.

considered as equivalent (see Fig 4). Thus, we have four different coarse-grained sites on the proline molecule.

The RDFs between the coarse-grained sites of the prolines dissolved in DMSO were first determined in atomistic simulations. They were used to compute the effective solvent-mediated potentials between all the pairs of the coarse-grained proline sites. The enantiometric character of the proline molecules was kept by using intra-molecular bond potentials, which were also reproduced from the distribution of the corresponding bond lengths. Also, the charges at the nitrogen and COO sites were kept in order to maintain the zwitterionic character of the coarse-grained model. In the coarse-grained model, the charges interact *via* Coulombic potential with the dielectric permittivity of DMSO $\varepsilon = 46$.

In simulations without explicit DMSO solvent the derived effective potentials provide the very same structural properties of proline clusters as those observed in atomistic simulations. An example of such clustering is shown in Fig. 5, where a similar aggregate is quickly formed of the D- and L-prolines both in all-atom
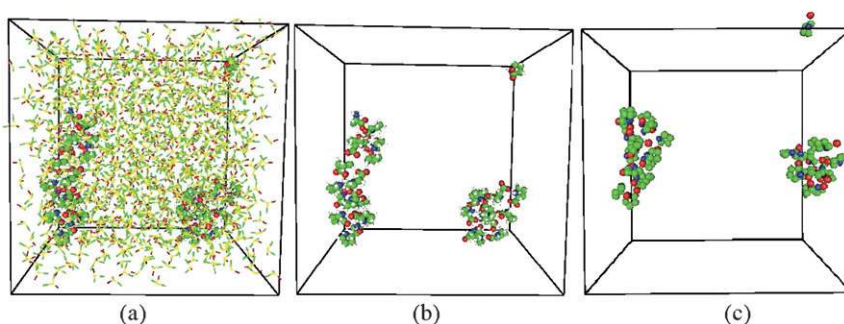

Fig. 5 Cluster of proline molecules obtained in molecular dynamics with explicit DMSO solvent: (a) DMSO are shown, (b) DMSO are not shown, and cluster formed in coarse-grained simulation with implicit solvent: (c).
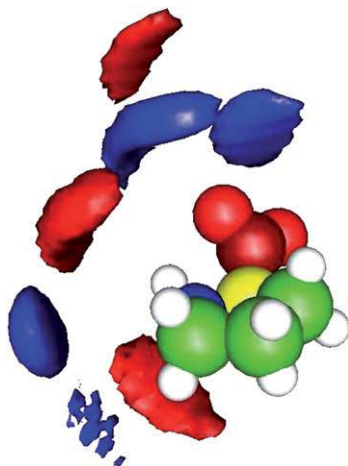
Fig. 6 Spatial distribution functions of CO (red) and NH (blue) CG-atoms around D-proline (red spheres - CO group, green - CH$_2$ groups, blue - NH group, sphere - CH group). Results are based on full atomistic MD calculation. CO- and NH- SDF shown in the figure has intensity value of 250 with maximum close to 2000.

simulations with explicit DMSO solvent (a and b) and in corresponding simulations using the six-site coarse-grained model with implicit DMSO solvent (c). Similar to the example of the one-site water model, we can suggest that effective potentials, derived from atomistic simulations of the "phase separated" system (that is, when the proline molecules were gathered to a cluster) would provide not only local structure of the proline cluster, but also a reasonable description of their osmotic pressure and affinity. To illustrate that the CG model contains the atomistic features of proline in implicit solvent we display the spatial distribution functions for the carbonyl and amine groups. In Fig. 6 these are calculated in all-atom simulations while in Fig. 7 they are calculated in CG simulations, correspondingly.

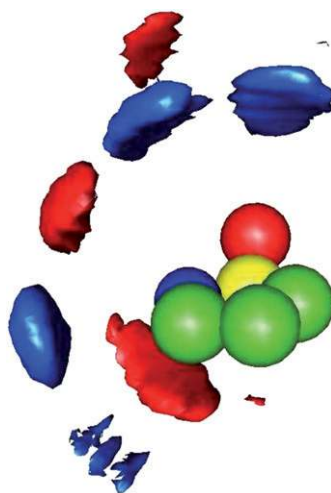
Fig. 7 Spatial distribution functions of CO (red) and NH (blue) CG-atoms around D-proline (red sphere CO group, green - CH$_2$ groups, blue - NH group, yellow - CH group). Results of MC calculation based on CG-potentials obtained using IMC procedure. CO-related SDF shown on the figure has intensity value of 220 with maximum of 1406 and NH-SDF has intensity value of 200 with maximum of 890.

3.3 Coarse-grained phospholipid model for bilayers and vesicles

Simulations of lipid membranes have attracted much attention during the last decade due to the fact that such membranes form outer shells of living cells. However, atomistic simulation of even a small piece of membrane consisting of about 100 lipids and surrounding water is a computational challenge, while many actual biophysical problems, such as studies of membrane mechanical properties, fusion, morphology, rafts formation, *etc.*, require consideration of substantially larger membrane fragments. For investigation of all these phenomena in molecular simulations, coarse-grained level of modeling provides practically the only possible choice. In recent years, coarse-grained lipid model based on Martini force field was used to model self-assembly of lipids into bilayers, micelles and vesicles.[8,41]

As another example of our multiscale modeling scheme, the effective solvent-mediated potentials for 10-sites CG model of DMPC lipid molecule (see Fig. 8) have been constructed.[14] The CG site-site RDFs were computed from atomistic molecular dynamics simulations of 16 lipids dissolved in water and described by the CHARMM-27 force field. In the subsequent development, the effective potentials derived in work[14] were re-parameterized after re-computation of CG site-site RDFs according to the recent modification of the CHARMM27 force field described in ref. 42. This modification of the CHARMM force field has been done with the primary aim to improve agreement with experiments for atomistic simulations of lipid bilayers, and to reproduce correctly the average area per lipid in particular. The principle point of this modification was recalculation of the atomic charges on the lipid head group by the *ab initio* Hartree–Fock method. The reparametrized DMPC model was simulated in 100 ns atomistic simulations, at a 1 : 100 DMPC/water molar ratio.

Similar to the described above cases of water simulated at low density, or proline molecules in DMSO, the simulated lipids were gathered in an unordered cluster of irregular structure, from which the site-site RDFs were determined. They were used in the inverse procedure to generate the effective potentials. The described above modification of the atomistic lipid potentials had a rather limited effect on the computed effective potential and they appear rather similar to those given in a previous paper.[14]

Large-scale Langevin molecular dynamics simulations of the CG lipid model, with the number of lipids in the range of 400–5000 and the system size of 200–500 Å, have been carried out using an ESPResSo package.[43] First, it was demonstrated that the coarse-grained model provides the same structure of a plane bilayer as the atomistic model. Then a number of other simulations were performed. It was shown that depending on conditions, lipids being initially thrown randomly in the simulation cell, organize themselves in different structures. If the number of lipids is small (less that 1000) the prevailing resulting structure is a bicell (a piece of bilayer of
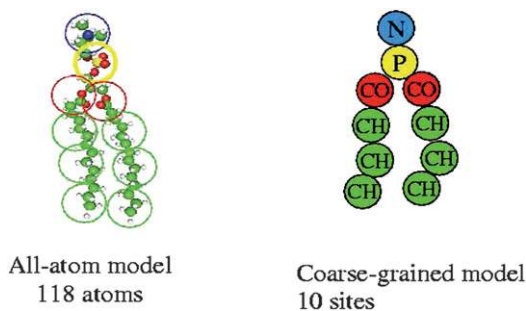

All-atom model
118 atoms

Coarse-grained model
10 sites

Fig. 8 Atomistic and coarse-grained representations of DMPC. Lipid-equivalent sites are shown in the same color.
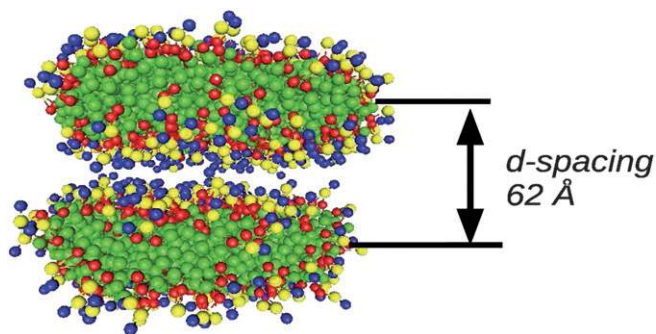
Fig. 9 Two discoid bicells, one over another, formed from an initially random state of 400 CG lipids in a 200 Å box. The choline groups of lipids are shown in blue, phosphates yellow, ester groups red and hydrocarbon tails green.

discoid shape). As an intermediate structure, several bicells in a stack one above the other (Fig. 9) were observed. It is quite remarkable that the average distance between the middle planes of two layers was about 62 Å, in perfect agreement with the experimental *d*-spacing of a multi-lamellar DMPC bilayer stack.[44]

A larger number of lipids (with 1500 lipids in a 400 Å box and 3500 lipids in a 500 Å box) was found to spontaneously form a spherical vesicle. The observed vesicle formation took place on a 100 ns time scale. Note however that in these simulations, the Langevin friction parameter γ was set to the very low value 0.001. With such a value of the parameter, the equations of motion are close to the Newtonian ones, while the role of the friction parameter is only to control the temperature. The dynamics in this case is artificially accelerated, which allows reaching equilibrium faster. It might be possible to adjust the Langevin friction parameter to reproduce lipid diffusion observed in atomistic simulation and thus also reproduce real-time description of lipid dynamics.

Another simulation has been performed at higher lipid concentration corresponding to 5000 CG lipids in a smaller periodic box, 300 Å. The size of the system does not allow formation of a single bicell or vesicle. The starting structure again was random. Initially, during the first 100 ns, a tendency to form multi-lamellar bilayer structures was observed, see Fig. 10a. This structure has been evolving further, and finally, after about 300 ns, a multi-lamellar vesicle consisting of almost ideally
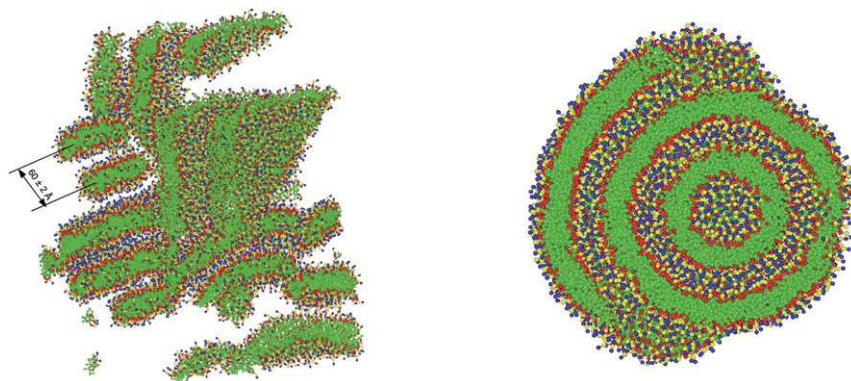

Fig. 10 Formation of multi-lamellar structures in a Langevin dynamics simulation of 5000 coarse-grained DMPC lipids in a periodic cubic box of 300 Å. Snapshots after 50 ns (left) and 300 ns (right; cut-plane through the middle of the vesicle is shown).

spherical layers was formed, see Fig. 10b. The third outer layer is not completely filled because there were not enough lipids to fill it. This structure was stable during the remaining 200 ns of simulation.

Again, it should be noticed that both in the case of bicells (Fig. 9) and multi-lamellar structures (Fig. 10), the distance between the mid-planes of neighboring layers is 60–62 Å, which means that there is space of about 20 Å between the surfaces of neighboring layers, filled by water. Also, the multi-lamellar vesicle in Fig. 10b occupies about one third of the total cell volume, which means in fact phase separation on a dense multi-lamellar phase and empty space representing bulk water. It is known that in real life solutions, lipids undergo such phase separation, on pure water phase and concentrated, lamellar phase with about 30 water molecules per lipid, which corresponds to about 20 Å water layers between lipid bilayers. Although our CG model does not include explicit water, it perfectly reproduces experimental phase behavior of phospholipid solutions, including a proper molar ratio of lipids in the lamellar phase. Thus the coarse-grained model, developed exclusively from all-atomic simulation data, reproduces well all the basic features of lipids in water solution.

4 Conclusions

A hierarchical true multiscale modelling approach presented here links together three levels of molecular modelling: *ab initio* molecular dynamics, classical molecular dynamics and meso-scale simulations. The method providing the link between these levels is the Newton inversion which in fact is equivalent to the inverse Monte Carlo approach previously developed by us. On the *ab initio* level, Car–Parrinello simulations, or alternatively, a highly efficient and accurate *ab initio* tight-binding (AITB) scheme, recently developed by us,[45] can be used in the future. The described methodology provides a consistent scheme to build molecular models for different scales without the need of empirical fitting of parameters. Some elements of this scheme were also demonstrated. Of course, there is still a very long way to go to define properties of molecules or materials exclusively *in silico*, and some tuning of the models against available experimental data is always an option. Also, transferability of the coarse-grained potentials needs to be checked in every case.

An interesting feature of the inverse procedure, which we demonstrated in the examples, is that when the detailed system was forced into the condition of phase separation, the derived "coarse-grained" system tried to reproduce the same phase separated structure, which implied the maintaining of some basic thermodynamic properties.

We hope that the suggested approach would increase the fraction of "*ab initio*" derived features in molecular models in expense of "*ad hoc*" or "empirically fitted" ones, which would enhance reliability of molecular simulations, increase their predictive power and open possibilities to address new, truly "large-scale" problems which are not yet considered in the molecular simulation domain.

Acknowledgements

References

1 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.
2 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
3 M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700.
4 P. Espanol, *Phys. Rev. E*, 1995, **52**, 1734.
5 R. Benzi, S. Succi and M. Vergassola, *Phys. Rep.*, 1992, **222**, 145.

6 R. D. Groot, *Langmuir*, 2000, **16**, 7493.
7 H. Noguchi and M. P. Takasu, *J. Chem. Phys.*, 2001, **115**, 9547.
8 S. J. Marrink, A. H. de Vries and A. E. Mark, *J. Phys. Chem. B*, 2004, **108**, 750.
9 A. P. Lyubartsev and Laaksonen, *Phys. Rev. E*, 1995, **52**, 3730.
10 A. P. Lyubartsev and A. Laaksonen, *Chem. Phys. Lett.*, 2000, **325**, 15.
11 A. P. Lyubartsev, K. Laasonen and A. Laaksonen, *J. Chem. Phys.*, 2001, **114**, 3120.
12 A. P. Lyubartsev and Laaksonen, *Phys. Rev. E*, 1997, **55**, 5689.
13 A. P. Lyubartsev and A. Laaksonen, *J. Chem. Phys.*, 1999, **111**, 11207.
14 A. P. Lyubartsev, *Eur. Biophys. J.*, 2005, **35**, 53.
15 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.*, 2008, **128**, 244114.
16 F. Ercolessi and J. Adams, *Europhys. Lett.*, 1994, **26**, 583.
17 S. Izvekov, M. Parrinello, C. J. Burnham and G. A. Voth, *J. Chem. Phys.*, 2004, **120**, 10896.
18 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469.
19 G. Toth, *J. Phys.: Condens. Matter*, 2007, **19**, 335222.
20 R. L. Henderson, *Phys. Lett. A*, 1974, **49**, 197.
21 A. P. Lyubartsev and A. Laaksonen, On the reduction of molecular degrees of freedom in computer simulations in *Novel Methods in Soft Matter Simulations*, ed. M. Karttunen, I. Vattulainen, A. Lukkarinen, Springer, Berlin, 2004, pp. 219–244.
22 S. L. Njo, W. F. van Gunsteren and F. Müller-Plathe, *J. Chem. Phys.*, 1995, **102**, 6199.
23 A. K. Soper, *Chem. Phys.*, 1996, **202**, 295.
24 J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay and M. L. Klein, *J. Phys. Chem. B*, 2001, **105**, 4464.
25 D. Reith, M. Putz and F. Muller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624.
26 D. Reith, M. Putz and F. Muller-Plathe, *J. Phys. Chem. B*, 2005, **109**, 2469.
27 V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt and K. Kremer, *Macromolecules*, 2006, **39**, 6708.
28 L. Reatto, D. Levesque and J. J. Weis, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1986, **33**, 3451.
29 A. P. Lyubartsev and S. Marcelia, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **65**, 041202.
30 M. Ostheimer and H. Bertagnolli, *Mol. Simul.*, 1989, **3**, 227.
31 Y. Rosenfeld and G. Kahl, *J. Phys.: Condens. Matter*, 1997, **9**, L89.
32 G. Toth and A. Baranyai, *J. Mol. Liq.*, 2000, **85**, 3.
33 R. L. C. Akkermans and W. L. Briels, *J. Chem. Phys.*, 2001, **114**, 1020.
34 N. G. Almarza and E. Lomba, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2003, **68**, 011202.
35 R. D. Groot, Applications of Dissipative particle Dynamics in *Novel Methods in Soft Matter Simulations*, ed. M. Karttunen, I. Vattulainen, A. Lukkarinen, Springer, Berlin, 2004, pp. 5–37.
36 A. Eriksson, M. N. Jacobi, J. Nyström and K. Tunstrøm, *J. Chem. Phys.*, 2008, **129**, 024106.
37 M. Praprotnik, L. D. Site and K. Kremer, *Annu. Rev. Phys. Chem.*, 2008, **59**, 545.
38 H. Wang, C. Junghans and K. Kremer, *Eur. Phys. J. E*, 2009, **28**, 221.
39 K. Toukan and A. Rahman, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1985, **31**, 2643.
40 A. P. Lyubartsev and A. Laaksonen, *Comput. Phys. Commun.*, 2000, **128**, 565.
41 S. J. Marrink and A. E. Mark, *J. Am. Chem. Soc.*, 2003, **125**, 15233–15242.
42 C.-J. Högberg, A. M. Nikitin and A. P. Lyubartsev, *J. Comput. Chem.*, 2008, **29**, 2359.
43 H. Limbach, A. Arnold, B. A. Mann and C. Holm, *Comput. Phys. Commun.*, 2006, **174**, 704.
44 N. Kucerka, Y. Liu, N. Chu, H. I. Petrache, S. Tristram-Nagle and J. F. Nagle, *Biophys. J.*, 2005, **88**, 2626.
45 Y. Tu, S. P. Jacobsson and A. Laaksonen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **74**, 205104.