

On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties

Kevin R. Hadley[†] and Clare McCabe^{*,‡}

Department of Chemical and Biomolecular Engineering and Department of Chemistry,
Vanderbilt University, Nashville, Tennessee 37235-1604

Received: December 16, 2009; Revised Manuscript Received: February 8, 2010

Developing accurate models of water for use in computer simulations is important for the study of many chemical and biological systems, including lipid bilayer self-assembly. The large temporal and spatial scales needed to study such self-assembly have led to the development and application of coarse-grained models for the lipid–lipid, lipid–solvent, and solvent–solvent interactions. Unfortunately, popular center-of-mass-based coarse-graining techniques are limited to modeling water with one water per bead. In this work, we have utilized the K-means algorithm to determine the optimal clustering of waters to allow the mapping of multiple waters to single coarse-grained beads. Through the study of a simple mixture between water and an amphiphilic solute (1-pentanol), we find a four-water bead model has the optimal balance between computational efficiency and accurate solvation and structural properties when compared to water models ranging from one to nine waters per bead. The four-water model was subsequently utilized in studies of the solvation of hexadecanoic acid and the structure, as measured via radial distribution functions, for the hydrophobic tails and the bulk water phase were found to agree well with experimental data and their atomistic target.

1. Introduction

Water is unquestionably the most common solvent in experimental and computational studies,¹ particularly for biological systems. It serves as the reference fluid for many properties (e.g., specific gravity is measured relative to water density, and heat capacity was originally defined in terms of the heat needed to raise the temperature of water) and is the most abundant chemical on earth. Water also has many unique properties, such as expansion upon freezing and a high surface tension, which complicates its dynamics and physics compared to other solvents.² Both its ubiquitous use and unusual properties make water an interesting and challenging system to study computationally and has led to the development of many different models to describe the properties and behavior of water (see, for example, the review of Guillot³ for water models proposed up to 2002 and the more recent work of Vega et al.⁴).

The models for water essentially vary in terms of their choice of interaction potential and purpose. For example, commonly used water models, such as TIP3P and SPC, are three site models with rigid bonds and angles, and point charges on the oxygen and hydrogens of the water to describe polarity. While these models reproduce many experimental properties such as enthalpy and liquid densities, other properties such as self-diffusion coefficients are not well reproduced.⁵ More recent models have included additional complexity to improve the ability of the model to reproduce other properties, such as the density maximum at 4 °C and isothermal compressibilities, by incorporating flexible bonds and angles,⁶ polarizability,⁷ and delocalized charges;^{8,9} however, more detail does not always lead to increased accuracy. For example, Hess and van der Vegt¹⁰ compared the SPC, SPC/E, TIP3P, TIP4P, and TIP4P-

Ew water models to study the solvation of amino acids and found accuracy in reproducing experimental thermodynamic properties did not necessarily correlate with the complexity of the water model.¹⁰ In particular, it was found that the more complex TIP4P-Ew model underestimated the hydration heat capacity compared to the SPC and TIP3P models, which were in good agreement with the experimental values; overall, however, the SPC/E model,¹¹ which, similar to SPC and TIP3P, is a three-site, rigid, nonpolarizable model, was found to be most accurate over the widest range of structural and thermodynamic properties.

An alternative approach to the development of increasingly realistic water models is to explore the ability of very simple models to predict certain features of the properties of water. For example, work done by Nezbeda and co-workers¹² has focused on the so-called primitive model, which describes water as a spherical segment with an attractive square-well potential to model hydrogen-bonding interactions and a hard sphere potential for all other interactions. Although the Nezbeda model is not suited to molecular dynamics simulations because of the discontinuous nature of the interaction potential, it is computationally efficient and able to quantitatively describe the PVT behavior of water.¹² In a similar vein, Dill and co-workers developed the Mercedes-Benz¹³ water model in which water is described as a planar Lennard-Jones disk with three orientation-dependent hydrogen-bonding arms. While the Mercedes-Benz model can capture, at least qualitatively, a number of the anomalous properties of water, the original version is only a two-dimensional model. A three-dimensional version has been developed by Bizjak et al.;¹⁴ however, implementation within common open source codes is difficult due to the orientation-based hydrogen-bonding potential.

Although these models, compared to TIP3P⁸ and SPC¹⁵ water models, for example, involve simple interactions, they are still very computationally expensive when used in studies of biological systems that involve large quantities of water to be

* Corresponding author. E-mail: c.mccabe@vanderbilt.edu. Phone: 615 322 6853. Fax: 615 343 7951.

[†] Department of Chemical and Biomolecular Engineering.

[‡] Department of Chemistry.

simulated over long time scales. As a result, an additional class of water model has been developed for use in coarse-grained, rather than atomistic, simulations,¹⁶ in which each bead or site in the model represents the collective interactions of a group of atoms. Coarse-grained models are typically developed using center-of-mass-based methods, in which the trajectory from an atomistic simulation is mapped to the coarse-grained level and the center of mass of the atoms within each bead determines the coordinates of that coarse-grained bead. By fitting the interactions of the coarse-grained model to reproduce specific target properties or aspects of the atomistic system, an accurate model on the coarse-grained level can be derived. While several different center-of-mass-based techniques, including force-matching,¹⁷ reverse Monte Carlo,¹⁸ and the iterative method developed by Reith, Pütz, and Müller-Plathe (RPM),^{19,20} can be used to develop coarse-grained models, they all result in numerical potentials for the coarse-grained model interactions.

For typical solute molecules, mapping schemes develop simply and naturally from the covalently bonded structure of the molecule. Subsequently, one can readily produce a target coarse-grained trajectory from an atomistic trajectory based on the center of mass of the heavy atoms within each coarse-grained bead. While for computational efficiency and consistency, one would like to map several water molecules to a single bead, center-of-mass-based methods are limited to one water molecule per bead due to the loose association of water molecules through hydrogen-bonding interactions. Therefore, while three or four water molecules may be associating in an atomistic simulation and could be assigned to a single bead centered on the center of mass of the clustered water molecules, the waters are only loosely associated and so will move somewhat independently of each other during the atomistic simulation, requiring reassignment of the waters to different coarse-grained beads. Without an efficient method to dynamically reassign atomistic waters to the coarse-grained beads, to date, center-of-mass-based methods map water on a 1:1 basis, thus limiting the improvement in computational efficiency obtained through coarse-graining. We note that, while coarse-grained models in which multiple waters are represented by a single coarse-grained bead have been developed, for example, by Marrink²¹ and Klein,²² such models are not compatible with center-of-mass-based coarse-graining methods (i.e., an atomistic trajectory cannot be mapped to the coarse-grained level and then used to derive target properties for parametrization of the cross-interactions with solute molecules).

In this work, in order to develop an efficient coarse-grained model for water with multiple water molecules mapped to one bead, we utilize a clustering algorithm, called the K-means algorithm.²³ Several degrees of coarse-graining have been studied and tested to determine the optimum balance between computational efficiency and accuracy. In sections 2–4, we describe the new coarse-grained water model, provide details of the simulations performed to develop and test the model, and then discuss the general strategy and methodology adopted to develop and parametrize the model. In section 5, we present the results for the different multiwater models, consider the most appropriate level of coarse-graining, and then apply the chosen model to a simple system of biological interest.

2. Coarse-Grained Model and Force Field Development

All coarse-grained models are designed to retain key features from the atomistic simulations on which they are based, typically at the cost of accuracy in other properties. We are primarily interested in developing a coarse-grained water model to be used

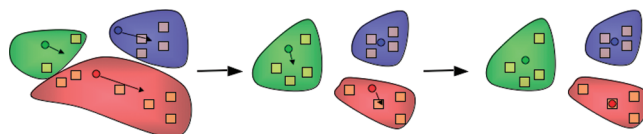


Figure 1. Schematic illustration of the K-means algorithm. Circles represent cluster locations, squares represent water locations, and shaded regions represent the allocation of waters to each cluster in a color-coded fashion.

in biological simulations of self-assembly; therefore, the model will be optimized to retain structural features, rather than focus on properties such as phase behavior or transport properties. To aid in the model development, atomistic simulations have been performed for pure water, selected pure solutes, and solute/water mixtures, and the trajectories mapped to the coarse-grained level, using the center of mass of the atoms contained within a specified site or bead to define the location of the coarse-grained bead, as described above in Section 1. To ensure the structural behavior of water and the solutes are retained on the coarse-grained level, radial distribution functions (RDF) from the atomistic trajectory mapped to the coarse-grained level serve as the target properties for the optimization. The coarse-graining method of Reith, Pütz, and Müller-Plathe (RPM)¹⁹ has been used to determine the coarse-grained force field, which iteratively optimizes the interactions until the coarse-grained RDF matches its target.

In order to develop a coarse-grained model for water that maps multiple water molecules to one bead, and so is on an equal footing with typical coarse-grained beads that contain three to five heavy atoms per bead, we have explored the use of the K-means algorithm.^{23,24} The K-means algorithm finds the optimal grouping of a large number of data points, which, in our application, corresponds to the coordinates of the atomistic water molecules. The K-means procedure locates which waters (the data points) are clustered together and determines the coordinates for the cluster (the location of the coarse-grained bead). As such, the K-means algorithm allows for a dynamic mapping scheme; i.e., it allows for the allocation of specific waters to a coarse-grained bead to change from frame to frame of the atomistic trajectory. The algorithm is schematically illustrated in Figure 1; the first step requires allocation of the number of clusters, k , to be used. For our model development, k is equal to the number of beads used to model water on the coarse-grained level and so relates to the degree of coarse-graining of the water. As shown in the example in Figure 1, if we map 4 waters (the squares) to each bead (the circles), a system with 12 waters would contain three beads (i.e., k equals three). The next step is to determine an initial location for each of the k clusters, which is chosen from the coordinates of random water oxygen atoms found within the simulation. After initializing the positions of the beads, each data point (the location of water oxygens) is allocated to the cluster with the smallest distance between the cluster and data point. In the example provided, two waters would be allocated to the green cluster, six waters grouped within the red cluster, and four waters placed in the blue cluster. Once the allocation is determined, the center of mass of the waters within a cluster is calculated as the new coordinates for that cluster, as indicated by the arrows in Figure 1. These steps are repeated until the termination criterion

$$\sum_{n=1}^k (r_n^{i+1} - r_n^i)^2 < \text{tol} \quad (1)$$

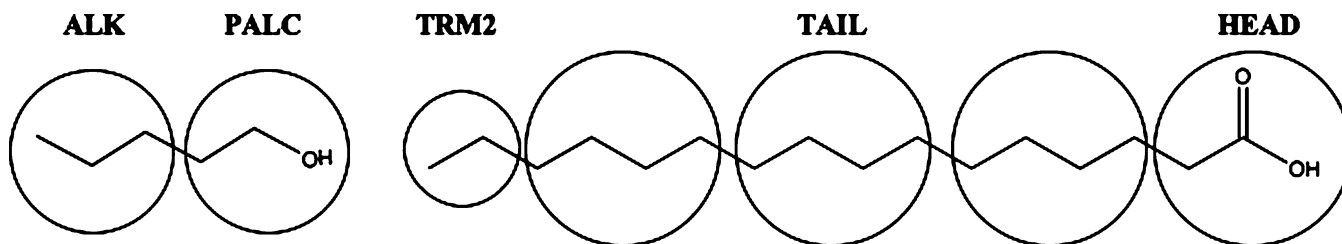


Figure 2. Schematic illustrating the mapping of the atomistic 1-pentanol (left) and C16:0 (right) to the coarse-grained level.

has been satisfied, where r_n^i represents the location of bead n at iteration i and tol is the tolerance set by the user (0.01 Å was used in this work). In the example provided, convergence is achieved in two iterations. For each subsequent frame of the trajectory, the locations of the clusters from the previous frame are used as the initial guess for the next frame in order to reproduce a more continuous target trajectory.

Using the K-means algorithm, a user-defined number of water molecules can be assigned to each bead. Thus, a goal of this work is to determine the degree of coarse-graining that provides an optimum balance between accuracy and computational efficiency. Water models containing 1, 3–6, 8, and 9 atomistic waters mapped to each bead (denoted H2OX, where X is the number of waters mapped to each bead) were parametrized and studied through simulations of pure water and simple mixtures of water and amphiphilic solutes; 1-pentanol was chosen as a representative amphiphilic solute because of its size and simplicity, and hexadecanoic acid was chosen as a model lipid. The coarse-grained mapping scheme for both molecules is shown in Figure 2. Pentanol is mapped to two beads, with PALC representing the hydrophilic region of the molecule containing the alcohol group, and ALK the hydrophobic alkane portion. For hexadecanoic acid, three bead types are used following earlier work:²⁵ HEAD to represent the acid headgroup, TAIL for the beads in the hydrocarbon tail, and TRM2 for the terminal bead that contains the last two carbons in the hydrocarbon tail. It is anticipated that, if the water model can properly solvate these coarse-grained molecules as mapped in Figure 2, it should provide the correct solvation and necessary driving forces for self-assembly in more complex biological systems.

3. Simulation Details

All simulations, atomistic and coarse-grained, were performed within the open source molecular dynamics program, DL_POLY 2.14.²⁶ The CHARMM force field²⁷ was utilized for the solutes due to its accuracy with respect to biological molecules, and TIP3P⁸ was used as the model for water, since solvation within the CHARMM force field is based on this water model. The Nosé–Hoover²⁸ algorithm was used for temperature and pressure control as needed and the reaction field method²⁹ employed to calculate the electrostatic interactions.

The pure TIP3P water simulations initially contained 901 water molecules and the pure solute systems contained 100 molecules, each at a density comparable to their experimental values under the same state conditions. The mixture systems were equiatom; 1-pentanol and water contained 75 solute molecules and 452 solvent molecules, and the acid mixture contained 50 solute molecules and 833 waters. All systems were equilibrated for 500 ps to 1 ns, first in the NVT ensemble and then in the NPT ensemble, to verify that the correct density was obtained.³⁰ The production runs for pure water and pure pentanol were then further simulated in both ensembles, while the mixture systems were studied only in the NVT ensemble

as the RPM pressure correction process was not required (since the pure potentials were previously pressure-corrected). All simulations were then run using a 1.0 fs time step for an additional 1.0 ns, which was found to produce enough sampling for these system sizes as defined by minimal changes in the target radial distribution functions.

4. Methodology

In previous work,²⁵ a coarse-grained model for fatty acids was developed using the RPM method to fit atomistic target RDFs based on the mapping shown in Figure 2. These potentials were used for the hexadecanoic acid studied in this work and as starting points in the optimization of the 1-pentanol model (i.e., the HEAD potential was used as an initial guess for the PALC potential and the TAIL bead served as the starting potential for the 1-pentanol ALK interaction). Studies have shown that the choice of initial potential has a negligible effect on the final optimized potential but can reduce the number of iterations required for convergence.³¹ For the water models, the HEAD–HEAD potential was used as the initial guess. Using these initial interaction potentials, RDFs were calculated from the production runs and the initial potentials subsequently updated via a Boltzmann inversion

$$V_{j+1}(r) = V_j(r) + kT \ln \frac{g_j(r)}{g^*(r)} \quad (2)$$

where $V_j(r)$ is the potential, $g_j(r)$ is the coarse-grained RDF, and $g^*(r)$ the target RDF, at distance r and iteration number j . A negligible change in the potential determines when convergence occurs. Given that the potentials for pentanol are fitted, only the RDF between the ALK beads is shown in Figure 3 as a representative result to illustrate the agreement achieved

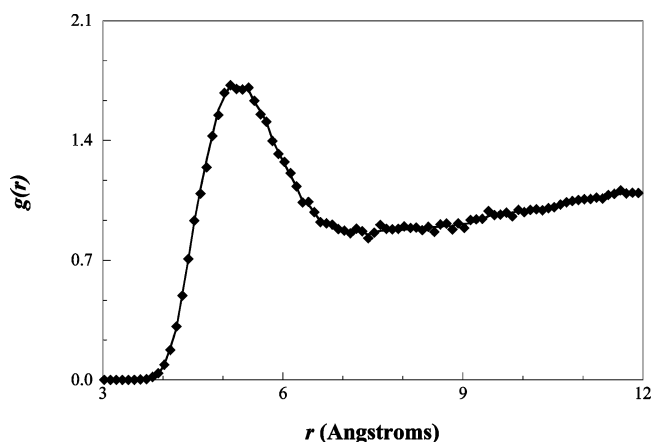


Figure 3. Radial distribution function between two hydrophobic beads (ALK–ALK) from a coarse-grained simulation (diamonds) and the target atomistic simulation (solid line) of pure 1-pentanol.

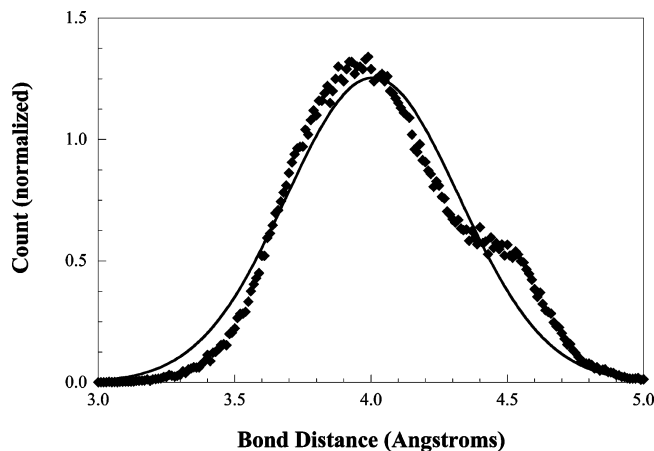


Figure 4. Probability distribution of a PALC-ALK bond length from an atomistic trajectory (diamonds) mapped to the coarse-grained level and fitted by a Gaussian curve (solid line).

between the target RDF and the RDF obtained from the coarse-grained model.

Once the pure model potentials are optimized by fitting to the RDFs extracted from the target atomistic trajectory, the cross-interactions in the mixed solute/solvent systems must be determined. In all of the mixtures, the potentials from the pure simulations (i.e., H2OX-H2OX for water and ALK-ALK, PALC-PALC, and ALK-PALC for pentanol) were used in a transferable fashion in the mixed systems, while the cross-interactions (H2OX-ALK and H2OX-PALC) were optimized to fit the target RDFs. In addition to the water models already mentioned, the cross-interactions between the coarse-grained alcohol and fully atomistic TIP3P water were also optimized to assess the difference between a high and low detail model with respect to solvating a coarse-grained solute.

Finally, the force field for the solute molecules requires intramolecular potential parameters, which are also derived from the atomistic target trajectory. For the bond lengths and bond angles, the distribution of distances between two bonded sites, or angles between three angled sites, is determined and fitted to a single-peak Gaussian. Before normalization, following the original RPM method,²⁰ the angle distributions are weighted by a factor of $\sin(\theta)$, as given by

$$P(\theta) = f_n p(\theta) / \sin(\theta) \quad (3)$$

where f_n is the normalization factor, $p(\theta)$ the distribution, $P(\theta)$ the normalized distribution, and θ the angle. A Boltzmann inversion is then applied to the fitted Gaussian distribution, and the parameters of a harmonic oscillator emerge from the simplification of the inversion, as seen in eqs 4–6.²⁰

$$P(\theta) = \frac{A}{w\sqrt{\pi/2}} \exp^{-2(\theta - \theta_{eq})^2/w^2} \quad (4)$$

$$V(\theta) = -kT \ln \left(\frac{A}{w\sqrt{\pi/2}} \exp^{-2(\theta - \theta_{eq})^2/w^2} \right) \quad (5)$$

$$V(\theta) = \frac{2kT}{w^2} (\theta - \theta_{eq})^2 + \text{const} \quad (6)$$

where A is the Gaussian area, w the Gaussian width, θ_{eq} the equilibrium angle, k Boltzmann's constant, and T the temper-

ature. For hexadecanoic acid,²⁵ the bonded potential was developed as described above in previous work.²⁵ For 1-pentanol, the distribution of bond lengths obtained from the atomistic simulation trajectory mapped to the coarse-grained level, along with its Gaussian fit, for the single coarse-grained bond is shown in Figure 4. The molecule is flexible on the atomistic level, and as a result produces a very wide bond-length distribution; therefore, a fairly low force constant of 35.6 kJ/Å²/mol on the coarse-grained level is measured.

To determine how many water molecules should be mapped per bead on the coarse-grained level, several models, all with different numbers of atomistic water molecules within a coarse-grained bead, were optimized and their individual accuracy assessed through a comparison of the results for three key properties, namely, computational efficiency, density, and solvation. Computational efficiency is obviously an important factor and was measured by the reduction in simulation time for the pure coarse-grained water models compared to the atomistic simulation. The computational speed-up is represented as the ratio of the coarse-grained (CG) simulation time to the simulation time for pure atomistic water, as shown in eq 7 and reported by DL_POLY

$$\text{speed scale} = \frac{\text{CG time}}{\text{atomistic time}} \quad (7)$$

To judge the accuracy of the model with respect to reproducing the bulk phase density of water, the percentage difference between the atomistic TIP3P water density and that from the coarse-grained water models was determined from constant pressure simulations. Finally, the solvation properties of the water model were verified by comparing the difference between the RDF for two ALK beads (the key solute interaction) from the coarse-grained mixture simulation to the target RDF via a merit function

$$f = \int w(r)(g(r) - g^*(r))^2 dr \quad (8)$$

where $w(r)$ is a weighting function described by

$$w(r) = g^{*-2}(r_{\max}) \exp \left(\frac{-|r - r_{\max}|}{r_{\max}} \right) \quad (9)$$

and r_{\max} is the distance where the highest peak is located. The quantity within the exponential term provides the greatest weight to values near the highest peak and is normalized by r_{\max} . The $g^{*-2}(r_{\max})$ term was added to normalize eq 8 so the values of the merit function could be compared between different systems.

These three criteria for assessing the most appropriate level of coarse-graining water were equally weighted and combined into a scoring function described by

$$S_{\text{fxn}} = \frac{\text{speed scale}}{\% \text{ diff}(\rho_{\text{pure}}) \times f} \quad (10)$$

where $\% \text{ diff}(\rho_{\text{pure}})$ is the percentage difference between the atomistic TIP3P water density and that from the coarse-grained water model. The model with the highest value of S_{fxn} was then utilized in a mixture of hexadecanoic acid and water, designed to mimic a simple biological lipid system.

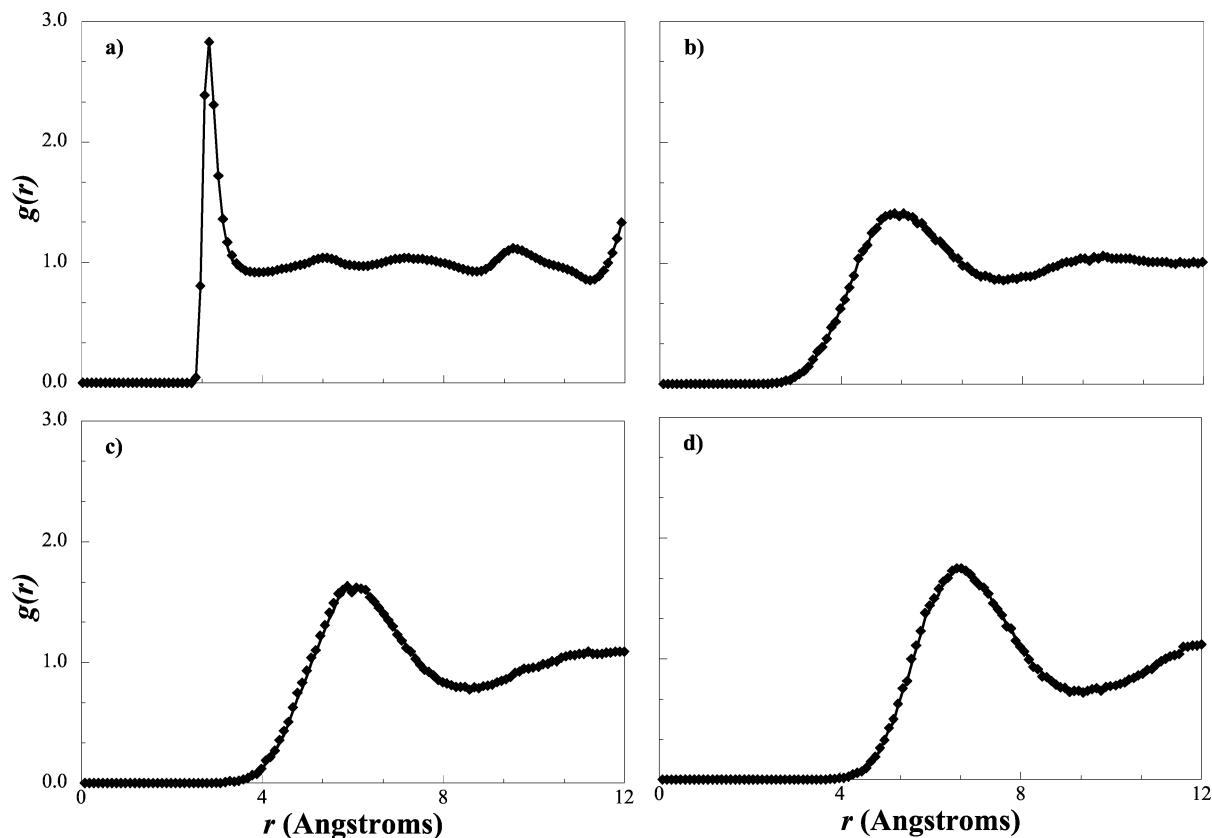


Figure 5. Radial distribution function between (a) one-water beads (H2O1–H2O1), (b) four-water beads (H2O4–H2O4), (c) six-water beads (H2O6–H2O6), and (d) eight-water beads (H2O8–H2O8), from a coarse-grained simulation (diamonds) and from the atomistic target simulation (solid line).

5. Results and Discussion

We first discuss the results from the development of the nonbonded potential for the different pure water models studied followed by the coarse-grained model development for the solute molecules. We then consider the results for the solute–solvent mixtures studied.

5.1. Pure Water. Using the RPM method, the target RDF for each water model was fitted to within line thickness. As an example, the water–water RDF for the H2O1, H2O4, H2O6, and H2O8 models is presented in Figure 5. Similar RDFs for the remaining water models are not shown, as they exhibit the same general shape and similar degrees of fit with respect to the target RDFs, as shown in Figure 5. From Figure 5a, we note the H2O1 model produces an RDF with a first peak that is much narrower and higher than the other models, indicating a much smaller bead and higher degree of structure. The high degree of structure most likely comes from the retention of hydrogen bonds between the water sites. Although without explicit hydrogens the coarse-grained beads do not hydrogen bond, the atomistic hydrogen bonds do directly affect the target RDFs, so those interactions are implicitly found within the coarse-grained nonbonded potential. In the H2O1 model, this implicit interaction has a larger contribution to the nonbonded potential than that found between the multiple waters per bead models. In the multiple water models, the hydrogen bonding is within the boundaries of the water bead and so the hydrogen bonding between beads is of less significance compared to the sum of the nonbonded interaction between beads, i.e., on the atomistic level, two water molecules mapped to the coarse-grained level and assigned to two different four-water beads may be hydrogen-bonded, but the potential between these two

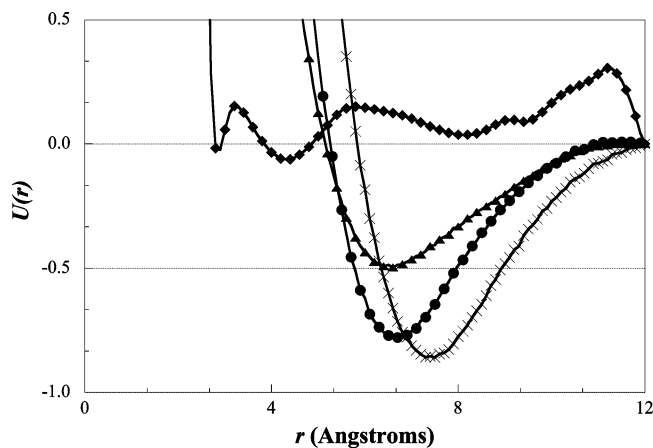


Figure 6. Interaction potential between one-water beads (H2O1–H2O1) (diamonds), four-water beads (H2O4–H2O4) (triangles), six-water beads (H2O6–H2O6) (circles), and eight-water beads (H2O8–H2O8) (crosses) fitted from the pure water system.

molecules is small compared to the potential shared between the beads that represent four water molecules. Wang et al.³² observed similar behavior when they applied the RPM method to the TIP3P, SPC, and SCP/E water models to develop coarse-grained models with a single water molecule per coarse-grained bead. They found that a one-water bead with an isotropic potential was not as accurate as models where orientation is incorporated into the coarse-grained model representation, such as the PM¹² or Mercedes-Benz water models,^{13,14} and concluded that this was due to the isotropic nature of the interactions in one-water models that do not allow for the orientation-specific hydrogen bonding observed atomistically.³²

TABLE 1: Comparison of the Results Obtained from the Different Water Models Studied

waters/bead	speed scale	density of pure water	% difference in density	RDF merit function	S_{fix}
atomistic	1.00	0.9983	0.0%	1.88×10^{-3}	n/a
1	16.7	0.9741	-2.4%	1.86×10^{-2}	3.76×10^2
3	152	0.9343	-6.4%	3.38×10^{-3}	7.04×10^3
4	254	0.9996	0.1%	7.75×10^{-3}	3.28×10^5
5	401	0.9815	-1.7%	2.35×10^{-2}	1.01×10^4
6	562	1.0942	9.6%	5.00×10^{-2}	1.17×10^3
8	979	1.0190	2.1%	5.18×10^{-2}	9.00×10^3
9	1234	0.9412	5.7%	5.35×10^{-2}	4.05×10^3

The difference in the models can also be seen in the fitted potentials, as shown in Figure 6. With the multiwater models, the potentials are found to be very similar to a Lennard-Jones potential in form, with the depth of the well increasing with the degree of coarse-graining. The location of the energy minimum also shifts to larger r , indicating that the size of the bead is increasing as the degree of coarse-graining increases. However, from the figure, we note that the H2O1 potential is very different in shape to that seen for the multiwater potentials but follows the same trend in terms of the depth and position of the potential well with respect to the degree of coarse-graining. The multiple wells and erratic shape are most likely a product of the high degree of structuring and hydrogen-bonding features found in the target RDF, as discussed above.

The increase in speed for the coarse-grained water models compared to the atomistic model is reported in Table 1, from which we see, for example, that H2O9 provides a factor of 1234 speed-up. However, the speed scale plateaus with increasing coarse-graining, indicating that the gain in calculation efficiency is greater between H2O4 and H2O3 (a ratio of 1.67) than between H2O9 and H2O8 (ratio of 1.26). Also reported in Table 1 is the accuracy of each model in reproducing the density of pure water. H2O4 is found to provide the most accurate water density with a difference of 0.1% compared to the atomistic value, which is reasonable if one takes into account the fact that the coordination number of water is 4.35.³³

5.2. Water/Pentanol. We now consider the coarse-grained pentanol-solvent simulations to further test and evaluate the different coarse-grained water models. As described above, the ALK-ALK and PALC-PALC potentials used in the water-pentanol simulations were taken from the simulation of pure pentanol and used transferably, while the cross-interactions were fitted. In Figure 7, the RDF between the ALK beads of pentanol is compared from simulations using each of the different water models to the atomistic target. From the figures, we see that the agreement between the RDF and its target generally deteriorates with an increase in the coarse-graining of the water; i.e., the highest detail model (atomistic TIP3P) provides the best solvation for the coarse-grained solute, but the H2O9 model does not properly solvate the 1-pentanol. The deterioration can also be seen numerically in Table 1 where we report the RDF merit functions. Although not shown, the same trend is seen in the corresponding RDFs for the PALC-PALC and ALK-PALC interactions. An exception is seen for the pentanol-H2O1 system, shown in Figure 7a, as more coarse-grained models provide better agreement. We believe this also reflects the differences in hydrogen bonding for the H2O1 and multiwater models, as discussed previously.

In Figure 8, we compare the water-water RDF from simulations of the water-pentanol mixture with the single-water coarse-grained model (H2O1) and a representative multiwater coarse-grained model (H2O4) against the atomistic target data. From the figures, we can see that the level of interaction between H2O4 beads (Figure 8b) is in good agreement with the target

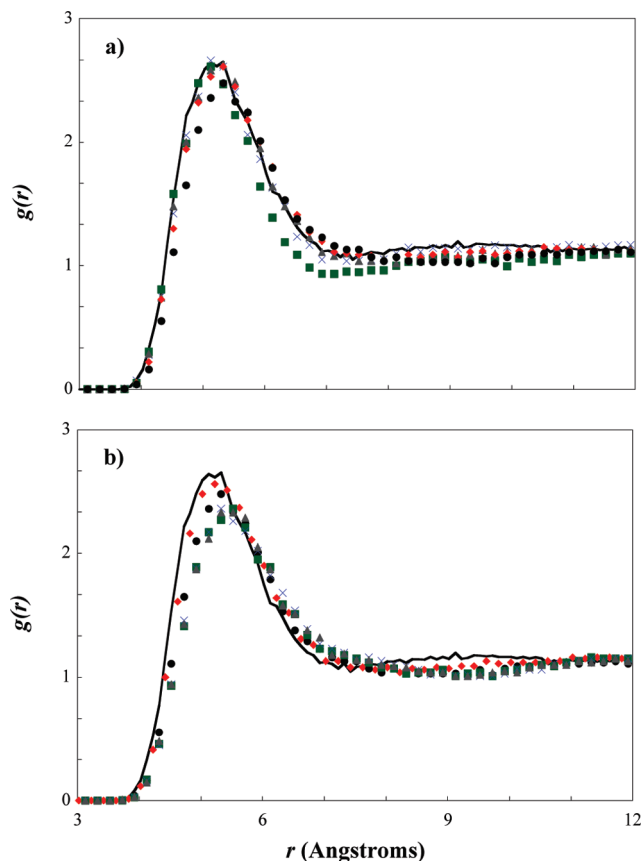


Figure 7. Radial distribution function between two hydrophobic beads (ALK-ALK) from simulations of 1-pentanol with (a) atomistic (crosses), H2O1 (squares), H2O3 (triangles), H2O4 (diamonds), and H2O5 (circles) water models, compared to the target atomistic simulation (solid line). (b) Radial distribution function between two hydrophobic beads (ALK-ALK) from simulations of 1-pentanol with H2O4 (diamonds), H2O5 (circles), H2O6 (squares), H2O8 (crosses), and H2O9 (triangles) water models, compared to the target atomistic simulation (solid line).

data, as indicated by the agreement in the RDFs, while the pure H2O1 potential (Figure 8a) produces a first peak that is much lower than that for the target RDF. Although the RDF in the pure water simulations of H2O1 is much taller than the other water models because of its hydrogen bonding (Figure 5a), it is not structured enough to produce an appropriately strong interaction in the mixed system. The multiwater model on the other hand has the necessary water-water interaction strength, since it does not require the pure water potential to account for hydrogen bonding; i.e., the hydrogen bonding is predominately found within the coarse-grained bead rather than between beads. The relative strength of attraction between water beads can also be seen in Figure 6, where the well is found to be much deeper for the H2O4 model compared to H2O1. As a result, an isotropic interaction accurately accounts for the water-water structure in the multiple-water coarse-grained model.

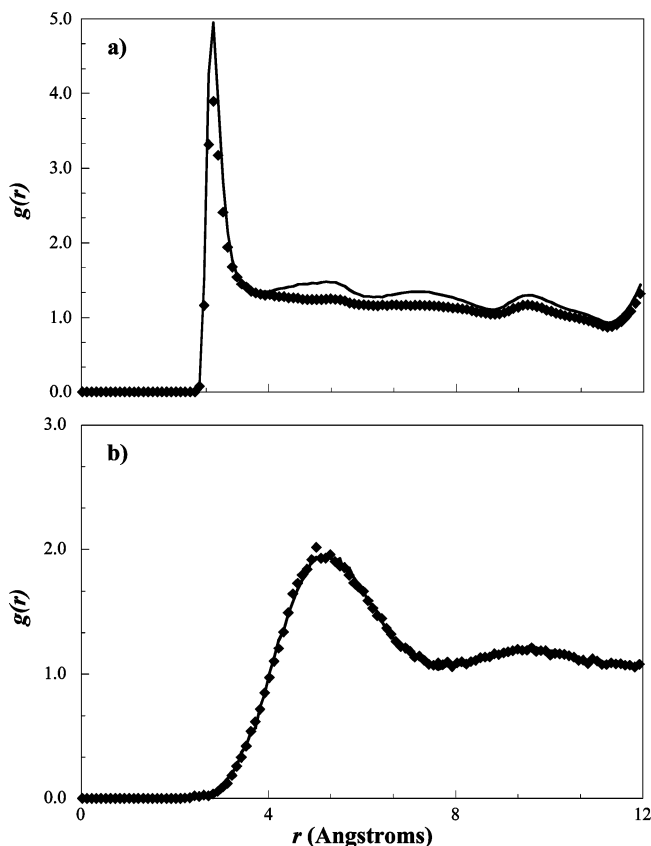


Figure 8. Radial distribution function between (a) one-water beads (H2O1–H2O1) and (b) four-water beads (H2O4–H2O4) from a coarse-grained simulation (diamonds) and an atomistic simulation (solid line) of the water–1-pentanol mixture.

We note that Wang et al. also observed a lack of structure in one-water coarse-grained models developed for the atomistic TIP3P, SPC, and SPC/E water models.³² This was attributed to the fact that RDFs fitted via the RPM method did not reproduce water's tetrahedral packing; if the potential was modified to reproduce the tetrahedral packing, the resulting RDF exhibited a much higher degree of structure. On the basis of this work, better agreement for the water RDF in the pentanol–H2O1 mixture could likely be achieved by reoptimizing the H2O1 model to reproduce the tetrahedral packing in pure water; however, a discrepancy in the pure RDF would result. Wang et al.'s findings also support our conclusion that the sharpness of the H2O1 RDF peak in Figure 5a can be attributed to hydrogen bonding. By fitting the tetrahedral packing, the existence of the hydrogen bonds is reinforced, and as a result, the coarse-grained peak becomes taller and thinner than the target.

From the results for the pure and mixed systems studied and reported in Table 1, we can now determine the scoring function, S_{fn} , for each water model as defined by the computational efficiency (speed scale), the percentage difference in the density for the coarse-grained models compared to the atomistic model, and the RDF merit function. From the table, we note that the H2O4 model has the highest scoring function, indicating it has the optimal trade-off between speed and accuracy, even though H2O3 is the most accurate with respect to the transferred RDF for the pure hydrophobic bead interaction.

5.3. Water/Hexadecanoic Acid. With the highest scoring function, the four-waters-per-bead model (H2O4) was chosen as the coarse-grained water model for use in the water–hexadecanoic acid system. In this preliminary study of a solvated lipid system,

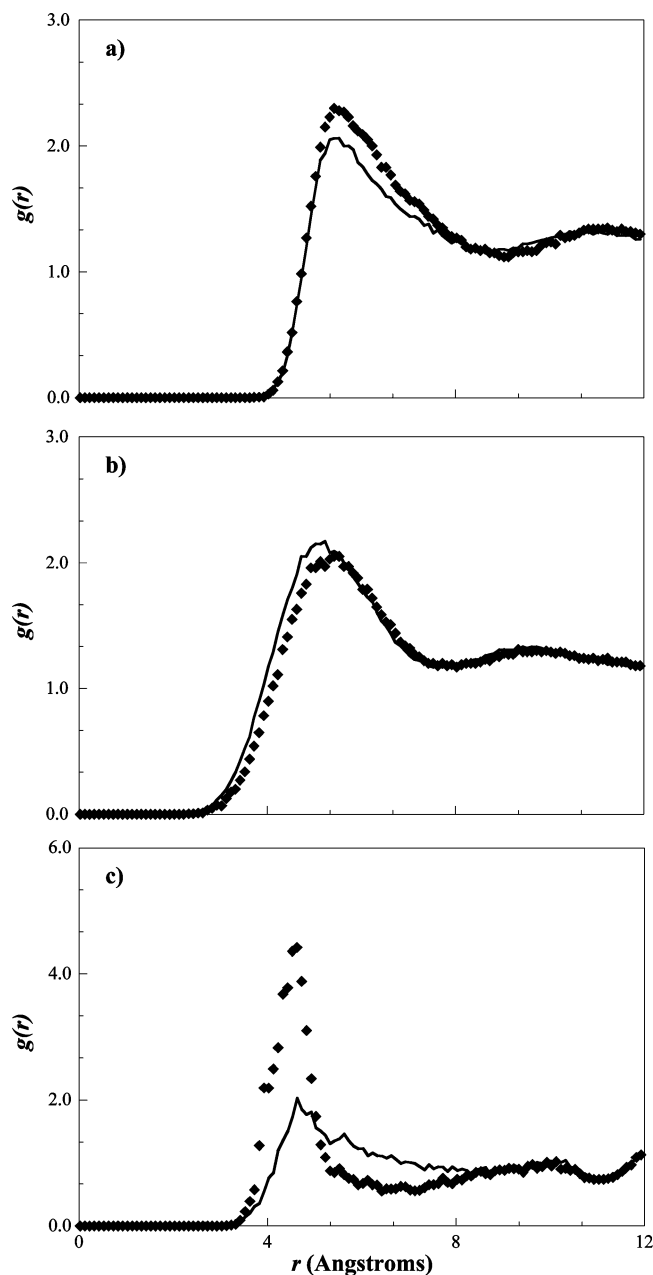


Figure 9. Radial distribution function between (a) tail beads (TAIL–TAIL), (b) four-water beads (H2O4–H2O4), and (c) head beads (HEAD–HEAD), for the hexadecanoic acid–water mixture from a coarse-grained simulation (diamonds) and from the target atomistic simulation (solid line).

the cross-interactions between water and hexadecanoic acid (H2O4–HEAD, H2O4–TAIL, and H2O4–TRM2) were optimized at 298 K and 1.0 bar. The RPM method was able to optimize the potentials such that the coarse-grained RDF and its target value agree within line thickness; the results are therefore not shown, and we focus instead on the transferred RDFs.

In Figure 9a, we compare the transferred pure TAIL–TAIL interaction to its target. The coarse-grained and target RDF is in good agreement in terms of the location of the peak, indicating the hydrocarbon tails are structuring themselves on the coarse-grained level in the same way as on the atomistic level; however, the height of the coarse-grained RDF is somewhat higher than its target, indicating a higher degree of clustering of the tail beads in the coarse-grained model compared to the target. In a

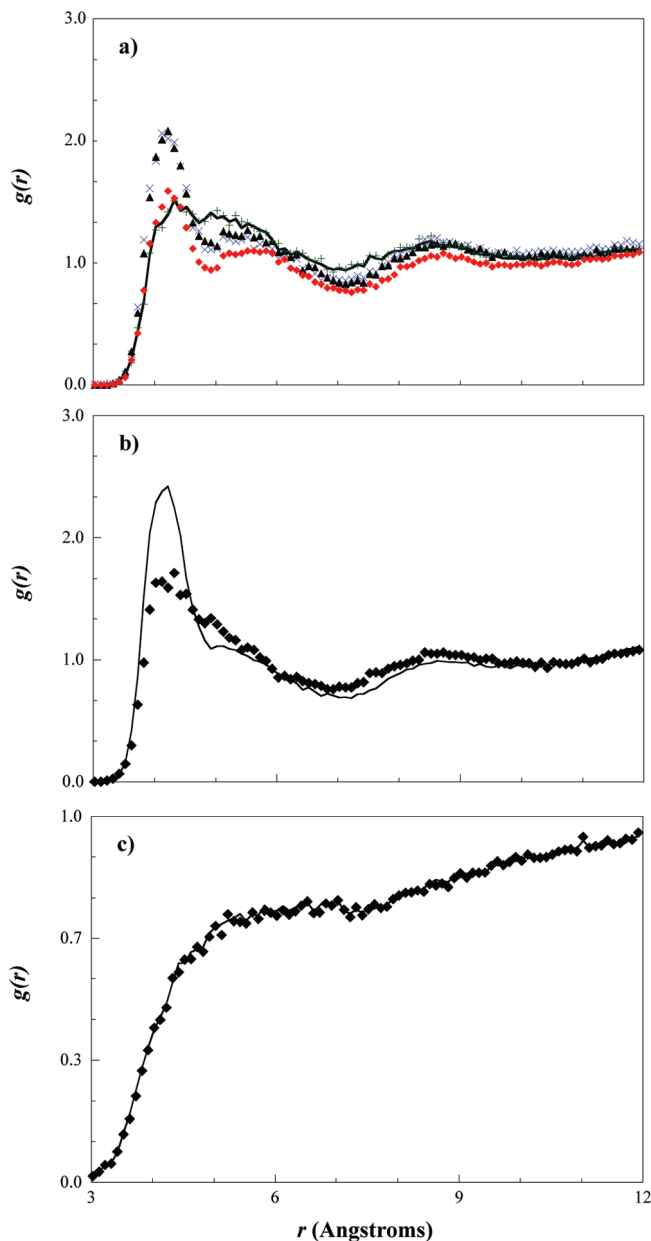


Figure 10. Radial distribution function between (a) alcohol beads (PALC–PALC) from a coarse-grained simulation of pentanol with atomistic water (crosses), H2O4 using the pure potential transferred to the water–pentanol mixture (triangles), H2O4 using the potential fitted to the pentanol mixture (plusses), and H2O4 using the ALK–H2O4 interaction replaced by the attractive PALC–ALK potential (diamonds), compared to the target atomistic simulation (solid line). (b) Radial distribution function between alcohol beads (PALC–PALC) in a simulation of pure pentanol using the potential fitted from a mixture simulation (diamonds) and the atomistic target (solid line). (c) Fitted radial distribution function between the hydrophobic bead of pentanol and a four-water bead (ALK–H2O4) from a coarse-grained simulation of the water–pentanol mixture (diamonds) and the target atomistic simulation (solid line).

bilayer system, this behavior could lead to a stronger tendency to phase separate from water and induce structure in the hydrophobic region of the bilayer. In Figure 9b, we see a similar trend for the H2O4–H2O4 RDF using the water interaction transferred from the pure simulation, indicating that the structure of water in the bulk phase is retained on the coarse-grained level. Finally, the RDF for the transferred HEAD–HEAD interaction

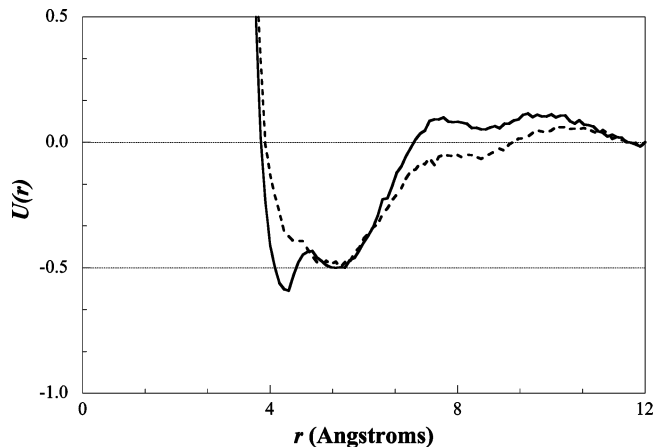


Figure 11. Interaction potential between alcohol beads (PALC–PALC) fitted from a pure simulation (solid line) and fitted from the mixture simulation (dashed line).

is shown in Figure 9c, where the coarse-grained RDF is found to possess a distinctive peak not seen in the target.

To investigate the possible reasons for the observed discrepancy, we consider the simpler (and computationally cheaper) alcohol system. As shown in Figure 10a, similar behavior is observed, in that the PALC–PALC interaction in the H2O4–pentanol mixture is also higher than the atomistic target. Similar behavior is seen for the other coarse-grained water models studied and in simulations with the TIP3P model (also shown in Figure 10a), indicating that the discrepancy is independent of the water model used and most likely dependent on the solute potential. To investigate the effect of the PALC potentials, all of the interactions (pure and mixed) for the coarse-grained H2O4–pentanol system were reoptimized against the target RDFs measured from the atomistic simulation. The RDF from the reoptimized PALC–PALC interaction is also presented in Figure 10a and, as expected, can be fitted within line thickness using the RPM method. If these reoptimized interactions (PALC–PALC, PALC–ALK, and ALK–ALK) are then transferred and used in a simulation of pure pentanol, the coarse-grained PALC–PALC peak is found to be much lower than the target value, as shown in Figure 10b. As such, the interaction fitted from the mixture does not exhibit the same interaction strength, or, most likely, the same hydrogen-bonding network as the potential fitted in the pure system. The depleted level of hydrogen bonding is supported when considering the trends for the pure water potential, where more explicit hydrogen bonding resulted in taller RDFs. The differences in the PALC–PALC potentials fitted from the pure RDFs and the mixture RDFs (seen in Figure 11) also support the reduced structuring with the mixture fitted potential. While the potentials are very similar, the potential from the pure simulation possesses an additional well, similar to that found for the pure H2O1 potential shown in Figure 6, suggesting that this is the hydrogen-bonding component of the coarse-grained potential. In addition, this well is deeper than the large well at larger r and steeper on both sides of the minimum, indicating the beads most likely get caught in the first minimum to produce the peak seen in Figure 10a.

Izvekov and Voth³⁴ observed similar results in their study of cholesterol within a dipalmitoylphosphatidylcholine (DPPC) bilayer using coarse-grained force fields derived through the force-matching procedure, in that they found that the cholesterol headgroup, while having the same forces on both the atomistic

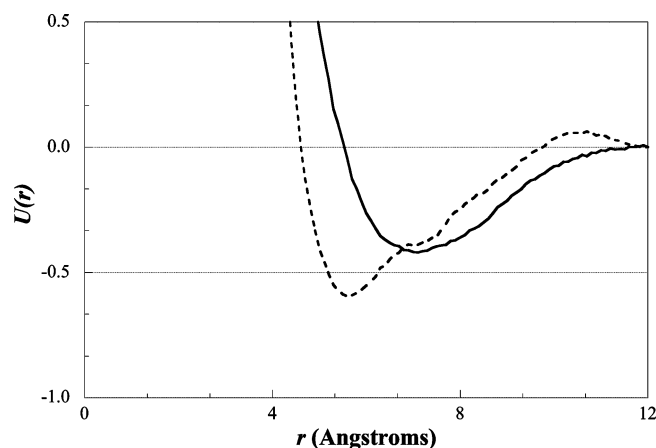


Figure 12. Interaction potentials between hydrophobic (ALK) and water (H2O4) beads (solid line) and hydrophilic (PALC) and hydrophobic (ALK) beads (dashed line).

and coarse-grained level, produced a strong peak in the coarse-grained RDF while only a weak peak was found on the atomistic level. As a result, the interaction between hydrophobic sites and water had a repulsive component at larger separation distances compared to the interaction between the cholesterol head beads and the interaction between the cholesterol head bead and water. The authors concluded that the discrepancy in the RDFs was due to the fact that the forces are being derived from a system where the hydrophobic beads maintain a larger average separation distance from the water beads than the hydrophilic beads and so have minimal direct contact with water. On the basis of their work, we hypothesize that the observed disagreement in the HEAD–HEAD RDF is independent of the coarse-graining methodology and also due to unavoidable sampling issues. This is further supported by the fact that the RDF between the ALK and H2O4 bead, as shown in Figure 10c, indicates that the fitted potential will possess a larger repulsive region. This can also be seen in Figure 12, where the fitted potential for ALK–H2O4 is compared to the more attractive PALC–ALK potential; although there is an attractive component in the ALK–H2O4 potential, the range of attractive separation distances is much smaller than that for PALC–ALK interaction.

In summary, if the PALC–PALC interaction is fitted to the atomistic mixture RDFs, the interaction strength is too low to produce the RDF peak seen in the pure pentanol system. In addition, if the ALK–H2O4 interaction is replaced by a potential with a larger attractive interaction region (i.e., PALC–ALK, as shown in Figure 12), the RDF of the PALC–PALC interaction transferred from the pure system to the mixture is in better agreement with the target RDF (Figure 10a). This indicates that the hydrophilic headgroup exhibits two priorities: to hydrogen bond with itself and to provide hydrophobic shielding between the water and the carbons in the hydrocarbon tail. The hydrophobic sites maintain minimal contact with water, because of the large repulsive interaction region in the ALK–H2O4 interaction. As a result, the PALC beads do not need to shield ALK beads from water and instead associate with each other at a separation distance corresponding to the first well in the potential shown in Figure 11, resulting in a high RDF peak. Atomistically, there is a level of attraction between the atoms of the hydrophilic bead and the water, so the atoms of the hydrophilic bead must sacrifice optimal hydrogen bonding to shield the water from the rest of the hydrophobic solute.

6. Conclusions

The K-means algorithm was successfully used to develop a coarse-grained model for water with multiple water molecules mapped to a single bead by enabling dynamic assignment of the water molecules to coarse-grained beads from analysis of atomistic trajectory data. Coarse-grained water models representing one to nine waters per bead were optimized and studied in both the pure state and in a mixture with 1-pentanol, a representative amphiphilic solute. On the basis of the increase in computational efficiency, the ability to reproduce the density of pure water, and to solvate the 1-pentanol solute, as measured by the RDF for the key ALK–ALK solute potential relative to the target RDF, the four-water model (H2O4) was chosen as the optimal water model.

The H2O4 model was then used in simulations of a water–hexadecanoic acid binary mixture. The solvation properties of the four-water model provided for the correct structuring of the hydrophobic component of hexadecanoic acid, as shown by the RDF between the TAIL beads derived from the mixture simulation; however, the RDF between HEAD beads appeared to be more attractive and structured than the atomistic target. A similar behavior was also observed for the hydrophilic interaction of the 1-pentanol/water mixture. By altering the potentials in the alcohol/water mixture, it was found that the discrepancy was independent of the water model, and the hydrophobic shielding and hydrogen bonding behavior of the atomistic alcohol are mutually exclusive properties on the coarse-grained level. In developing the coarse-grained model, the hydrophobic shielding is essentially sacrificed in order to retain the hydrogen bonding, which will force the hydrophilic RDFs to have higher structure than is seen atomistically, but will ensure the clustering of the HEAD beads. As a result, the hydrophilic groups will tend to associate and promote self-assembly; therefore, this property of the model must be retained.

Acknowledgment. The authors gratefully acknowledge support from the National Institute of Arthritis and Musculoskeletal and Skin Diseases through Grant No. R21 AR053270-02.

References and Notes

- (1) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219. Holt, J. K. *Microfluid. Nanofluid.* **2008**, *5*, 425.
- (2) Malenkov, G. *J. Phys.: Condens. Matter* **2009**, *21*, 35.
- (3) Guillot, B.; Guissani, Y. *J. Chem. Phys.* **2001**, *114*, 6720.
- (4) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragonés, J. L. *Faraday Discuss.* **2009**, *141*, 251.
- (5) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954.
- (6) Liew, C. C.; Inomata, H.; Arai, K. *Fluid Phase Equilib.* **1998**, *144*, 287.
- (7) Paricaud, P.; Predota, M.; Chialvo, A. A.; Cummings, P. T. *J. Chem. Phys.* **2005**, *122*.
- (8) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926.
- (9) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.
- (10) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616.
- (11) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem. B* **1987**, *91*, 6269.
- (12) Vlcek, L.; Nezbeda, I. *Mol. Phys.* **2004**, *102*, 485.
- (13) Silverstein, K. A. T.; Haymet, A. D. J.; Dill, K. A. *J. Am. Chem. Soc.* **1998**, *120*, 3166.
- (14) Bizjak, A.; Urbi, T.; Vlachy, V.; Dill, K. A. *Acta Chim. Slov.* **2007**, *54*, 532.
- (15) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
- (16) Ayton, G. S.; Tepper, H. L.; Mirijanian, D. T.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 4074. Izvekov, S.; Swanson, J. M. J.; Voth, G. A. *J. Phys. Chem. B* **2008**, *112*, 4711. Molinero, V.; Moore, E. B. *J. Phys. Chem. B* **2009**, *113*, 4008.
- (17) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.

- (18) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.
- (19) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.
- (20) Milano, G.; Goudeau, S.; Muller-Plathe, F. *J. Polym. Sci., Part B: Polym. Phys.* **2005**, *43*, 871.
- (21) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.
- (22) Shinoda, W.; DeVane, R.; Klein, M. L. *Soft Matter* **2008**, *4*, 2454.
- (23) MacQueen, J. B. *Some Methods for Classification and Analysis of Multivariate Observations*; 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1967.
- (24) Steinley, D. *Br. J. Math. Stat. Psychol.* **2006**, *59*, 1.
- (25) Hadley, K. R.; McCabe, C. *J. Chem. Phys.*, in press, 2010.
- (26) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136.
- (27) Schlenker, M.; Brickmann, J.; MacKerell, A. D., Jr.; Karplus, M. Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications. In *Biological Membranes: A Molecular Perspective from Computation and Experiment*; Roux, K. M. a. B., Ed.; Birkhauser: Boston, MA, 1996. Feller, S. E.; MacKerell, A. D. *J. Phys. Chem. B* **2000**, *104*, 7510. Feller, S. E.; Gawrisch, K.; MacKerell, A. D. *J. Am. Chem. Soc.* **2002**, *124*, 318.
- (28) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (29) Neumann, M. *J. Chem. Phys.* **1985**, *82*, 5663.
- (30) Inglese, A.; Robert, P.; DeLisi, R.; Milioto, S. *J. Chem. Thermodyn.* **1996**, *28*, 873. Lide, D. R. *CRC Handbook of Chemistry and Physics*, 87th ed.; Taylor & Francis: Boca Raton, FL, 2006.
- (31) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. *J. Chem. Phys.* **2007**, *126*. Chan, E. R.; Striolo, A.; McCabe, C.; Cummings, P. T.; Glotzer, S. C. *J. Chem. Phys.* **2007**, *127*.
- (32) Wang, H.; Junghans, C.; Kremer, K. *Eur. Phys. J. E* **2009**, *28*, 221.
- (33) Soper, A. K. *J. Chem. Phys.* **2000**, *258*, 121.
- (34) Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637.

JP911894A