Systematic Implicit Solvent Coarse Graining of Dimyristoylphosphatidylcholine Lipids

Alexander Mirzoev* and Alexander P. Lyubartsev

We have used systematic structure-based coarse graining to derive effective site–site potentials for a 10-site coarse-grained dimyristoylphosphatidylcholine (DMPC) lipid model and investigated their state point dependence. The potentials provide for the coarse-grained model the same site–site radial distribution functions, bond and angle distributions as those computed in atomistic simulations carried out at four different lipid–water molar ratios. It was shown that there is a non-negligible dependence of the effective potentials on the concentration at which they were generated, which is also manifested in the properties of the lipid bilayers simulated using these potentials. Thus, effective potentials computed at low lipid concentration favor to more condensed and ordered structure of the bilayer with lower average area per lipid, while potentials obtained at higher lipid concentrations provide more fluid-like structure. The best agreement with the reference data and experiment was achieved using the set of potentials derived from atomistic simulations at 1:30 lipid:water molar ratio providing fully saturated hydration of DMPC lipids. Despite theoretical limitations of pairwise coarse-grained potentials expressed in their state point dependence, all the resulting potentials provide a stable bilayer structure with correct partitioning of different lipid groups across the bilayer as well as acceptable values of the average lipid area, compressibility and orientational ordering. In addition to bilayer simulations, the model has proven its robustness in modeling of self-aggregation of lipids from randomly dispersed solution to ordered bilayer structures, bicelles, and vesicles. © 2014 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23610

Introduction

Modeling of lipid bilayers and other lipid self-assembled structures is one of the important areas of large-scale biomolecular and soft matter simulations.[1–4] Cell membranes play an especially important role in biology, serving as shells of eucaryotic cells and separating organelles inside cells. They consist of many components: lipids, proteins, sterols, carbohydrates, and other associated molecules. Phospholipids are the most common type of lipids present in the cell membrane, being the main building material of the membrane. Due to the hydrophobic nature of phospholipid's tails, they prefer to self-aggregate in aqueous solution and form a number of different structures such as micelles, inverse micelles, flat bicelles, spherical vesicles, and flat bilayers which serve as a base for biological membranes. The typical length and time scale required to model various lipid aggregations and self-assembled structures have an order of 100 nm^{-1} μm and 100 μs^{-1} ms, respectively. Such sizes are far beyond capacities of atomistic molecular dynamics (MD), which in the present state of art can cover systems up to 10 nm in size and 1 μs in time. Even an optimistic estimate of the development in computing facilities says that it will take more than 30 years for the atomic MD to reach the required scales.[5] The problem can be overcome by neglecting unimportant details in the system description for the price of computation speedup. The idea of uniting atoms into coarse-grain (CG) sites (coarse graining) to reduce the number of degrees of freedom turned out to be very fruitful in application to lipid bilayers. A number of various models were reported during the past two decades.[6–27] In general, they can be divided into two groups, so called top-down and bottom-up methodologies.

The first one requires the CG-model to reproduce experimentally measurable macroscopic properties of the system without considering much of microscopic details.[6–16] Typically, in this group of models lipids are represented as rigid rods or bead-spring molecules, with a soft core repulsion or Lennard–Jones potential for the nonbonded interactions. Electrostatic interactions are often omitted. Such models are extremely effective, allowing one to consider systems of few thousand lipids. Drouffe et al.[6] in their pioneering work simulated self-aggregation of about 4000 lipids into a vesicle. At that time, such a large system was affordable only due to the crude level of coarse graining: each particle represented a pair of lipids. A similar detalization was suggested later by Brannigan and Brown.[11] Later models represent a lipid molecule with more details: three beads connected in a rigid[8,9] or flexible[15] rod, or a polymer-like five-bead chain with constrained[12] or elastic bonds.[13,14] More advanced models take into account chemical structure of phospholipids, that is, two hydrocarbon tails introduced explicitly.[7,10,16] The probably most known and widely used empirically parameterized model is formulated in

A. Mirzoev A. P. Lyubartsev
Division of Physical Chemistry, Department of Materials and Environmental Chemistry, Stockholm University, Stockholm SE-10691, Sweden
E-mail: alexander.lyubartsev@mmk.su.se
Contract grant sponsor: Swedish Research Council (Vetenskapsrådet); Contract grant number: 70525601
© 2014 Wiley Periodicals, Inc.

terms of Martini force field, with coarse-grained beads representing groups of about four heavy atoms and interactions parameterized using experimentally known thermodynamic data.[21,22]

Methods of the second group[17–27] derive parameters of the model aiming to reproduce some important properties of an underlying reference (typically atomistic) system taking into account chemical specificity. Such properties can be effective mean forces acting on beads of the coarse-grained system[23–25] or structural properties, that is, radial distribution function (RDF) of the beads, bond length distributions, and so forth.[17–20,26,27]

Alongside with coarse graining of lipid molecules, description of solvent also requires simplification which is usually introduced either using "coarse-grained" solvent or considering it implicitly via effective solvent mediated potentials acting between lipid molecules. Among explicit-solvent models, we can mention works,[17,21,23] having one, three, or four molecules per solvent coarse-grained particle, respectively. In early solvent-free studies,[6,8] water was introduced as additional multibody interaction between lipids, which is expensive in terms of computation time. Later, implicit water was described in terms of pairwise interactions[15] defined within a top-down approach. Other implicit solvent models were introduced within bottom-up approach.[19],[24–27] In paper by Lu and Voth,[25] a bilayer composed of a mixture of CG lipids with the interaction potentials defined by the force matching was simulated at constant-volume conditions. A number of other implicit solvent CG models simulated at other conditions faced difficulties with the hydrophobic behavior of the model, which were corrected by introduction of additional interactions. Thus, Murtola et al.[19] introduced a surface tension constrain to the model, whereas Wang and Deserno[27] brought additional cohesion force to the model to stabilize the bilayer. In the study of Izvekov and Voth,[24] CG-lipids self-aggregated to bicelles having exposed lipid tails on the boundary, which might indicate luck of tail-hydrophobicity in the model and which was corrected by an additional constraint for the partial virial associated with CG sites. It is worth to note, that in the most of mentioned bottom-up studies, a preassembled bilayer was chosen as a reference atomistic system, which could be the reason for the reported problems with hydrophobic interactions.

In the present work, we report recent development of a bottom-up derived coarse-grained solvent-free model of dimyristoylphosphatidylcholine (DMPC) lipid which was introduced previously.[26] It is known from other examples that a general feature of bottom-up effective potentials is their state dependence.[28] Transferability of the bottom-up derived CG potentials with respect to the system composition and concentration of components has been investigated, with varying conclusions, in a number of studies for simple polymer models,[29,30] mixtures of small molecules,[31–33] and ion solutions.[34–36] Transferability studies for more complex molecules such as lipids with several CG sites of different types and many different intramolecular interactions are less frequent though investigation of structure-based CG peptide models can be mentioned.[37] Also, the cited transferability studies were typically addressed to the question: how well the effective potentials obtained at certain thermodynamic conditions can be used to describe other thermodynamic conditions? Here, we address to a related but slightly different question: how the effective potentials depend on the thermodynamic conditions of the atomistic simulations from which they were derived?

Table 1. Molecular composition of the reference systems.

Label	N_{mol} (DMPC)	N_{mol} (water)	Ratio DMPC:water
DMPC16–100	16	1600	1:100
DMPC30–50	30	1500	1:50
DMPC60-30	60	1800	1:30
DMPC60-20	60	1200	1:20

Therefore, in this work, special attention was paid to investigation of the concentration dependence of the reference system, that is, how the derived effective potentials and subsequent results of CG-simulations depend on the lipid/solvent ratio in the reference atomistic simulations. Four compositions were studied: 16 DMPC + 1600 H_2O, 30 DMPC + 1500 H_2O, 60 DMPC + 1800 H_2O, and 60 DMPC + 1200 H_2O, and a separate set of effective potentials was generated for each composition. Additionally, the original model,[26] which had only pairwise bonds, was augmented with angle bending interactions (shown on Fig. 1-right with dashes). To investigate the effect of the angle bending, we have used both models and the results are referred to as "angles" and "no-angles." Effective interactions between beads in the model were derived from fully atomistic simulations by the inverse Monte–Carlo (IMC) method[38] implemented in our recently released software package MagiC.[39] The potentials are derived in a tabulated form, which does not limit interactions to a specific functional form. The model that was found to fit both atomistic simulation data and experiment best, was further used for mesoscale simulations, showing clear presence of hydrophobicity of lipid tails leading to self-assembly of lipids into both flat structures (bilayer, bicell) and spherical structures (vesicles).

Method and Model

Atomistic modeling

Reference atomistic MD simulations were performed for four compositions listed in Table 1. In each case, a uniform solution of lipid and water molecules placed and oriented randomly was taken as a starting geometry. All other parameters were kept the same within all four systems.

The simulations (after initial short pre-equilibration) were performed in NPT-ensemble, with the temperature and pressure fixed to 303 K and 1 atm, respectively, by the use of Nosé–Hoover thermostat[40,41] and Parrinello–Rahman barostat.[42,43] Periodic boundary conditions with a cubic periodic cell were used. Differently from previous work,[26] a modified CHARMM force field[44] was used for lipid description together with rigid TIP3P[45] water model. The cited force field improves agreement with experiment for a number of important

properties of hydrated lipid bilayer such as average area per lipid at zero tension and structure factor compared to the original CHARMM27 force field.[44] In the present simulations, all the bonds with hydrogen atoms were constrained using LINCS algorithm,[46] thus, the high frequency bond vibrations were removed, which allowed us to use an integration time step of 1 fs. The electrostatic interactions were treated by particle-mesh Ewald summation method with real-space cutoff of 13 Å. The same cutoff was used for the Lennard–Jones short-range interactions. Each atomistic MD simulation resulted in a 400 ns trajectory of which first 100 ns were disregarded as equilibration. The MD-simulations were performed using the GROMACS 4.5.3[47] software.

In addition to the mentioned four systems, a MD simulation of a preassembled bilayer made of 98 DMPC lipids hydrated by 2800 water molecules has been carried out. In this case, a semi-isotropic barostat was used, which decoupled box size normal to the membrane from two other box dimensions. All other parameters of the simulation as well as the force field were kept the same as in the simulations described above. Results of this 100-ns-long simulation were used as a reference for comparison with the results of bilayer simulations within the coarse-grained models.

Coarse-grained model

Each DMPC molecule consisting of 118 atoms was mapped onto a CG model consisting of 10 beads as shown on Figure 1a. Polar head of the lipid is represented by two charged groups: choline (blue) and phosphate (yellow). The phosphate group is connected to two ester groups (red) which are also bound to each other, and which are further connected to lipid tails represented by two hydrocarbon chains made of three beads (cyan) each. The position of each bead is defined as a center of masses of all the atoms included in the bead, and the total charge of the atomic group is assigned to the charge of the CG bead. Thus, the highest partial charges lie on the N and P groups, with CO-groups slightly charged and CH-groups completely neutral (Fig. 1b). To keep the number of different potentials reasonably low, we did not distinguish terminal CH group from middle-chain CH groups (and prescribing them the same potentials), and also considered both ester groups as identical species in bonded and short-range interactions, however, they had unequal charges. In average, one CG bead substituted four to five heavy atoms, providing same level of resolution as other similar CG lipid models.[17,21,24,27]

Since we defined four bead types, 10 different intermolecular pairwise interactions need to be defined in the CG-model. Besides that, "covalent" pair bonds and angular bonds were defined, which are shown in Figure 1b with solid and dashed lines, respectively. We assume also that bonds between same pair of species can be considered identical, thus we have five different bonds: N—P, P—CO, CO—CO, CH—CO, CH—CH; and five angles: N—P—CO, P—CO—CH, CO—CH—CH, CO—CO—CH, CH—CH—CH. All together, this gave us 20 different effective potentials to determine, of which 10 are bonded and 10 are nonbonded (Fig. 1b). In "no-angle" CG models only five pair

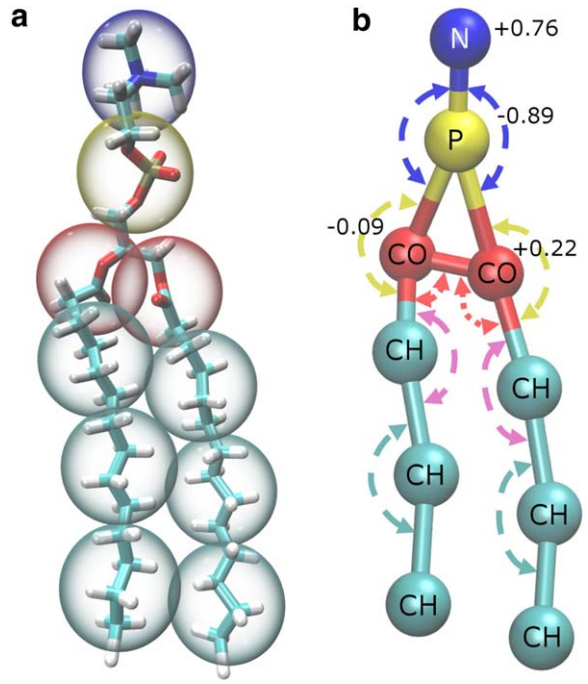

Figure 1. (a) Coarse-grained representation of a DMPC molecule. Blue bead—choline, yellow bead—phosphate group, red beads—ester groups, and cyan beads—hydrocarbon tails. Beads of the same color belong to same bead type. (b) Intramolecular bonds introduced in the CG model. Solid lines denote covalent bonds, dashed arrows denote angle bending interactions. Bonds of the same color are assumed to be identical. Positive and negative numbers indicate partial charges of the head-beads. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

bonds were considered. All bead pairs connected by simple bonds and, in case of "angle" models, bending angles were excluded from nonbonded interactions (both short range and electrostatic) as well as from computations of the RDFs.

Once the CG lipid model (e.g., bead-mapping scheme, bead types and bonds) was defined, each reference atomistic trajectory was translated (mapped) onto a CG-trajectory, which was further used for calculation of the reference structural properties needed for parameterization of the CG-model. Thus, 10 RDFs for nonbonded interactions and distributions of bond lengths (five bond types) and bended angles (five angle types) for intramolecular interactions were computed. The RDFs were computed up to cutoff distance $r_{cutoff} = 20$ Å with resolution of 0.1 Å. The bond length and bended angle distributions had resolution of 0.02 Å and 1.0°, respectively. Such set of reference distribution functions was calculated for each of the four reference systems.

Effective potentials: RDF inversion

The RDFs and bond length/angle distributions were used as input to the inverse procedure, which returns a set of effective coarse-grained potentials, reproducing within the coarse-grained model, the same set of distribution functions of the reference system. In general, the procedure works as follows: First, a Monte–Carlo simulation of the CG-system described by trial potentials is performed, and a set of RDFs and bond length/

angle distributions is obtained. After that, the potentials are updated, giving us a new set of trial potentials. The update is defined by the deviation between the calculated distribution functions and the reference ones. Then a new sampling starts with the new set of trial potentials, and repeated until the desired agreement between the reference and sampled distributions is reached. The update can be calculated within two approaches: iterative Boltzmann inversion[48,49] or by the inverse Monte–Carlo[38] (known also as Newton inversion[50]).

At each iteration of the inverse procedure, the Monte–Carlo sampling was carried out in the NVT ensemble conditions, using the same number of DMPC molecules and temperature as in the corresponding atomistic simulation. The box size was set to the average value of the periodic cell size of the reference atomistic system. Thus possible size effects on the RDFs should be largely cancelled out in direct and inverse simulations. The electrostatic interactions were treated by the Ewald method[51] using partial charges of the coarse-grained sites shown on Figure 1b. As the solvent was removed at this stage, a relative dielectric permittivity $\varepsilon=70$ was used to describe effective dielectric media. The system's partial charges were kept constant, thus the long range electrostatic potentials are taken out of the update. The short-range interactions were set to have the same cutoff distance of 20 Å as the reference RDF. Ranges of distances (angles) with zero values of the reference distributions were considered as prohibited for MC transitions. At each iteration, a new starting configuration of the coarse-grained system was randomly generated, thus memory effects in the sampling were avoided. We have used three kinds of Monte–Carlo transitions: single bead displacement, whole molecule displacement, and whole molecule rotation. The maximum displacements within each kind of the MC moves were adjusted to provide acceptance ratio for each of them at the level of about 0.5.

The potentials of mean force were used as starting trial potentials for the inverse procedure. The first 10 iterations were run using the iterative Boltzmann inversion (IBI)[48,49] method. It provides fast but rough initial optimization of the trial potentials. After that, the IMC algorithm was switched on. The IMC algorithm takes the cross correlation terms into account and by this reason it has higher demand to the sampling quality comparing to the IBI method. Also, higher accuracy of sampling is needed at the final stages of the potential tuning, when the difference between the reference and computed RDF became small. Therefore, we were increasing a number of MC sampling steps by factor 2 every 10–20 iterations while making about 50–100 iterations in total. To improve efficiency of sampling, simulations were run in the parallel mode employing (up to) 48 processors and collecting data for averaging from each processor.

CG simulations

As a result of the inverse procedure, eight sets of effective potentials (with and without account of angle bending interactions for each of the four compositions) were obtained. To facilitate further use of these potentials in mesoscale MD simulations, they were smoothed providing also continuous first derivative at the cutoff distance. Soft harmonic-like walls were added to bond potentials, restricting the bonds to stay in the region where bonding potential is defined. Nonbonded potentials were also extended with repulsive quadratic potential in the core region. Resulting extended potentials were used in three kinds of MD simulation: model validation, preassembled bilayer simulation, and self-aggregation simulation. All the CG molecular dynamics simulations with effective potentials were carried out using a stochastic Langevin dynamics integrator mimicking the friction and random forces arising from the solvent. At the moment, we did not fit the friction parameter (which was set to 1.0 ps^{-1} in all CG simulations) to match the diffusion in CG model with the diffusion in atomistic simulation, since we were primarily interested in the structural properties of lipid assemblies.

Model validation. We have first verified that the set of effective potentials, obtained within Monte–Carlo sampling, reproduced the reference properties if sampled by molecular dynamics with smoothed potentials. For that, the systems of the same number of lipids and periodic box sizes were simulated in the NVT ensemble at temperature of 303 K by GROMACS. In the validation CG Langevin MD simulation, the water was also accounted implicitly by dielectric permittivity of $\varepsilon=70.0$, thus, the same simulation conditions were used for MC-sampling and the verification run. These MD simulations were run for 500 ns with a MD time step of 3 fs. The structural distributions obtained have shown very close agreement with the reference distributions (in most of cases within the thickness of line on the plot), thus a simple proof-test for the effective potentials was fulfilled successfully.

Bilayer simulation. Once the reproducibility of the structural properties obtained within CG models was validated, several simulations of preassembled bilayers were performed. These simulations started from a flat bilayer, constructed of either 98 or 800 lipid molecules (49 and 400 molecules in each leaflet, respectively), packed in a periodic rectangular cell. This study was performed in NPT ensemble with semi-isotropic Parrinello–Rahman barostat at pressure 1 bar. Because of the absence of explicit solvent, the box size in *Z*-direction was fixed (at 8.5 nm), whereas box sizes in *X*- and *Y*- directions (along the membrane plane) were allowed to change. The temperature was kept constant by the Langevin thermostat. Most of such simulations were run at 303 K while several runs were made at other temperatures to investigate the temperature effects. The starting box sizes were set to values corresponding to artificially stretched bilayer, with average area per lipid 0.8 $nm^{[2]}$/lipid compared to the experimental value of 0.6 $nm^{[2]}$/lipid equal also to the value observed in atomistic simulations.[44] In total, 100 ns of production trajectory integrated with 3 fs time step was generated for bilayers of both sizes and each considered CG-model and concentration.

Lipids self-aggregation. Two self-aggregation studies of initially randomly distributed lipids were carried out: self-assembly into bilayer and vesicle self-aggregation. First, it was

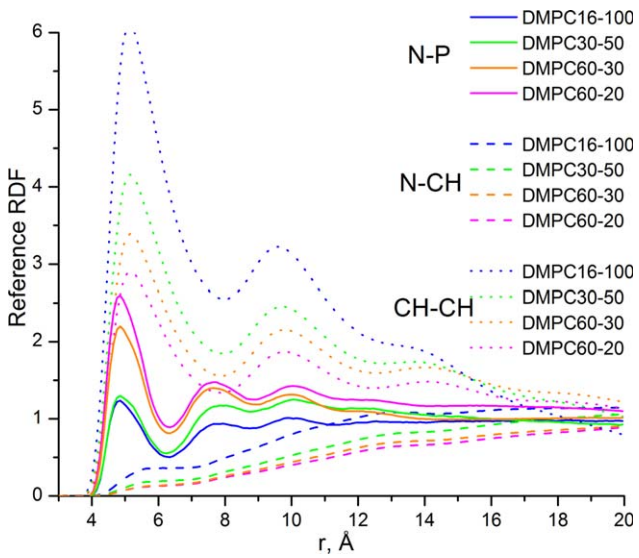

Figure 2. Reference radial distribution functions obtained in atomistic MD simulation: N-P sites—solid lines; N-CH sites—dashed lines; and CH-CH sites—dot lines. Results for DMPC16–100 are drawn in blue, DMPC30–50 in green, DMPC60-30 in orange, DMPC60-20 in purple, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

tested how the lipids were able to self-aggregate into a flat bilayer. For that, a system of 800 coarse-grained DMPC lipids was considered. The periodic box was kept constant with sides of 15.5 nm, which provided natural area per lipid (0.6 $nm^{[2]}$/lipid)[52] for a membrane oriented along the box *XY* plane. Uniform distribution of lipids in the box was taken as starting point. The study was performed in NVT ensemble, where temperature of 303 K was maintained by stochastic Langevin dynamics with the time step of 3 fs.

Another large-scale MD simulation was made with the purpose to model lipid self-aggregation in excess of water. This system consisted of 1000 coarse-grained DMPC molecules initially randomly distributed in a cubic box with a fixed side length of 30 nm. The larger box length excluded formation of a flat bilayer going through periodic conditions. Instead, formation of bicelles or vesicles could be observed. The time step of this simulation carried out by Langevin dynamics was set to 5 fs and the total simulation time was extended to about 3 μs.

Results and Discussion

Effective potentials: RDF inversion

The reference RDFs obtained from atomistic MD simulations are shown in Figure 2. To keep the number of figures reasonable, we, in the main text, report only data for N—P, N—CH, and CH—CH CG sites, that is, with one example for head–head, head–tail, and tail–tail pairs. Results involving other pairs are given in the Supporting Information Figure 1. As one could expect, RDFs obtained in atomistic simulation show clear trend with the change of concentration. It is best seen with higher association of CH—CH beads at low concentrations, which demonstrates an increased hydrophobic effect. N—P interactions show an opposite trend, with decrease of head-to-head contacts at lower concentration. Concentration dependence of N—CH RDF is somewhat weaker except the case of the lowest concentration.

Each of four sets of RDFs, together with bond distributions, was used as an input to the inverse Monte–Carlo procedure in both "angle" and "no-angle" models, resulting in eight sets of effective potentials. As for RDFs, here we only report N—P, N—CH, and CH—CH potentials (Fig. 3), leaving all other potentials including intramolecular bonded for Supporting Information Figures 2–4. The potentials also show concentration dependence, with the case DMPC60-20 differing (for some site pairs) more significantly from others. A probable reason of this specific case may be insufficient sampling which could happened due to small free volume available for every lipid at high concentrations. An important observation is that potentials for "angle" models are generally less concentration dependent than for "no-angle" models, which indicates that inclusion of intramolecular angle interactions between coarse-grained beads improves transferability of the effective potentials. The effect of angular intramolecular interactions on the nonbonded potentials appears because in the absence of angular potentials, the 1–3 CG neighbors interact by the corresponding nonbonded potentials, which, in this case, become affected by the need to fit distributions of 1–3 neighbors.

Preassembled bilayer simulation

Each of eight sets of the potentials was used for a separate MD simulation of a preassembled small and large lipid bilayer constructed of 98 and 800 CG-lipids, respectively. Density profiles of the headgroups (N beads) calculated for the small bilayers were compared with results of the atomistic simulation, where the bilayer of the same size was used, see Figure 4. All CG simulations resulted in distributions being approximately as wide as the reference atomistic distribution, and results for "angle" models appear to be somewhat closer to the reference ones than "no angle" models. The bilayer thickness, estimated as a distance between the density maxima, was found to be about equal for all CG models and in close agreement with the atomistic simulations. The density profiles of the headgroups in the systems consisting of 800 CG lipids show the same positions of maxima but wider distributions due to undulations of the large bilayer fragment (Fig. 6 in Supporting Information).

The small bilayer simulations were also used for calculation of the lipid tail order parameter defined as $S=(3<\cos^2\theta>-1)/2$, where θ is the angle between the bilayer normal and vector connecting two lipid sites. Two order parameters defined by CO—CH and CH—CH beads, as shown on Figure 5 were computed. The lipid tails demonstrate a change of orientational ordering with potentials obtained at different lipid concentration. The potentials based on low concentration systems show a higher order parameter, while the lowest order is obtained with the high concentration based potentials. The potential corresponding to composition of DMPC60-30 with angle bending interactions was found to provide tail order

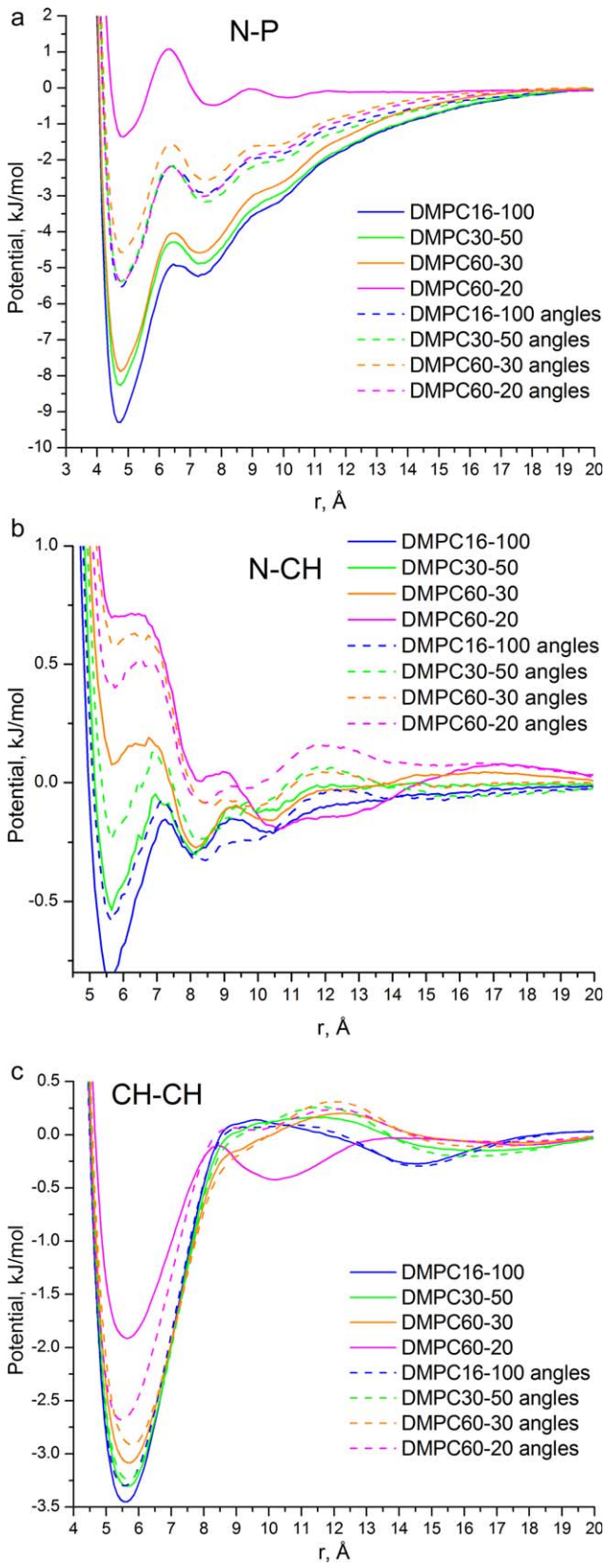

Figure 3. Effective short-range pairwise potentials: Potentials obtained for pairwise only model are drawn with solid lines and potentials obtained for model with bended angles are drawn with dashed lines. Different concentrations are shown by color: DMPC16-100—blue, DMPC30-50—green, DMPC60-30—orange, and DMPC60-20—purple.

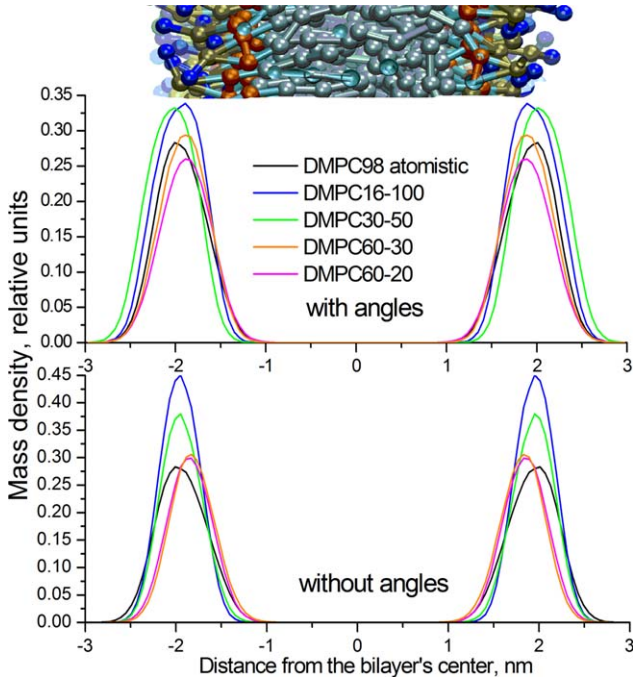

Figure 4. N-bead density profile calculated for the bilayer of 98 lipids. Black line—atomistic bilayer simulation; blue, green, orange, and purple lines—CG simulation with a set of potentials based on reference system DMPC16-100, DMPC30-50, DMPC60-30, and DMPC60-20, respectively. Lower plot shows results obtained with DMPC model based on pairwise bonds only. Upper plot shows results obtained with additional bending angle bonds in CG DMPC model.

parameters almost equal to the results of the reference atomistic simulation, shown by dashed lines in Figure 5.

An important property of a lipid bilayer is its projected area per lipid. The areas per lipid calculated for both small (98 lipids) and large (800 lipids) bilayers are shown in Table 2. There are only small, almost within error limits, differences in the

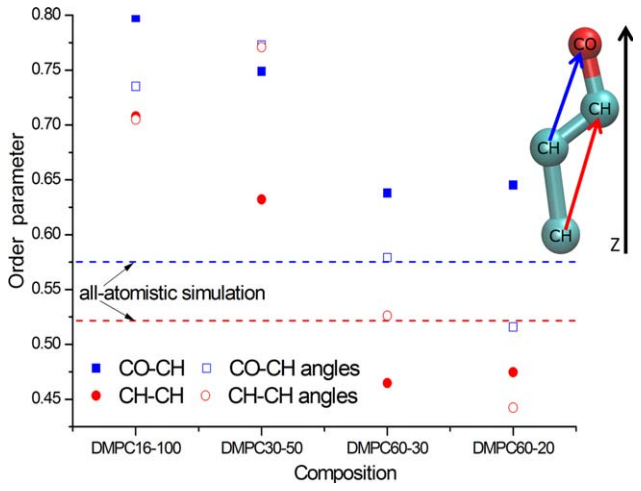

Figure 5. Order parameter of lipid tails calculated for the bilayer of 98 lipids. Blue squares represent order parameter for CO—CH groups, whereas red circles represent order parameter for CH—CH groups defined by the arrows in an inset showing lipid tail. Filled symbols denote "no-angle" model; open symbols denote model with bending angles. Dashed horizontal lines show results of atomistic bilayer simulation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. Physical properties of coarse-grained lipid bilayers with respect to lipid/water composition of the reference atomistic system.

System	Area (Å$^{[2]}$/lipid)		K_A (mN/m)	D (10^{-6}cm$^{[3]}$/s)
N lipids	98	800	800	800
DMPC16-100, angles	49.1 ± 0.7	49 ± 0.3	910 ± 50	1.4 ± 0.4
No angles	48.1 ± 0.7	48.1 ± 0.3	940 ± 40	1.8 ± 0.2
DMPC30-50, angles	50.1 ± 0.7	49.8 ± 0.3	910 ± 40	1.2 ± 0.2
No angles	55.3 ± 0.8	53.3 ± 0.3	710 ± 30	2.4 ± 0.1
DMPC60-30, angles	59.7 ± 1	59.3 ± 0.4	370 ± 30	1.6 ± 0.1
No angles	64.5 ± 1	62.4 ± 0.6	240 ± 40	3.4 ± 0.2
DMPC60-20, angles	65.7 ± 1	65.1 ± 0.6	190 ± 20	2.5 ± 0.1
No angles	65.1 ± 1	64.5 ± 0.6	330 ± 30	3.9 ± 0.3
Atomistic	60 ± 0.7	–	250 ± 20	0.09 ± 0.003

areas between large and small systems described within the same set of potentials. There is, however, noticeable dependence of the lipid area both on the concentration at which the potentials were generated, and whether angle bending interactions were included. Potentials based on low lipid concentration systems, result in a low area per lipid with values typical for a bilayer in a gel phase. Visual analysis of the corresponding snapshots shows, however, that the lipids remain essentially unordered (snapshots in Supporting Information). Oppositely, potentials defined at high lipid concentrations lead to higher values of the area per lipid, which corresponds to a more disordered fluid phase. The best agreement with the atomic simulation (as well as with experiment) is reached for effective potentials computed at composition of 30 waters per lipid (DMPC60-30).

Bilayer compressibility was also estimated from the fluctuations of the area according to $K_A = k_B T \langle A \rangle / (\langle A^2 \rangle - \langle A \rangle^2)$, where A is the total area of the simulated bilayer patch. The results (computed for the large systems) are shown in Table 2. As in the case of average area, the best agreement with atomistic simulations was achieved for DMPC60-30 composition, while low concentration based potentials provide significantly higher values indicating more rigid bilayer.

While effective potentials obtained by the structure-based coarse graining reproduce accurately pair distribution functions, three-particle correlations are not necessarily correctly reproduced.[30,33] We have compared three-body correlations expressed in terms of angular distribution between a choline (N) site of one molecule and two other nearest sites (N or P) belonging to two other lipid molecules. Results are presented in Figure 6. All four coarse-grained models reproduce reasonably well the atomistic angular distribution, with the system DMPC60-30 showing the best agreement. Thus, the CG model provides correct higher-order structure of the mosaic of positive and negative charges at the bilayer surface, which may be of importance in using this CG model for studies of interaction of other charged or polar objects with the membrane. It is worth to note that in previous studies of three-body correlations in the single-site water model[53] angular distributions between neighboring molecules were found different in the CG and atomistic model. In the case of lipids, CG lipid molecules interact by many site–site pair potentials (a not by a single center-to-center potential) reproducing the whole set of site–site RDFs, which leaves less freedom for possible changes of the three-body distributions, which by this reason occur similar in atomistic and CG representations.

Lateral diffusion illustrates dynamical properties of a coarse-grained bilayer (Table 2). The diffusion coefficients were

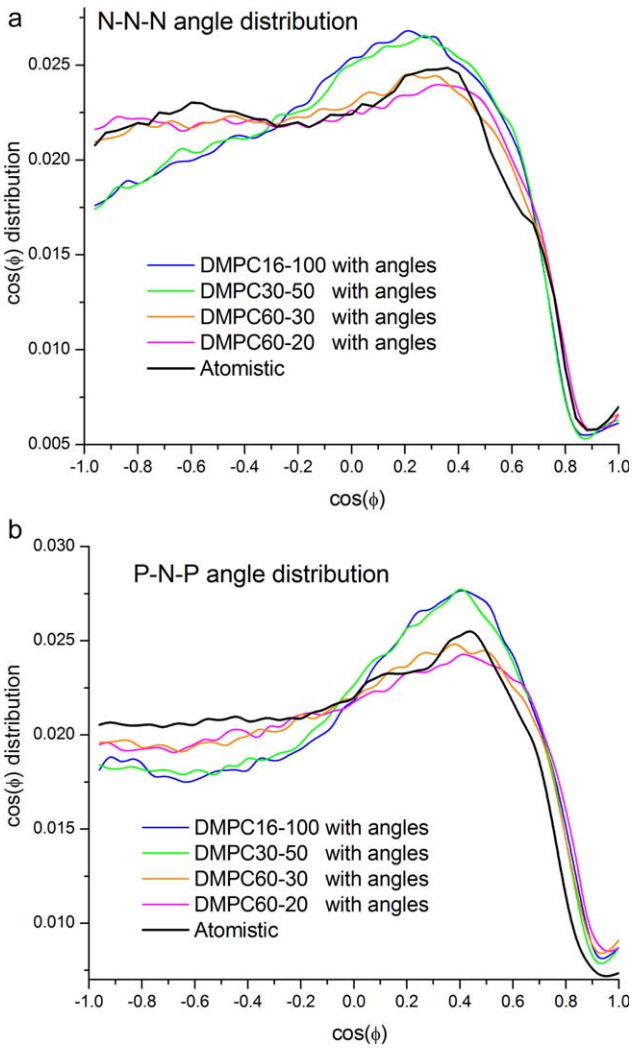

Figure 6. Distribution of $\cos\phi$ where ϕ is the angle formed by "N" site of a lipid (which is the central site of the angle) and two "N" (a) or two "P" (b) nearest sites belonging to two other molecules. Comparison of four "Angle" CG models and atomistic simulation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

obtained from mean square displacement (MSD) linear fitting using Einstein relation. The lateral diffusion obtained in CG simulation, as it was also observed in many other studies, is significantly higher than in atomistic simulations, which means that the internal time reported in CG simulation shall be scaled to get the actual real time corresponding to atomistic MD simulation. The speedup of the dynamics due to coarse-graining effects (i.e., smooth potential energy surface and absence of solvent) can be estimated from ratio D_{CG}/D_{atomic}, which is about 20–50 times. This additional speedup (alongside with speedup obtained from the reduction of degree of freedom and possibility to use longer time step) is favorable, if the aim of the simulation is to find equilibrium state of the system and provide fast sampling of the configurational space.

As the friction parameter of the Langevin dynamics was the same in all the CG systems, relative diffusion in different CG models can be compared. The data show a significant grow of the diffusion rate with lipid concentration in the reference system. Such trend exists for both "angle" and "no-angle" models, however, the later provides faster diffusion. Suppressed diffusion of the systems parameterized at low concentrations (DMPC16–100, DMPC30–50) can be considered as an additional evidence that the bilayer described by these sets of potentials has a preference to a more ordered phase with a more restricted mobility of the lipids.

Choice of the best CG potential

Summarizing comparison of CG bilayer systems derived from atomistic simulations performed at different concentrations, we can conclude that there is an overall trend that the potentials derived at lower concentration of lipids provide a more ordered, condensed, and tightly bound CG system with less mobile lipids, whereas potentials derived from concentrated atomistic lipid system provide more fluid-like coarse-grained systems. This can be observed from the behavior of the average lipid area, compressibility, tail order parameters, and diffusion. The likely origin of such behavior is concentration dependence of the effective hydrophobic interactions: while in the limit of low concentration, the hydrocarbon chains are effectively attracted because water molecules prefer to form water–water bonds instead of water-hydrophobic contacts, at high concentration of hydrocarbon chains where the water molecules are practically excluded from the hydrophobic phase this effect disappears. One can see from Figure 3c that the attractive depth of CH—CH effective potential increases with decrease of the concentration at which the potential was derived. Even if other site–site potentials show varying behavior, the CH—CH site–site potentials stay for 36% of all site–site contacts in a 10-site DMPC model, and they completely determine behavior of the hydrophobic membrane interior. We can note in light of this that existing practice of using the mean-force potentials obtained in the limit of low concentration as CG potentials can lead to overestimation of the hydrophobic attraction and aggregation phenomena.

The best agreement between CG and atomistic model describing bilayer structure was observed for DMPC60-30 system with angles. It is remarkable that just this ratio of water molecules per lipid (30:1) is considered as condition of full hydration of phosphatydylcholine lipid headgropus.[54] At this molar ratio, the lipid headgroups are properly hydrated in the atomistic system and provide water-mediated effective potentials which reproduce realistic behavior of lipid polar groups in the CG model. At the same time, water molecules are mostly excluded from mediation of interactions between hydrophobic lipid tails in the atomistic model leading to CG tail–tail interactions similar to those in the membrane interior. Furthermore, the geometry of DMPC60-30 system and amount of lipids in it does not allow formation of an unperturbed bilayer through the (cubic) periodic box, leading to relatively frequent contacts of type "hydrocarbon tail"—"hydrated polar group" or "hydrocarbon tail"—nonbound water, thus providing sampling of all important CG interaction types. All together this explains why the conditions of DMPC60-30 system results in CG potentials which best of others match the atomistic simulations for the bilayer phase. In the following sections, we use CG potential obtained from DMPC30–60 system with angles.

CG bilayer phase transition

As it was discussed above, the CG lipids models obtained in lower-concentration atomistic simulations demonstrate tendencies of going to gel phase, which is manifested by a lower average area per lipid, higher orientational ordering and slow diffusion. However, it cannot be stated that these systems (DMPC16–100, DMPC30–50) attained a true gel phase because the lipids remain still unordered and diffusion is far from being zero. We investigated whether the CG bilayer described by the same set of potentials can undergo liquid crystalline-gel transition under change of temperature. Therefore, a set of simulation of DMPC60-30 system with angles was carried out in the temperature range between 200 and 350 K with step of 25 K.

The results for area per lipid and lateral diffusion coefficient are depicted on Figure 7. Both of them show a well seen jump between 225 and 250 K, which indicates a gel–liquid phase transition. Snapshot of the system at 225 K in Figure 8 (top) demonstrates clearly gel-phase ordering of lipids at 225 K, whereas higher temperature 250 K system shows disordered structure of the lipid tails, see Figure 8 (bottom). The experimental phase transition temperature for DMPC bilayer is 297 K, and lower phase transition temperature observed in simulation can be prescribed to both coarse-grained description of the system and to possible deficiencies of the atomistic force field.

Lipids self-aggregation

Bilayer self-aggregation. For further studies of lipid self-aggregation, we choose set of potentials DMPC60-30 with angles, which demonstrated the best agreement for the average area per lipid and the tail orientational ordering. The self-aggregation was carried out in a constant-volume cubic cell with the size providing same area per lipid for a membrane made of 800 lipids and oriented along any of its sides. The aggregation happened relatively fast: the bilayer was formed already after 18–20 ns (unscaled time) and remained further

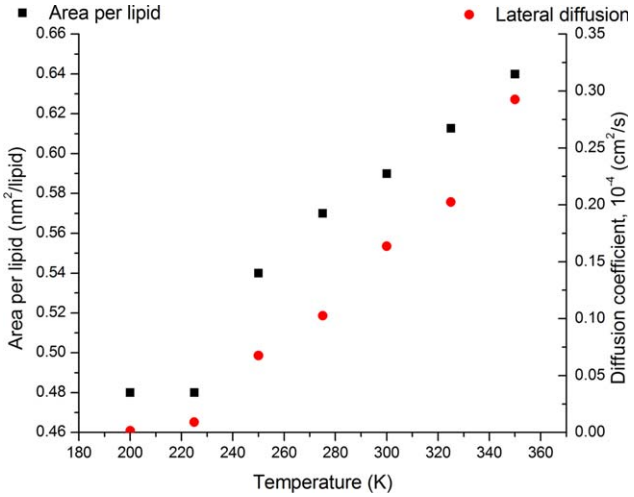

Figure 7. Temperature dependence of bilayer properties obtained within same potentials set "DMPC60-30 with angles." The bilayer consists of 800 lipids. Black squares denote area per lipid; red circles denote lateral diffusion. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

stable for following 80 ns of the simulation. While aggregating, the system, starting from uniform random distribution of lipids (Fig. 9a), passed through several characteristic stages: first, almost immediately (about 30–50 ps) separate lipids merged into a number of dense clusters, which further joined together and formed a distorted band-like structure (Fig. 9b). This stage took about 300–500 ps after which the band branches began to unite into irregular sheet with holes and breaks (Fig. 9c). After 3–5 ns, the sheet got formed and began relaxation to a flat bilayer (Fig. 9d), which proceeded during 15–18 ns. The resulting bilayer remained stable during next 80 ns until the end of the simulation.

Bicell and vesicle formation. Another case of self-aggregation is assembly of lipid molecules in bicell- or vesicle-like structures in excess of water. In the example, a system of 1000 coarse-grained DMPC lipids was put into fixed periodic cell of

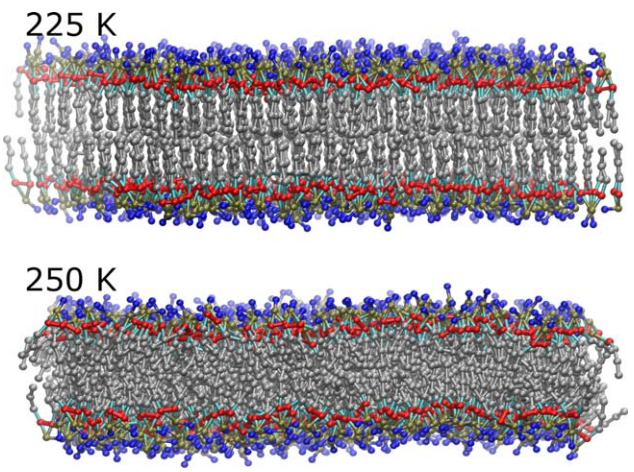

Figure 8. Lipid tail ordering due to gel–liquid phase transition in lipid bilayer. Top bilayer—gel phase state at 225 K. Bottom bilayer—Liquid phase state at 250 K. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

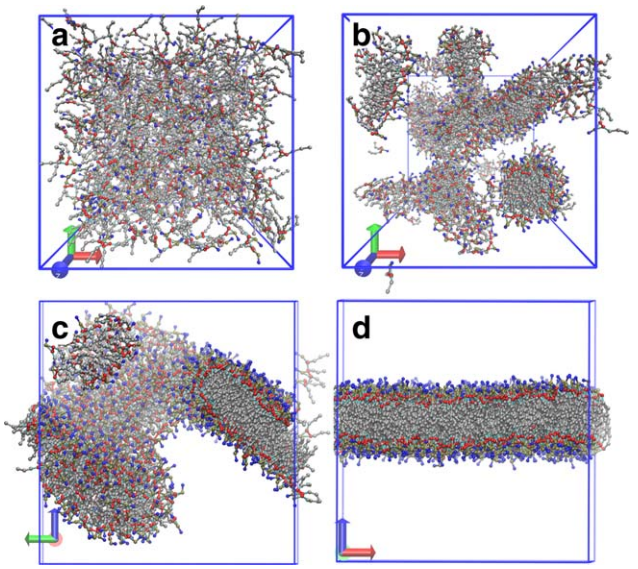

Figure 9. Bilayer formation: (a) Initial uniform distribution of lipids (800 lipids); (b) formation of dense clusters after 500 ps; (c) formation of distorted bilayer after 5 ns; (d) formation of stable bilayer after 20 ns. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

size 30 × 30 × 30 nm. In this initial configuration, the lipid molecules were randomly distributed in the box (Fig. 10a). After start of the simulation, a fast clustering of lipids into

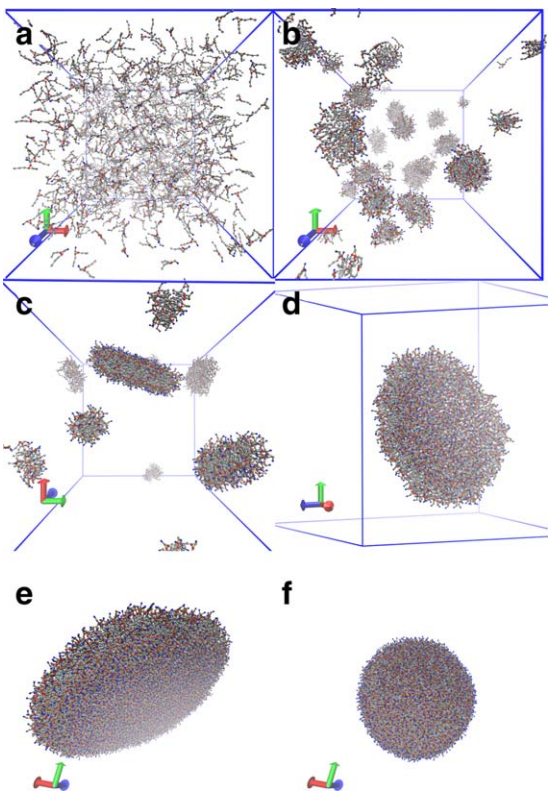

Figure 10. Vesicle formation: (a) initial uniform distribution of lipids (1000 lipids); (b) formation of small spherical clusters; (c) formation of small round-flat bilayers; (d) formation of large flat-round bilayer (1000 lipids); (e) Preassembled large round-flat bilayer (4000 lipids); and (f) stable spherical vesicle (4000 lipids). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

small semispherical droplets (Fig. 10b) was observed. These droplets merged further (in about 500 ns) in a small disk-like structures, so-called bicelles (Fig. 10c), which after 2 μs formed one large "pancake" (Fig. 10d). During next 3.5 μs, the "pancake" faced some undulations but did not changed structure significantly, and kept staying in a flat formation as on Figure 10d. When in another simulation, a similar bicell was formed by 4000 lipids (Fig. 10e), after period of some undulations the system collapsed into a spherical vesicle during a 100 ns transition (Fig. 10f) and remained in this phase during next 1 μs until end of the simulation. In DMPC-type lipid self-assembly, the resulting structure (bicell or vesicle) is defined by two competing factors: the energy expenses to form a curved surface (in a vesicle) and the presence of hydrophobic lipid tail-water contacts at the bicell edge. Typically, vesicles can be formed from bicelles if the number of lipids in it exceeds a few thousand,[26] though cases of smaller vesicle formation with 864 lipids were also reported.[55]

Conclusions

We have systematically applied structure-based coarse-graining approach to derive effective solvent-mediated potentials for a 10-site coarse-grained lipid model. The effective potentials, obtained by the inverse Monte–Carlo technique, provide for the coarse-grained model the same site–site RDFs, bond and angle distributions as those computed in atomistic simulations. Specifically, we investigated the role of the reference (atomistic) system composition on the coarse-grained potentials and on the properties of the lipid bilayer systems simulated with these potentials. It was shown that there is a non-negligible concentration dependence of the effective potentials and properties of lipid bilayers simulated using these potentials. Thus, potentials based on low lipid concentration overestimate the effective hydrophobic attraction of the lipid tails, which favors to more gel-like and more ordered structure of the bilayer. The potentials based on higher lipid concentration in the atomistic simulations provide more fluid-like structure with larger area per lipid. Introduction of bending angle interaction into coarse-grained model makes effective potentials somewhat less concentration dependent. The best agreement with reference data as well with experiment was achieved with a set of potentials derived from atomistic simulations at 1:30 lipid:water molar ratio (DMPC60-30) providing full saturating hydration of DMPC headgroup[54] in bilayer.

Despite the theoretical limitations of the structure-based approach to derive effective potentials, expressed in their state point dependence, all the derived potentials obtained in conditions of unordered lipid–water mixture, provided a stable bilayer structure with correct partitioning of different lipid groups across the bilayer as well as with acceptable values of the average lipid area, orientational tail ordering, and compressibility. This behavior was reached without use of any pressure correction of the coarse-grained potentials, which according to previous investigations might be necessary in some coarse-grained models, for example, one-cite water.[53] Another important property demonstrated by the model is liquid/gel phase transition with lowering of the temperature which demonstrates that the model's capacity is not limited just to structural properties at the state point. In addition to bilayer simulations, the model has proven its robustness in modeling of self-aggregation of bilayer/vesicle from a dispersed lipid solution, showing realistic behavior without introduction of any empirically fitted parameters or additional artificial interactions. If compared to fully atomistic model such coarse-grained model provides acceleration of two to four orders of magnitude, due to reduced number of particles as well as increased time step and diffusion rate. Due to the absence of explicit solvent, such solvent-free CG models are computationally superior the explicit-solvent CG models while providing at least similar level of accuracy and having potential to reproduce qualitatively dynamics by proper tuning of parameters of a Langevin (or Lowe–Anderson) thermostat.

Acknowledgments

Authors are thankful to the Swedish National Infrastructure for Computing (SNIC) for granting access to high performance computing facilities at the PDC Center for High Performance Computing, Stockholm, High Performance Computing Center North (HPC2N), Umeå and National Supercomputing Centre (NSC), Linköping. AM is also grateful to Joakim Jämbeck for fruitful discussions and useful hints given while working on the project.

Keywords: bottom-up coarse graining · inverse Monte Carlo · lipid bilayers · dimyristoylphosphatidylcholine lipids · implicit solvent potentials

How to cite this article: A. Mirzoev, A. P. Lyubartsev *J. Comput. Chem.* **2014**, *35*, 1208–1218. DOI: 10.1002/jcc.23610

Additional Supporting Information may be found in the online version of this article.

[1] E. Lindahl, M. S. Sansom, *Curr. Opin. Struct. Biol.* **2008**, *18*, 425.
[2] S. J. Marrink, A. H. de Vries, D. P. Tieleman, *Biochim. Biophys. Acta* **2009**, *1788*, 149.
[3] A. P. Lyubartsev, A. L. Rabinovich, *Soft Matter* **2011**, *7*, 25.
[4] K. Yanga, Y. Ma, *Soft Matter* **2012**, *8*, 606.
[5] F. L. H. B. Grace Brannigan, L. C.-L. Lin, *Eur. Biophys. J.* **2006**, *35*, 104.
[6] J. M. Drouffe, A. C. Maggs, S. Leibler, *Science* **1991**, *254*, 1353.
[7] R. Goetz, R. Lipowsky, *J. Chem. Phys.* **1998**, *108*, 7397.
[8] H. Noguchi, M. Takasu, *Phys. Rev. E* **2001**, *64*, 041913.
[9] O. Farago, *J. Chem. Phys.* **2003**, *119*, 596.
[10] M. J. Stevens, *J. Chem. Phys.* **2004**, *121*, 11942.
[11] G. Brannigan, F. L. H. Brown, *J. Chem. Phys.* **2004**, *120*, 1059.
[12] G. Brannigan, P. F. Philips, F. L. H. Brown, *Phys. Rev. E* **2005**, *72*, 011915.
[13] O. Lenz, F. Schmid, *J. Mol. Liq.* **2005**, *117*, 147.
[14] F. Schmid, D. Düchs, O. Lenz, B. West, *Comput. Phys. Commun.* **2007**, *177*, 168.
[15] I. R. Cooke, K. Kremer, M. Deserno, *Phys. Rev. E* **2005**, *72*, 011506.
[16] A. J. Sodt, T. Head-Gordon, *J. Chem. Phys.* **2010**, *132*, 205103.
[17] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, M. L. Klein, *J. Phys. Chem. B* **2001**, *105*, 4464.
[18] T. Murtola, E. Falck, M. Patra, M. Karttunen, I. Vattulainen, *J. Chem. Phys.* **2004**, *121*, 9156.
[19] T. Murtola, E. Falck, M. Karttunen, I. Vattulainen, *J. Chem. Phys.* **2007**, *126*, 075101.

[20] T. Murtola, M. Karttunen, I. Vattulainen, *J. Chem. Phys.* **2009**, *131*, 055101.
[21] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750.
[22] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812.
[23] S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2005**, *109*, 2469.
[24] S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2009**, *113*, 4443.
[25] L. Lu, G. A. Voth, *J. Phys. Chem. B* **2009**, *113*, 1501.
[26] A. P. Lyubartsev, *Eur. Biophys. J.* **2005**, *35*, 53.
[27] Z. Wang, M. Deserno, *J. Phys. Chem. B* **2010**, *114*, 11207.
[28] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodriges-Ropero, N. van der Vegt, *Soft Matter* **2013**, *9*, 2108.
[29] A. A. Louis, P. G. Bolhius, J. P. Hansen, E. J. Meijer, *Phys. Rev. Lett.* **2000**, *85*, 2522.
[30] A. A. Louis, *J. Phys.: Condens. Matter* **2002**, *14*, 9187.
[31] J. W. Mullinax, W. G. Noid, *J. Chem. Phys.* **2009**, *131*, 104110.
[32] A. Villa, C. Peter, N. F. A. van der Vegt, *J. Chem. Theory Comput.* **2010**, *6*, 2434.
[33] S. Izvekov, *J. Chem. Phys.* **2011**, *134*, 034104.
[34] B. Hess, C. Holm, N. van der Vegt, *J. Chem. Phys.* **2006**, *124*, 164509.
[35] B. Hess, C. Holm, N. van der Vegt, *Phys. Rev. Lett.* **2006**, *96*, 147801.
[36] J. Shen, C. Li, N. F. A. van der Vegt, C. Peter, *J. Chem. Theory Comput.* **2011**, *7*, 1916.
[37] O. Engin, A. Villa, C. Peter, M. Sayar, *Macromol. Theory Simul.* **2011**, *20*, 451.
[38] A. P. Lyubartsev, A. Laaksonen, *Phys. Rev. E* **1995**, *52*, 3730.
[39] A. Mirzoev, A. P. Lyubartsev, *J. Chem. Theory Comput.* **2013**, *9*, 1512.
[40] S. Nosé, *Mol. Phys.* **1984**, *52*, 255.
[41] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.
[42] M. Parrinello, A. Rahman, *Phys. Rev. Lett.* **1980**, *45*, 1196.
[43] S. Nosé, M. L. Klein, *Mol. Phys.* **1983**, *50*, 1055.
[44] C.-J. Högberg, A. M. Nikitin, A. P. Lyubartsev, *J. Comput. Chem.* **2008**, *29*, 2359.
[45] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
[46] B. Hess, H. Bekker, H. J. C. Berendsen, J. C. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463.
[47] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, 4, 435.
[48] A. K. Soper, *Chem. Phys.* **1996**, *202*, 295.
[49] D. Reith, M. Pütz, F. Müller-Plathe, *J. Comput. Chem.* **2003**, *24*, 1624.
[50] A. Lyubartsev, A. Mirzoev, L. Chen, A. Laaksonen, *Faraday Discuss.* **2010**, *144*, 43.
[51] M. P. Allen, D. J. Tildesley, Computer Simulation of Liquids; Oxford University Press, Oxford UK, **1989**.
[52] J. P. M. Jämbeck, A. P. Lyubartsev, *J. Phys. Chem. B* **2012**, *116*, 3164.
[53] H. Wang, C. Junghans, K. Kremer, *Eur. Phys. J. E* **2009**, *28*, 221.
[54] J. F. Nagle, R. Zhang, S. Tristram-Nagle, W. Sun, H. I. Petrache, R. M. Suter, *Biophys. J.* **1996**, *70*, 1419.
[55] S. J. Marrink, A. E. Mark, *J. Am. Chem. Soc.* **2003**, *125*, 15233.

Received: 21 January 2014
Revised: 20 March 2014
Accepted: 26 March 2014
Published online on 29 April 2014