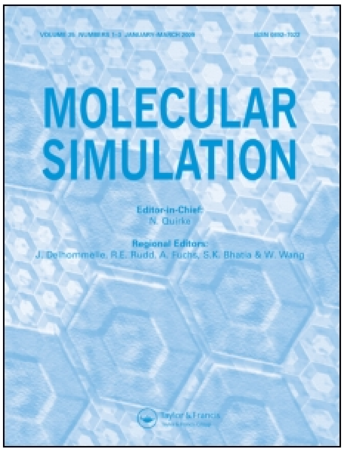

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants

W. Shinoda[ab]; R. DeVane[a]; M. L. Klein[a]
[a] Department of Chemistry, Center for Molecular Modeling, University of Pennsylvania, Philadelphia, PA, USA [b] Research Institute of Computational Sciences, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

To cite this Article Shinoda, W. , DeVane, R. and Klein, M. L.(2007) 'Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants', Molecular Simulation, 33: 1, 27 — 36

To link to this Article: DOI: 10.1080/08927020601054050

URL: http://dx.doi.org/10.1080/08927020601054050

Molecular Simulation, Vol. 33, Nos. 1–2, 15 January–15 February 2007, 27–36

Taylor & Francis
Taylor & Francis Group

Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants

W. SHINODA†‡*, R. DEVANE†¶ and M. L. KLEIN†

†Department of Chemistry, Center for Molecular Modeling, University of Pennsylvania, 231 S. 34th Street, Philadelphia, PA 19104-6323, USA
‡Research Institute of Computational Sciences, National Institute of Advanced Industrial Science and Technology, Central 2, 1-1-1, Umezono, Tsukuba, Ibaraki 305-8568, Japan

(Received August 2006; in final form October 2006)

A new systematic approach to build coarse-grained (CG) molecular models for surfactants/water systems is proposed. A step-by-step approach using several molecular systems for the parameterization makes the CG model versatile and transferable. The intramolecular bond potentials are determined to reproduce the bond and angle distributions obtained from all-atom (AA) molecular dynamics (MD) simulations. A careful choice of the potential function for nonbonded interactions is essential for better structural properties. Density and surface/interfacial tension data are used for parameter fitting, because these thermodynamic properties are of key importance in characterizing the self-organized surfactant structure. Solvation (hydration) and transfer free energies, which play an essential role in determining the partition of solute molecules, are also taken into account in the model.

Keywords: Coarse graining; Molecular dynamics; Polyethylene glycol; Surfactant; Hydration free energy; Surface tension

1. Introduction

In the last decade, considerable effort has been devoted to develop coarse-grained (CG) molecular models to study polymers, biomembranes, and surfactants [1–11]. A CG simulation is indeed a promising tool to study such complex fluids that contain long-time relaxation processes, which are inaccessible by all-atom (AA) simulations. However, a coarse-graining process, i.e. a simplification of molecular description with fewer interaction sites, necessarily involves a significant change in thermodynamics because of the entropy difference. As a result, the "effective" potential useful for the CG simulations cannot be derived systematically in the framework of statistical mechanics except for a few simple cases. Thus, no unique answer is generally available for the CG mapping. A key question is what molecular properties are required with the CG model. Some properties are not compatible at the CG level, i.e. coarse-graining involves a selection of the properties to be reproduced. Keeping this in mind, in this paper, we will discuss how we can derive a CG model for surfactant/water systems in a systematic way.

The minimal requirement for the model to have, for example, a bilayer structure for lipid molecules is rather simple. Some of the CG units (beads) should "like" CG water particles and the others "dislike" them. There have been several efforts to make a CG model for lipid membranes along this line [4,12]. Recently, Marrink *et al.* developed a CG model that treats a wide variety of lipids based on an extension of the bead-spring model with a LJ12–6 function for all non-bonded interactions and a single value of σ [7]. The non-bonded interaction is basically characterized just by choosing one of five ε values. Nevertheless, the CG model successfully reproduces several key structural and thermodynamic properties. The minimalism, however, involves some negative aspects such as a higher melting point of the CG water and a much broader alkane–water interface. As shown later, a different choice of functional form for the non-bonded interaction can improve the situation. (see Section 3.1)

Shelley *et al.* developed a CG model for a dimyristoyl phosphatidylcholine (DMPC) bilayer in a different way [8]. After the parameterizations for water and alkanes based on thermodynamic data, all the other pair interactions have been simultaneously modelled in

---

*Corresponding author. Email: w.shinoda@aist.go.jp
¶National Science Foundation Biological Informatics Post Doctoral Fellow.

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020601054050

iterative molecular dynamics (MD) calculations to reproduce the pair distribution functions of AA-MD. The latter is known as the inverse Boltzmann method, being widely used in polymer CG modelling [13]. This structure-based CG model shows a rapid convergence in several iterations and is generally useful for any molecular system [8]. However, there are a few concerns with this approach. Firstly, thermodynamic properties are not taken into account in the parameterization, though some refinements are possible to improve the situation. Secondly, a simultaneous mapping for many pair interactions leads to multiple solutions, some of which can be counter-intuitive. The problem is similar to that found in reverse Monte Carlo [14], which has been also used for an implicit solvent CG model [15]. Finally, the model can be specific to the molecule (molecular conformation) in the sampled AA-MD system, then the resultant interaction may be less transferable.

More recently, a systematic approach to CG models has been proposed by Izvekov and Voth [9]. These authors successfully applied a force-matching (FM) method to build a CG phospholipids bilayer using AA-MD trajectories and forces. Since the CG model has been constructed totally based on the AA forces, one can build a unique model from AA-MD data. This is an advantage of this method. However, one possible issue for this approach is applicability for CG particles composed of multiple small molecules. Our short analysis for the CG water model will elucidate the limitation of the FM approach.

In this paper, we propose a systematic approach to build a new CG surfactant model using thermodynamic data in combination with AA-MD trajectories. This approach has been used previously for water and alkane [8]. Later, Nielsen *et al.* successfully applied the methods to make a CG model for a series of alkanes using surface tension data [16]. Herein, we extend the idea to surfactant (amphiphilic) molecules. In general, surfactant/water complex systems exhibit a variety of morphologies depending on the conditions such as water content and temperature. Morphology is mainly determined by the interfacial properties, so that the surface/interfacial tension is one of the key properties to characterize the system. Thus, the first priority is to reproduce surface/interfacial tension and density for the nonbonded potential parameterization while keeping the functional form simple. The functional form is selected to refine the structural data and compressibility (of water). A step-by-step parameterization has been undertaken using several molecular systems. Each system has been examined at the AA and CG levels. By keeping the number of unknown parameters smaller than the number of target properties at each fitting step, the parameters are fixed uniquely. Free energy (FE) calculations are also employed to determine and evaluate the interaction potential. An extensive use of many molecular systems for fitting is essential for a systematic parameterization. Here we focus on the CG modelling for polyethylene glycol (PEG) surfactants as an example, though our approach will be useful for many other systems, including lipids, peptides and drugs.

2. Methodology

2.1. Modelling procedure

AA-MD simulations have been carried out to obtain reference structural data. The CG intramolecular potential is constructed using the corresponding distribution functions obtained from AA-MD. For intramolecular interactions, bond stretching (1–2) and bond angle bending (1–2–3), interactions are described by harmonic potentials;

$$U_{\text{intra}} = \sum^{\text{Bond}} k_b(r - r_0)^2 + \sum^{\text{Angle}} k_a(\theta - \theta_0)^2, \qquad (1)$$

where k_b and k_a are spring constants, r_0 and θ_0 are the distance and angle at the energy minimum. To get the initial guess for these parameters, we calculate the effective potential from the AA distribution functions [11],

$$U_{\text{bond}}(r) \propto -k_B T \ln[P(r)/r^2]$$

$$U_{\text{angle}}(\theta) \propto -k_B T \ln[P(\theta)/\sin\theta],$$

where $P(r)$, $P(\theta)$ are the probability distributions from AA-MD for the bond and bend, respectively, k_B is the Boltzmann constant, and T is the simulation temperature. The Jacobian should be taken into account. By fitting to the effective potential, a reasonable initial guess for the parameters in equation (1) is obtained, though some refinements are usually required to reproduce the same average and dispersion of the distributions in the CG-MD [16]. This requires iterative MD calculations.

The nonbonded interaction is determined in a different manner. A pairwise additive potential is used. To keep the model simple and versatile, we select rather common functional forms for the pair-potential. Several functional forms have been examined and our final choice is to use the following two Lennard–Jones (LJ) functions;

$$U_{\text{LJ9-6}}(r) = \frac{27}{4}\varepsilon\left\{\left(\frac{\sigma}{r}\right)^9 - \left(\frac{\sigma}{r}\right)^6\right\}, \qquad (2)$$

$$U_{\text{LJ12-4}}(r) = \frac{3\sqrt{3}}{2}\varepsilon\left\{\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^4\right\}. \qquad (3)$$

With the prefactors in the functions, ε is the minimum energy and σ is the distance at $U = 0$. The LJ12–4 function is used only for the pairs involving water, while the LJ9–6 is used for any other pairs. The non-bonded interaction is calculated for all pairs of particles except for 1–2, 1–3 bonded pairs. With this choice, we fix the parameters, ε and σ, to reproduce thermodynamic properties, e.g. surface/interfacial tension and density. A simple truncation (cut-off) of non-bonded interactions is used without any tail-correction for potential or virial.

Note that the tail-correction assumes no structure in $r > R_c$, i.e. $g(r) = 1$. However, this is not true for many surfactant/lipid systems that have long-range order. As a result, the cut-off length is an influential parameter to the thermodynamic properties of the CG system. The cut-off length is fixed at 1.5 nm, which is long enough to prevent significant artefacts due to the truncation. Taking a series of molecular systems, the parameters are determined in a systematic way as we will see below. To fix the parameter for the interaction between hydrophilic groups and water, the interfacial tension is not useful. In this case, we tune the parameters to have a correct solvation (hydration) FE.

2.2. *Molecular dynamics simulations*

A series of MD simulations of dozens of different systems have been carried out at the CG as well as the AA levels. The system temperature is maintained at 303.15 K for the parameter fitting using the Nosé–Hoover chain of length five [17,18]. The reversible RESPA algorithm [19] is used to generate the molecular trajectories. All MD simulations have been carried out using our in-house program codes, MPDyn [20] and CM3D.

2.2.1. All-atom MD. The CHARMM PARAM27 force field [21] is used for AA-MD except for PEG headgroups. The interaction parameters for PEG headgroup are taken from Ref. [22,23]. The SHAKE/RATTLE (ROLL) algorithm [24] is used to apply the constraint for the bonds involving hydrogen atoms. The long-range coulomb interaction is calculated using the Ewald or particle mesh Ewald method [25]. The van der Waals interaction is truncated at 1.2 nm by applying the standard CHARMM smoothing function for the tail region of 1.0–1.2 nm. The time step size is 2 fs. For a bulk simulation of small molecules, we typically take the following procedure to make MD trajectories. After preparing the initial configuration in the simulation box of size $L \sim 3$ nm, the system is thermalized at a relatively high temperature, e.g. 500 K, for 500 ps. After cooling down the system to the target temperature, 303.15 K, with another 100 ps equilibration run, molecular trajectories are sampled from 1 ns-MD simulation in a canonical ensemble.

2.2.2. Coarse-grained MD. Two different time step sizes are used; 10 fs for updating long-range nonbonded forces ($0.6 \sim 1.5$ nm) and 2 fs for updating short-range nonbonded and bonded forces. Typically, the simulation box size is 4.0 nm for the bulk systems. We fix the nonbonded parameters in two steps. First, a random initial configuration is generated in the simulation box with the appropriate number of molecules to reproduce the experimental density. By checking the system pressure in a short MD run in a canonical ensemble, we obtain a reasonable parameter set to reproduce the density. Typically 100 ps-MD is long enough to check the pressure. Second, an MD simulation of a slab of the liquid is carried out to estimate the surface tension. Just by changing the cell length in z from 4 to 40 nm in the last configuration of the bulk system, a reasonable starting configuration is obtained. To see the convergence of surface tension with a precision of 1 mN/m, $5 \sim 10$ ns MD calculations are usually needed. The surface tension is computed by

$$\gamma = \frac{L_z}{2}\left\{P_{zz} - \frac{P_{xx} + P_{yy}}{2}\right\}, \tag{4}$$

where the factor of 1/2 is included to account for the two interfaces in the simulation box, and P_{ij} is the ij component of the averaged pressure tensor. To verify the system density, NPT-MD [26] is also carried out for 1 ns.

2.2.3. Free energy calculations. Hydration (solvation) FE is estimated in the slab water system by calculating the work needed to transfer the target molecule from the vacuum region to the centre of the water layer. A series of steered MD (SMD) [27] is used to estimate the work based on Jarzynski's equality [28]. Taking the harmonic spring constant, $k = 10\,\text{kcal/mol/Å}^2 (= 0.4184$ kJ/mol/nm^2), the target molecule is pulled at the constant rate of 1 nm/ns. With 15 SMD simulations, a sufficiently converged FE profile is obtained. Thermodynamic integration is also used to estimate the FE using a series of constrained MD simulations [25]. In this case, the target molecule is fixed at a position on the reaction path (*i.e.* a fixed distance from the air/water interface), and the mean force exerted on the target molecule from the surrounding molecules is calculated during the constrained MD. The calculation is repeated on the different fixed points with the grid size of 0.2 nm. About 3 ns/constraint is usually enough to see the convergence for the CG model. In FE-MD, each degree of freedom of the system is coupled to a different Nosé–Hoover chain (massive Nosé–Hoover chain) [17,18] to enhance the sampling efficiency. By integrating the mean force along the reaction path, hydration FE is obtained. To calculate the transfer FE of a hexane molecule from its bulk to water, we take a similar protocol.

3. CG modelling and results

We show here how the CG model is built in a systematic way using the procedure described above. Our CG model is described by a central force model; each CG site is typically composed of three heavy atoms and their accompanying hydrogen atoms (e.g. figure 1). In table 1, the definition of the CG sites are given. Before reaching the current model, we have analysed the AA systems and tried several possible CG models. The illustration of our analyses is helpful to understand our choice of the CG model. Therefore, we put some of the details for water in the next subsection.

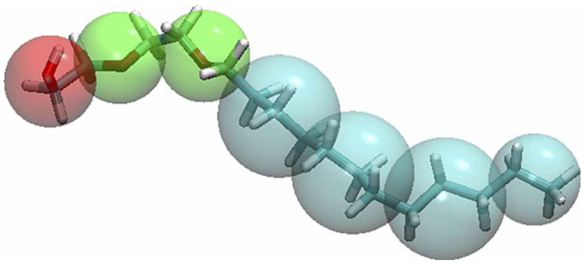

Figure 1. CG molecular model for $C_{12}E_2$. Transparent spheres represent CG particles. The molecular graphics are created with VMD [29].

Table 1. Segment name and definition.

Name	*All-atom*
W	$(H_2O)_3$
CT	$CH_3CH_2CH_2—$
CM	$—CH_2CH_2CH_2—$
CT2	$CH_3CH_2—$
EOT	$CH_3—O—CH_2—$
EO	$—CH_2—O—CH_2—$
OA	$HOCH_2—$

3.1. Water

As shown in table 1, the CG water "W" site represents three water molecules. With this choice, we first tried to build a CG model using the effective interaction of the AA model in the sprit of FM. Using a 1 ns-MD trajectory of 4080 TIP3P [30] water system, we analyse the energy between "blobs" of three water molecules. The water "blob" should consist of three neighbouring water molecules, and their representative position is taken at their centre of mass(COM). In other words, the three water molecules should be the three closest water molecules from their COM. Figure 2 plots the average interaction energy, $\langle U_{\mathrm{b-b}}\rangle$, between the water "blobs" as a function of the distance.

$$U_{\mathrm{b-b}} = \sum_{i\in \mathrm{blob}I}\ \sum_{j\in \mathrm{blob}J\neq I} U(r_{ij}) \tag{5}$$

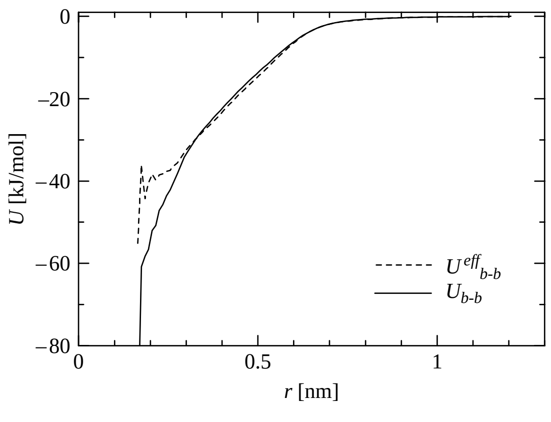

Figure 2. Effective interaction energy between water blobs calculated using AA-MD trajectory in a pure water system. The definitions for U are given in text.

The calculated interaction gives no repulsive force and grows more attractive as two "blobs" get closer. In this calculation, we cannot evaluate the energy loss due to the deformation of the water blob as a result of the close contact of two blobs. To take this into account, we also calculate the effective interaction energy using

$$U^{\mathrm{eff}}_{\mathrm{b-b}} = \sum_{i,j\in \mathrm{blob}\,I \mathrm{or}\, J} U(r_{ij}) - 2\left\langle U^{\mathrm{intra}}_{\mathrm{blob}}\right\rangle. \tag{6}$$

The first term includes the intra-blob interactions. By subtracting the average intermolecular energy within a water "blob", $\langle U^{\mathrm{intra}}_{\mathrm{blob}}\rangle$, the energy loss due to the structural correlation is evaluated. The effective interaction, which is also plotted in figure 2, shows a slight modification of the energy at a short distance, while it still gives no repulsive force. It should be noted that the attractive interaction energy is very strong at short range and rapidly decays, as a function of distance, to zero within ~1 nm. It is known that particles having short-ranged strong attractive interaction do not form a liquid state. Actually, the particles that interact via $U_{\mathrm{b-b}}$ give the solid phase at 303.15 K. In AA-MD, a water molecule takes a preferential orientation to make hydrogen bonds with the surrounding molecules at a contact distance. This causes a strong association on the blob–blob level. In this particular case, the entropic difference is so significant that we cannot make a reasonable model using a FM procedure. In addition, increasing the number of water molecules packed into one CG site makes the situation worse; the interaction becomes stronger and shorter-ranged compared with the particle size. FM is one of the most promising approaches to construct a CG model based upon AA-MD analysis; Izvekov and Voth have successfully applied FM to build a CG water particle that represent single water molecule [31]. Our analysis suggests that, however, the FM method is not effective when applied to the CG model composed of multiple water molecules.

To make an "effective potential" for CG water, we adopt the following guidelines. First of all, the LJ function is employed to keep the model simple. We design the model to be in the liquid state from 273 to 373 K and to have experimental density and surface tension at the target temperature of 303.15 K. For this purpose, Shelley's water model using the LJ6–4 potential actually works well. However, a possible problem with the model is that the compressibility is too high due to the soft repulsive potential. Since water is one of the major components in many applications including surfactants, this high compressibility may be problematic. This may affect the elastic property of vesicles, for example. To improve the compressibility, we have tried other functional forms for the W—W interactions, namely, LJ9–6, LJ8–4, LJ10–4 and LJ12–4. Except for the LJ9–6, we find a parameter set for each function to give liquid water with the correct surface tension and density. The steepest repulsion term of the LJ12–4 gives the best compressibility among them,

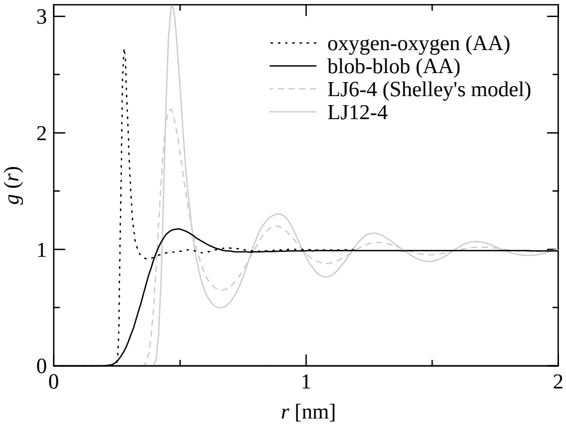

Figure 3. RDF for pure water systems. Black lines show AA-MD results; grey lines show CG-MD.

though it still overestimates the experimental compressibility by about 35% (Figure 4). We should note here that the compressibility is not a property compatible with structure. In general, to reproduce the same compressibility with a larger particle, we need a more repulsive, steeper pair-potential, because the larger cavity makes the fluid more compressible. On the other hand, a structural analysis of the grouping of molecules gives a totally different view. Figure 3 plots W—W radial distribution functions (RDFs). The blob–blob and oxygen–oxygen RDFs calculated using AA-MD trajectories are also plotted in the figure, demonstrating a remarkable structural change (softening) due to the coarse-graining. To reproduce the blob–blob RDF with a central force model, we need a very soft core potential. Even with the LJ6–4 function, which overestimates the compressibility three times higher, the model has too much structure. This is one of the inevitable consequences of coarse-graining, thus we have to make a selection. Here the W—W structure is beyond the scope of our CG model. Instead, we desire that the CG water model should sustain the volume properly as a major momentum carrier. Figure 4 plots density, surface tension, and compressibility as a function of temperature. As for the density and surface tension, the CG model shows a quite similar temperature dependence to the experiments. The compressibility shows a minimum at about 320 K, known as one of the abnormalities of water. The simplified CG model has no chance to reproduce this behaviour, though the new CG model with LJ12–4 function shows a significant improvement in compressibility. As we will see later, the new CG water model gives better interfacial properties at the boundary with an alkane.

3.2. Alkanes

Here, we follow a similar procedure to that used in previous work [16] on the alkanes. However, we have changed the cut-off length of the LJ9–6 function and ignored the tail correction. The parameters are fixed in a systematic calculation of a series of alkanes with different chain lengths. Again, we use surface tension and density data to fix the parameters assuming the LJ9–6 function. At first, we determine the parameters for a series of alkane chains with the carbon number of $3n (n = 2, 3, 4, 5, 6)$. We have two CG particle types, CT ($-CH_2CH_2CH_3$) and CM ($-CH_2CH_2CH_2-$), and three nonbonded interaction pairs (CT—CT, CM—CM, CT—CM) in these systems. The CT—CT interaction parameters are determined using hexane ($n = 2$). The other parameters are fixed using longer chains ($n = 3, 4, 5, 6$). The calculated surface tension and density show a good agreement with experimental values (table 2), suggesting that the CG model is accurate enough to characterize these thermodynamic quantities for different chain lengths of alkanes. A similar parameterization may be possible with a different functional form for the nonbonded interaction. The selection of the LJ9–6 function for alkane–alkane interactions is validated by checking the structural properties. For example, the RDF for the pure dodecane system are plotted in figure 5. Although a slight higher first peak is observed for CT—CT with the CG model, overall structure agrees reasonably with the AA results.

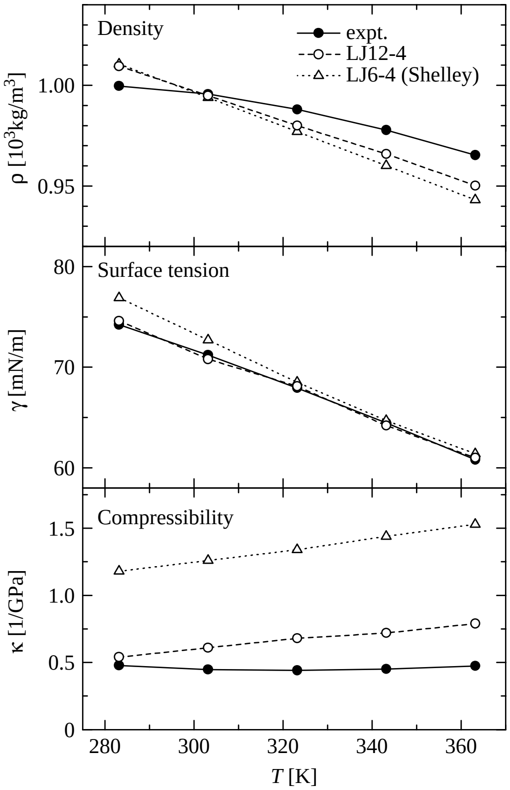

Figure 4. Temperature dependence of thermodynamic properties of CG models. Experimental data are taken from Ref. [33].

To treat chains of arbitrary length, a scaling formula for LJ parameters has been proposed [16]. The scaling rule also works very well for our alkane model. For the interaction with the other molecules, however, the same scaling rule is not useful. Efforts for a new scaling rule remind us that it is not easy to find a general scaling rule for the interactions. To give flexibility for the choice of the

Table 2. Surface tension and density of the CG model for pure component systems.

			$\gamma^{\dagger}$		$\rho^{\ddagger}$	
System	*Molecular structure*	*Interaction*	*Exp*[¶]	*MD*[§]	*Exp*[¶]	*MD*[‖]
Water						
Water	W	W—W	71.20	70.8	0.9957	0.9949
Alkanes						
Hexane	CT—CT	CT—CT	17.43	17.5	0.6518	0.6498
Nonane	CT—CM—CT	CM—CT, CM—CM	21.94	22.3	0.7114	0.7129
Dodecane	CT-(CM)$_2$-CT		24.48	24.5	0.7415	0.7422
Pentadecane	CT-(CM)$_3$-CT		26.23	25.9	0.7616	0.7603
Octadecane	CT-(CM)$_4$-CT		27.53	27.6	0.7722	0.7726
Heptane	CT2-CM-CT2	CT2–CT2, CT2–CM	19.27	19.4	0.6773	0.6791
Decane	CT2-(CM)$_2$-CT2		22.92	22.3	0.7247	0.7239
Oxyethylenes						
Dimethoxyethane	EOT—EOT	EOT—EOT	19.45	19.9	0.8593	0.8617
Diethyleneglycoldimethylether	EOT—EO—EOT	EOT—EO, EO—EO	28.60	25.6	0.9372	0.9374
Triethyleneglycoldimethlether	EOT—(EO)$_2$—EOT		27.83	29.8	0.9735	0.9804
Tetraethyleneglycoldimethylether	EOT—(EO)$_2$—EOT		32.88	31.7	1.0010	1.0060
Ethyleneglycols						
Ethylene glycol	OA—OA	OA—OA	49.01	50.2	1.1070	1.1060
Diethyleneglycol	OA—EO—OA	EO—OA	48.86	44.8	1.1100	1.0990
Triethyleneglycol	OA—(EO)$_2$—OA		45.80	45.2	1.1180	1.1150
Tetraethyleneglycol	OA—(EO)$_3$—OA		43.53	45.0	1.1170	1.1200
Ethers with long alkane chains						
Diethyleneglycoldi-n-butylether	CT—(EO)$_3$—CT	EO—CT	26.07	26.3	0.8774	0.8767
Dipropylether	CT2—EO—CT2	EO—CT2	19.46	19.1	0.7366	0.7379
Di-n-hexyl-ether	CT2—CM—EO—CM—CT2	EO—CM	24.91	24.9	0.7860	0.7858
Alcohols						
1-propanol	CT2—OA	CT2—OA	23.80	22.8	0.7950	0.7943
1-hexanol	CT2—CM—OA	CM—OA	25.48	25.7	0.8123	0.8121

[†] Surface tension in mN/m.
[‡] Density in 10^3kg/m^3.
[¶] Experimental data are taken from Ref.[32].
[§] The average value is taken from at least 5 ns-NVT-MD for the slab system. The standard error is about 1.0 mN/m.
[‖] The average value is taken from 1 ns-NPT-MD for the bulk system. The typical error is less than 0.001×10^3 kg/m^3.

alkane chain length, a new CG particle, CT2 (—CH$_2$CH$_3$), is adopted. To fix the parameters for CT2—CT2 and CT2—CM, heptane (CT2—CM—CT2) and decane (CT2—CM—CM—CT2) are used. (see table 2) By the combination of CT, CM, CT2, we can build a variety of alkane chains.

3.3. *Alkane/water interface*

For the two different molecular fluids showing a phase separation, the interfacial tension is useful to determine the interaction parameters between these two molecules. Alkane/water is the typical case. Similar to the case of pure alkane, a systematic parameterization is carried out for a series of alkanes starting from the hexane/water interface. Since we already have two functional forms for nonbonded interaction, the LJ12–4 and LJ9–6, we have to select the function for the W—CT interaction. Our current choice is the LJ12–4, though the LJ9–6 does not give a significant difference in the results. Taking hexane/water as an example, some important results will be shown here. The interfacial width is one of the important properties to characterize the interface. Due to the larger size of the CG particle, it is inevitable for

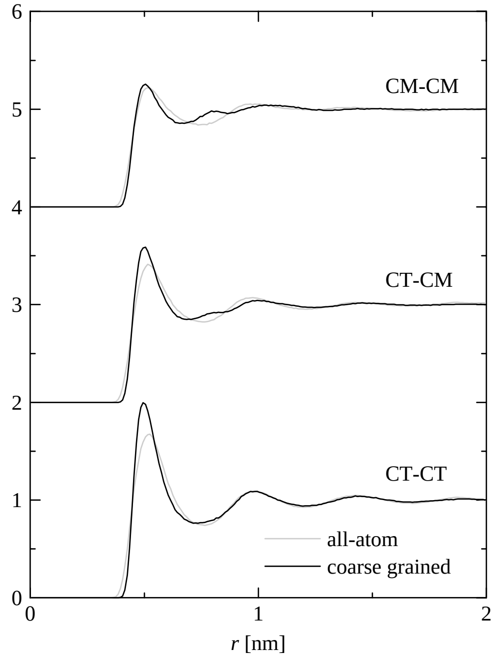

Figure 5. RDF in the pure dodecane system.

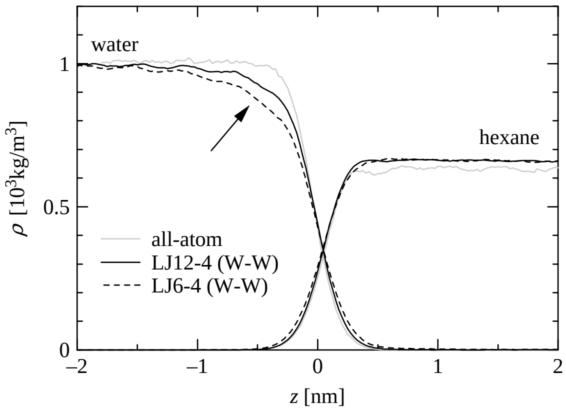

Figure 6. Density profiles across the hexane/water interface.

Table 3. Calculated interfacial tension for the alkane/water interface with the CG model.

		$\gamma^\dagger$	
Interface	*Interaction*	*Exp*‡	*MD*¶
Water/hexane	CT-W	49.96	50.0
Water/nonane		51.21	51.9
Water/dodecane	CM-W	52.14	52.9
Water/pentadecane		–	52.9
water/heptane	CT2-W	50.30	50.1

† Interfacial tension in mN/m.
‡ Experimental data are taken from Ref.[34].
¶ The average value is taken from at least 10 ns-NP$_n$AT-MD. The standard error is less than 1.0 mN/m.

the interface to broaden to some extent. Figure 6 plots the density profiles of water and hexane along the normal direction of the interface, z. AA-MD underestimates the hexane bulk density and overestimates the water density slightly. Taken this into account, the density profile of the CG hexane with the LJ12–4 function shows good agreement with that of AA-MD. However, we actually see the broadened decay in the CG water density profile near the interface. The situation is worse with the softer CG water potential (LJ6–4 for W—W). Interestingly, the broadened interface is not sensitive to the functional form for W—CT but for W—W. Thus, even from the viewpoint of the interfacial width, the LJ12–4 is preferable for the W—W interaction. Even though we select a functional form for this interaction, we have two unknown parameters to be fixed with one property, i.e. the interfacial tension here. As for σ, which represents a contact distance between the particles, we take the arithmetic average between the alkane particle and W. Although this assumption does not always work for CG models, in this particular case, it is shown that our parameters work well even with respect to solubility. Figure 7 shows the FE profile of a hexane molecule across the hexane/water interface calculated from 15 independent SMD simulations using Jarzynski's theorem. A large FE barrier is observed for hexane to enter the water phase near the interface, though a plateau in the FE profile is obtained deep in the water region. The FE value at the plateau gives the transfer FE of hexane from its bulk to water. The transfer FE is estimated to be 34 ± 2 kJ/mol for the CG model with the LJ12–4 function, which is in good agreement with an experimental value of 32.4 kJ/mol [35]. Again, the LJ9–6 for the CT—W interaction gives no difference in the transfer FE, though the LJ6–4 water gives the smaller transfer FE (figure 7).

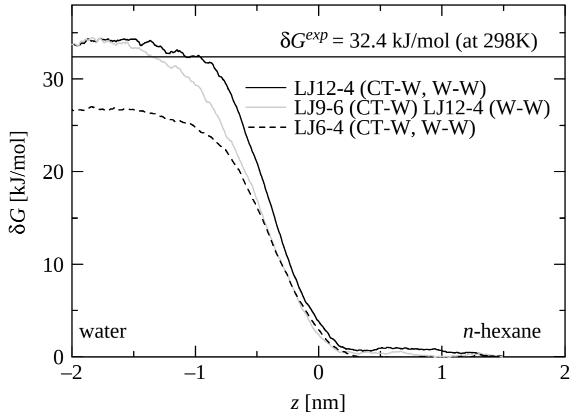

Figure 7. FE profile of a hexane molecule across the hexane/water interface.

3.4. PEG

The same protocol that we used for alkanes is useful to determine the interaction parameters in PEG. For the versatility, in this parameter fitting, oligo-oxyethylenes, i.e. methyl-terminated ethylene glycols, are also included. Again, all parameters for bonded interactions are determined to reproduce the same probability distribution of AA-MD. This demands a lot of computation with AA detail. However, a careful mapping is indeed required for the bond parameters because the bond interaction affects the effective molecular size, which plays an important role in characterizing thermodynamic properties of the molecular fluid. Then, assuming the LJ9–6 functional form, ε and σ are determined using surface tension and density. To parameterize the interaction between hydrophobic groups and ethylene glycols, a series of ethers and alcohols are also used. The CG-MD results for this mapping are listed in table 2, showing that the parameter fitting has been successfully done.

3.5. PEG/W interaction

Although interfacial tension is used to fix the parameters between water and alkanes, this approach is useful only for molecules or groups that make a well defined interface with water. Actually, ethylene glycols have high solubility in water. Thus, we need an alternative approach to fix the ethylene glycol/water interaction parameters.

One possible systematic way is to use the hydration FE data as a target property to be reproduced in the CG model. However, since we have two unknown parameters to be fixed, ε and σ, we need one more property to fix the parameters uniquely. Although the combination rule for σ may be used, we take a different approach to estimate the σ

value because the effective size of a CG site in water will change depending on the hydrophilicity and may be different from that in their bulk solution. To estimate the effective size of a CG site, we use an AA-MD simulation of the molecule (or fragment of molecule) corresponding to the CG site in bulk water. For example, to estimate σ for the OA site, AA-MD of —CH_2OH in TIP3P water has been performed. By analysing the RDF of water "blobs" (defined in Section 3.1) around the solute (—CH_2OH), a potential of mean force (PMF) is obtained. The PMF is not straightforwardly useful, though the shortest distance that the PMF reaches zero may give a reasonable guess for σ. Analysis of alkane segments (CT and CM) in water shows that the distance is almost in accord with the σ that is determined from surface tension and density data, though a slight shift by 0.017 nm is needed for the perfect agreement. Thus, after σ is fixed in this approach, ε is determined to reproduce the hydration FE. We need a series of FE calculations of hydration of the target solute to find a suitable parameter for ε. Although this sounds computationally demanding, it does not require so much. For example, only about 70 CPU hours are typically needed to estimate hydration FE for a CG dimer with the precision of ± 4 kJ/mol using a single processor (Pentium IV, 2.8 GHz). This approach is generally useful for a variety of molecules as long as the experimental hydration FE data is available. Thus, a systematic parameterization for a series of CG segments is now feasible. Here we fix the parameter for OA—W with this protocol.

Table 4. Bond parameters (1–2 interaction).

Atoms		k_b†	r_0‡
CT	CT	6.955	3.71
CM	CM	6.160	3.64
CM	CT	6.160	3.65
CT2	CM	9.000	3.13
OA	OA	63.000	2.16
EOT	EOT	5.500	3.33
EOT	EO	5.400	3.34
EO	EO	4.900	3.28
EO	OA	15.000	2.79
CT	EO	7.100	3.61
CM	EO	7.100	3.56
CT2	EO	10.000	3.07
CT	OA	6.955	3.05
CM	OA	7.500	3.01
CT2	OA	14.000	2.53

† Harmonic spring constants in kcal/mol/Å^2 (= 418.4 kJ/mol/nm^2).
‡ The bond length at the energy minimum in Å(= 0.1 nm).

To map out the PEG surfactant/water system in the CG description, we still lack the EO—W interaction. Again, σ is estimated from the effective size of the EO segments in water from AA-MD. Since the experimental hydration FE data useful for this purpose is not available, we have used structural data of the lamellar phase of the $C_{12}E_2$/water system [36]. Since the lamella spacing and the molecular area of $C_{12}E_2$ are available from the X-ray diffraction measurement, we fix the parameter using these quantities. A series of NP_nAT-MD simulations of the lamellar (bilayer) systems at the surfactant composition of 67 wt% have been carried out in the simulation box that has the cross-sectional area to reproduce the experimental molecular area of 0.30 nm^2. Now we need to find the ε value to give zero surface tension. This is essentially the same as we have done for pure molecular systems (i.e. surface tension mapping). After the parameterization, we have carried out a 10 ns-NPT-MD simulation of the $C_{12}E_2$/water system to assess the membrane properties. The average molecular area shows perfect agreement with the experimental value, 0.300 nm^2, while the lamellar spacing, 4.81 nm, is slightly overestimated, compared with the experiment (4.73 nm). Figure 8 plots the number density for each CG segment along the bilayer normal. A comparison with 10 ns-AA-MD in the NP_nAT ensemble is also made, showing reasonable agreement with the probability distribution. Taking into account the fact that we have not used structural data except for the molecular area of the surfactant to make the CG model, the agreement is fairly good. All interaction parameters are summarized in tables 4–6.

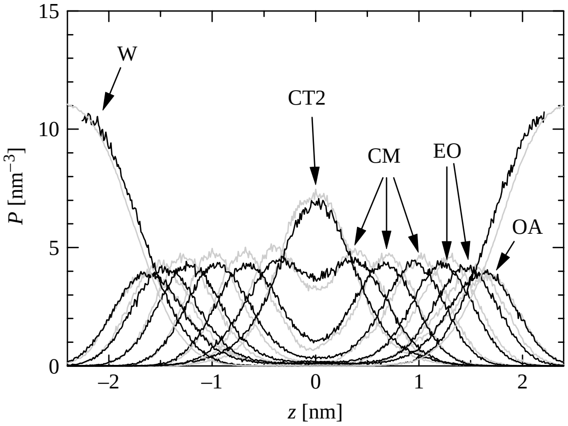

Figure 8. Density profile of the CG segments along the $C_{12}E_2$ bilayer normal. CG-MD and AA-MD results are given in black and grey lines, respectively.

Table 5. Angle parameters (1–2–3 interaction).

Atoms			k_a†	θ_0‡
CM	CM	CM	1.190	173.0
CM	CM	CT	1.190	175.0
CT2	CM	CM	1.600	172.0
CT2	CM	CT2	1.700	173.0
CT	CM	CT	1.093	175.5
EO	EO	EOT	3.500	135.0
EO	EO	EO	3.400	132.0
EOT	EO	EOT	3.500	134.0
OA	EO	OA	6.600	131.0
EO	EO	OA	3.000	131.0
CT	EO	EO	2.200	145.0
CT2	CM	EO	1.300	178.0
CM	EO	CM	1.800	171.0
CT2	EO	CT2	2.600	165.0
OA	EO	CT2	4.000	146.0
CM	CM	EO	1.500	172.0
CM	EO	EO	3.200	146.0
CT2	CM	OA	1.800	166.0

† Harmonic spring constants in kcal/mol/rad^2 (= 4.184 kJ/mol/rad^2).
‡ The angle at the energy minimum in degree.

Table 6. Nonbonded interaction parameters; ε is in kcal/mol($= 4.184$ kJ/mol), σ and R_c are in Å ($= 0.1$ nm).

Atoms		*Function*	ε	σ	R_c
W	W	LJ12–4	0.8950	4.3710	15
CT	CT	LJ9–6	0.4690	4.5850	15
CT	CM	LJ9–6	0.4440	4.5455	15
CM	CM	LJ9–6	0.4200	4.5060	15
W	CT	LJ12–4	0.3600	4.4780	15
W	CM	LJ12–4	0.3400	4.4385	15
CT2	CT2	LJ9–6	0.3120	4.2210	15
CT2	CM	LJ9–6	0.3620	4.3635	15
W	CT2	LJ12–4	0.2900	4.2960	15
OA	OA	LJ9–6	0.4491	3.7130	15
EOT	EOT	LJ9–6	0.4370	4.2500	15
EO	EO	LJ9–6	0.4050	4.2500	15
EO	EOT	LJ9–6	0.4200	4.2500	15
EO	OA	LJ9–6	0.4400	3.8900	15
EO	CT	LJ9–6	0.4100	4.3400	15
EO	CM	LJ9–6	0.3770	4.2740	15
EO	CT2	LJ9–6	0.3700	4.1400	15
W	OA	LJ12–4	0.7000	3.9500	15
W	EO	LJ12–4	0.5700	4.3100	15
CT	OA	LJ9–6	0.4372	4.0330	15
CM	OA	LJ9–6	0.3650	3.9870	15
CT2	OA	LJ9–6	0.3800	3.8400	15

4. Application: self-assembly to lamellar structure

One of the goals of MD using the CG molecular model is to pursue nonequilibrium molecular processes that take place on time-scales not accessible by AA-MD. A self-organized mesostructure of amphiphiles is one such example. We demonstrate here an application of our new CG potential to observe the self-assembling process of $C_{12}E_2$ molecules into the lamellar structure in water. The initial configuration is made with 1296 $C_{12}E_2$ molecules and 3456 W particles randomly packed into the cubic simulation box with the edge of about 10 nm. Lamellar formation has occurred in a 10 ns-CG-MD run (figure 9). The initial stage of the structural formation is a local rearrangement of molecular structure to reduce the contact area between hydrophilic and hydrophobic components. This process proceeds in a short time period, <0.5 ns, and brings about a large energy gain. A cubic-like structure is observed at 1 ns, showing a bicontinuous bilayer/water domain formation. At this stage, the hydrophobic core has been almost completely covered with hydrophilic components. After 4 ns, a lamellar frame shows up, though the two bilayers are still interconnected, showing a spiralled continuous membrane. The interconnection of the two bilayers is gradually thinning. This process takes place relatively slowly, which allows a symmetric partitioning of surfactant molecules into two adjacent bilayers; the resultant four layers (two bilayers) of the surfactants consist of 325, 327, 323 and 321 $C_{12}E_2$ molecules, respectively. Even at 7 ns, a water channel penetrating the membrane is observed in each bilayer, and these two channels disappear at the same time, 8.3 ns. A similar long-lasting water pore, during the self-assembling process of bilayer-forming lipids, has been observed in an AA-MD simulation previously [37], though the system may be too small to observe the interconnection among adjacent bilayers. In the present case, the water channels always coexist with a surfactant tube interconnecting two bilayers. After the 10 ns NVT-MD simulation, another 10 ns MD simulation has been carried out in the NPT ensemble to have fully relaxed bilayers. Again, we confirm that the average molecular area for a $C_{12}E_2$ molecule, 0.301 nm^2, showing excellent agreement with the experimental value [36].

5. Conclusions

A new CG model has been generated for PEG surfactants. Our step-by-step approach to construct the CG model is significantly different from previous methods [7–9]. By using several small molecular systems for the parameter fitting, the CG model has versatility and transferability. The intramolecular bond potentials are determined to reproduce the AA bond and angle distributions based on AA-MD data. A careful choice of the functional forms for water–water nonbonded interactions improves the compressibility and the structure at the boundary with alkanes. In most cases, the density and surface/interfacial tension are used as target properties for the fitting, because these are of key importance in characterizing the self-organized surfactant structure in the mesoscale that is the plausible target of CG-MD. Solvation (hydration) and transfer free energies, which play an essential role in determining of the partition of solute molecules, are also taken into account in our model. This approach should also be useful for the CG mapping of a variety of biological molecules, including peptides, cholesterol, and drugs. Our approach

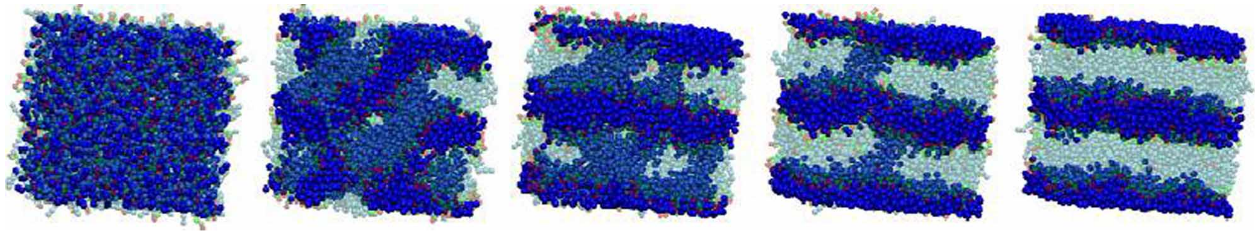

Figure 9. Snapshots of lamellar formation of the $C_{12}E_2$/water system from a random initial configuration. The time stamps are 0, 1, 4, 7 and 10 ns from left to right. Water is depicted in blue solid particles, whereas the $C_{12}E_2$ molecules are drawn in transparent representation. (see figure 1) The molecular graphics are created with VMD[29].

is useful for a general purpose and will open a way to build a systematic CG model.

Acknowledgements

We thank NSF for generous support. We are grateful to Steve Nielsen and Preston Moore for helpful discussion. R.D. is supported by a NSF Postdoctoral Fellowship under Grant No. DBI-0532800.

References

[1] W. Tschop, K. Kremer, J. Batoulis, T. Burger, O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.*, **49**, 61 (1998).

[2] F. Muller-Plathe. Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. *Chem. Phys. Chem.*, **3**, 754 (2002).

[3] R.D. Groot, K.L. Rabone. Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. *Biophys. J.*, **81**, 725 (2001).

[4] R. Goetz, G. Gompper, R. Lipowsky. Mobility and elasticity of self-assembled membranes. *Phys. Rev. Lett.*, **82**, 221 (1999).

[5] M. Venturoli, B. Smit, M.M. Sperotto. Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins. *Biophys. J.*, **88**, 1778 (2005).

[6] S.O. Nielsen, C.F. Lopez, G. Srinivas, M.L. Klein. Coarse grain models and the computer simulation of soft materials. *J. Phys. Condens. Matter*, **16**, R481 (2004).

[7] S.J. Marrink, A.H. de Vries, A.E. Mark. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, **108**, 750 (2004).

[8] J.C. Shelley, M.Y. Shelly, R.C. Reeder, S. Bandyopadhyay, M.L. Klein. A coarse grain model for phospholipid simulations. *J. Phys. Chem. B*, **105**, 4464 (2001).

[9] S. Izvekov, G.A. Voth. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, **109**, 2469 (2005).

[10] G. Srinivas, J.C. Shelley, S.O. Nielsen, D.E. Discher, M.L. Klein. Simulation of diblock copolymer self-assembly, using a coarse-grain model. *J. Phys. Chem. B*, **108**, 8153 (2004).

[11] H. Fukunaga, J. Takimoto, M. Doi. A coarse-graining procedure for flexible polymer chains with bonded and nonbonded interactions. *J. Chem. Phys.*, **116**, 8183 (2002).

[12] B. Smit, P.A.J. Hilbers, K. Esselink, L.A.M. Rupert, N.M. Vanos, A.G. Schlijper. Computer-simulations of a water oil interface in the presence of micelles. *Nature*, **348**, 624 (1990).

[13] Q. Sun, R. Faller. Systematic coarse-graining of a polymer blend: polyisoprene and polystyrene. *J. Chem. Theory Comput.*, **2**, 607 (2006).

[14] R.L. McGreevy. RMC: progress, problems and prospects. *Nucl. Instrum. Meth. Phys. Res. A*, **354**, 1 (1995).

[15] A.P. Lyubartsev. Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.*, **35**, 53 (2005).

[16] S.O. Nielsen, C.F. Lopez, G. Srinivas, M.L. Klein. A coarse grain model for n-alkanes parameterized from surface tension data. *J. Chem. Phys.*, **119**, 7043 (2003).

[17] S. Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, **81**, 511 (1984).

[18] G.J. Martyna, M.L. Klein, M. Tuckerman. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J. Chem. Phys.*, **97**, 2635 (1992).

[19] M. Tuckerman, G.J. Martyne, B.J. Berne. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, **97**, 1990 (1992).

[20] W. Shinoda, M. Mikami. Rigid-body dynamics in the isothermal-isobaric ensemble: a test on the accuracy and computational efficiency. *J. Comput. Chem.*, **24**, 920 (2003).

[21] A.D. MacKerell Jr, D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586 (1998).

[22] K. Shinoda, W. Shinoda, T. Baba, M. Mikami. Comparative molecular dynamics study of ether- and ester-linked phospholipid bilayers. *J. Chem. Phys.*, **121**, 9648 (2004).

[23] T. Yoshida, W. Shinoda, S. Tsuzuki, M. Mikami. unpublished.

[24] G.J. Martyna, M.E. Tuckerman, D.J. Tobias, M.L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, **87**, 1117 (1996).

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

[26] G.J. Martyna, D.J. Tobias, M.L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, **101**, 4177 (1994).

[27] S. Park, K. Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.*, **120**, 5946 (2004).

[28] C. Jarzynski. *Phys. Rev. Lett.*, **78**, 2690 (1997).

[29] W. Humphrey, A. Dalke, K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33 (1996).

[30] W.L. Jorgensen, J. Chandrashekar, J.D. Madura, R. Impey, M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926 (1983).

[31] S. Izvekov, G.A. Voth. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.*, **123**, 134105 (2005).

[32] C.L. Yaws. *Chemical Properties Handbook*, McGraw-Hill. Online version available at: http://www.knovel.com/knovel2/Toc.jsp?BookID = 49 (1999).

[33] D.R. Lide (Ed.). *CRC Handbook of Chemistry and Physics*, Chapman&Halls/CRC (2002).

[34] S. Zeppieri, J. Rodriguez, A.L.L. de Ramos. Interfacial tension of alkane + water systems. *J. Chem. Eng. Data*, **46**, 1086 (2001).

[35] S.H. White, W.C. Wimley. Membrane protein folding and stability: physical principles. *Ann. Rev. Biophys. Biomol. Struct.*, **28**, 319 (1999).

[36] S.S. Funari, G. Rapp. X-ray studies on the $C_{12}EO_2$/water system. *J. Phys. Chem. B*, **101**, 732 (1997).

[37] S.J. Marrink, E. Lindahl, O. Edholm, A.E. Mark. Simulation of the spontaneous aggregation of phospholipids into bilayers. *J. Am. Chem. Soc.*, **123**, 8638 (2001).