Coarse-Grained Potential Models for Phenyl-Based Molecules: I. Parametrization Using Experimental Data

Russell DeVane,*,† Michael L. Klein,† Chi-cheng Chiu,‡ Steven O. Nielsen,‡ Wataru Shinoda,§ and Preston B. Moore⊥

Institute for Computational Molecular Science and Department of Chemistry, Temple University, 1901 North 13th Street, Philadelphia, Pennsylvania 19122, Department of Chemistry, The University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080, Research Institute for Computational Sciences, National Institute of Advanced Industrial Science and Technology (AIST), Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, and Department of Chemistry & Biochemistry, University of the Sciences in Philadelphia, 600 South 43rd Street, Philadelphia, Pennsylvania 19104

Received: December 11, 2009; Revised Manuscript Received: March 9, 2010

A coarse-grained intermolecular potential has been parametrized for phenyl-based molecules. The parametrization was accomplished by fitting to experimental thermodynamic data. Specifically, the intermolecular potentials, which were based on Lennard-Jones functional forms, were parametrized and validated using experimental surface tension, density, and partitioning data. This approach has been used herein to develop parameters for coarse-grained interaction sites that are applicable to a variety of phenyl-based molecules, including analogues of the amino acid side chains of phenylalanine and tyrosine. Comparison of the resulting coarse-grain model to atomistic simulations shows a high level of structural and thermodynamic agreement between the two models, despite the fact that no atomistic simulation data was used in the parametrization of the coarse-grain intermolecular potentials.

I. Introduction

There is continuing interest in developing a coarse-grained (CG) potential for computational studies of soft matter and biological systems. CG models provide a means to expand the utility of existing computational resources by allowing the exploration of far greater temporal and spatial scales than is possible with traditional MD simulations utilizing full atomic (AA) detail. However, the added efficiency comes necessarily at a cost, namely, loss of resolution in the description of the system. Typically, several atoms comprise a single CG interaction site, thus reducing the number of interacting particles necessary to characterize the system. With this approach in mind, many groups have developed CG models using various approaches with numerous applications appearing in the literature over the past four decades.[1−15]

Herein, we present the application of a recently proposed methodology to develop CG MD models for phenyl-based molecules. This method is unique in that it removes the common dependence on AA MD simulations by making extensive use of experimental thermodynamic data. This approach builds on previous work by Nielsen et al.,[16] which was extended and applied to the development of CG parameters for PEG surfactants and amino acids.[10−12,17] Phenyl-based molecules are of particular importance from a biological standpoint as they are constituents of proteins; for example, phenylalanine (PHE), tyrosine (TYR), and tryptophan (TRP) and are building blocks for other biologically important molecules. These molecules have been the focus of previous CG parametrization approaches.[9,18−24] The CG models and parametrization methodology presented herein are currently being employed for the development of CG models for a wide range of systems including lipids, amino acids, and charged surfactants. In our companion paper, we show how to adapt the present CG model to enable studies of fullerenes.[25] The methodology used for the parametrization and validation of the CG models will be presented in section II. The details of the parametrization will be presented in section III. In order to validate the model, comparison to AA MD and experimental data will be presented in section IV. Finally, the conclusions will be presented in section V.

II. Methods

A. Coarse Grain Model Potential. For generality and ease of implementation, the present CG model employs Lennard-Jones (LJ)-style nonbonded potential functions. This work and previous work has demonstrated the ability of these potentials to model systems at the CG level with sufficient accuracy to predict phase behavior and interfacial properties of surfactants.[10−12] For the models developed herein, the CG beads interact via a LJ 9−6 or LJ 12−4 potential given as follows

$$v(r)_{9-6} = \frac{27}{4}\epsilon\left(\frac{\sigma^9}{r^9} - \frac{\sigma^6}{r^6}\right) \qquad v(r)_{12-4} = \frac{3\sqrt{3}}{2}\epsilon\left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^4}{r^4}\right) \tag{1}$$

The choice of prefactors for the LJ functions are selected such that $v(\sigma) = 0$ and ϵ is the well depth. The choice of the LJ functional form (for example the LJ 9−6 in eq 1) is essentially an adjustable parameter used in the fitting procedure. We have previously explored various LJ styles including 6−4, 8−4, 10−4, 9−6, and 12−4.[10] For alkane interactions, the choice of

* To whom correspondence should be addressed.
† Temple University.
‡ The University of Texas at Dallas.
§ National Institute of Advanced Industrial Science and Technology (AIST).
⊥ University of the Sciences in Philadelphia.

10.1021/jp9117369
Published on Web 04/28/2010

the 9−6 functional form was validated by comparison of CG liquid structure to AA MD results.[10] The adoption of the 9−6 functional form here provides the ability to reproduce several properties with the CG model and maintains consistency with existing CG models.[10,11] Finally, the CG water (W) model used herein was developed previously in our group and employs a LJ 12−4 functional form for the nonbonded interactions between the water sites. Details can be found in our previous work.[10]

A simple truncation, implemented at a distance of rcut = 15 Å, is employed for the long-range cutoff, with no smoothing or shifting. This length is sufficient to avoid gross artifacts resulting from discontinuity; however the cutoff length, rcut, does affect the thermodynamic properties and so is effectively treated as a CG fitting parameter. The parameters in the LJ function are fixed by reproducing known thermodynamic data. For single component systems, both ϵ and σ can be unambiguously fixed by using a combination of density and surface tension as targets. The cross interactions arising from nonidentical CG sites can be generated using the Lorentz−Berthelot combining rules given by eq 2.

$$\sigma_{ab} = \frac{\sigma_{aa} + \sigma_{bb}}{2} \qquad \epsilon_{ab} = \sqrt{\epsilon_{aa}\epsilon_{bb}} \tag{2}$$

Here, ϵ_{aa} and σ_{aa} represent the self-interaction values for ϵ and σ, and ϵ_{ab} and σ_{ab} represent those for the cross interactions. The Lorenzt−Berthelot mixing rules are not used without testing to evaluate the ability of the parameters to reproduce the correct target values (in this case, density and surface tension). If needed, further adjustments are made to the parameters to give better agreement with target data. Further, the use of mixing rules to develop parameters between particles with different functional forms (9−6, 12−4) for the self-interactions would not, in general, be expected to yield good results. Nonetheless, this approach could be used to generate an initial set of parameters, which would be followed by testing and, if needed, further refinement.

The bond and angle interactions are modeled via harmonic potentials given by eqs 3 and 4

$$V(r)_{bond} = k_b(r - r_0)^2 \tag{3}$$

$$V(r)_{angle} = k_a(\theta - \theta_0)^2 \tag{4}$$

where r_0 and k_b represent the equilibrium bond length and force constant for bonds; θ_0 and k_a represent the equilibrium angle value and force constant for angles.

Dihedrals are modeled using eq 5

$$V(r)_{dihed} = k_d[1 + \cos(n\theta - d)] \tag{5}$$

which is based on the LAMMPS implementation of the CHARMM dihedral potential. In that implementation, K_d is the force constant, n (integer ≥ 0) determines the periodicity, and d (integer value of degrees) is the equilibrium dihedral angle input as an integer. The terminology used here matches that used in the LAMMPS documentation.(http://lammps.sandia.gov/). In this work, all dihedrals are modeled with the same parameters, k_d = 60 kcal/mol, n = 1, and d = 180° with the weighting factor (preventing double calculation of 1−4 interactions) set to 0.0 (see LAMMPS documentation).

B. Molecular Dynamics Simulations. For work performed here, AA MD simulations were carried out using the NAMD MD package,[26] and analysis was performed using the visualization tool VMD.[27] The CHARMM PARAM27 force field was used with the van der Waals interactions truncated at 12 Å and the standard CHARMM smoothing function from 10 to 12 Å.[28] The long-range Coulomb interactions were handled via the particle mesh Ewald method.[29] Constraints were applied to bonds involving hydrogen atoms using the SHAKE/RATTLE (ROLL) method.[30] CG MD simulations were performed using the LAMMPS code developed at Sandia National Laboratory and extended by our group to implement our CG models.[31] This extension is now part of the standard LAMMPS release available at http://lammps.sandia.gov. A time step of 10 fs was used for the CG simulations, without using multitime step integrators, although this could be extended to a larger value by taking advantage of such algorithms. As a test, bulk properties calculated for benzene using 4 and 20 fs did not significantly diverge from those calculated using 10 fs. Simulations involving water and hexane were performed using parameters previously developed for those solvents.[10] The CG model for water maps three water molecules per CG site. Details of the CG water model were previously reported, where the CG type was labeled "W" (this type label will also be used here to identify the CG water model), with parameters provided by Shinoda et al.[10] Details and parameters for the CG sites used to model hexane and heptane were also previously reported.[10] The hexane model was produced by combining two "CT" CG sites (CT−CT) with a bond; heptane was modeled using "CT2" and "CM" CG sites (CT2−CM−CT2) including bonds and an angle. The parameters are provided here for convenience.

C. Surface Tension and Density. The surface tension, γ, was calculated from CG MD simulations via

$$\gamma = \frac{L_z}{2}\left[P_{zz} - \frac{(P_{xx} + P_{yy})}{2}\right] \tag{6}$$

where L_z is the box dimension in the z-direction and P_{aa} is the isotropic pressure tensor for each Cartesian direction, where a represents x, y, or z.[29] For pure systems, equilibrated simulation boxes were used as starting configurations. The z-dimension of the box was extended to create an interface with the vacuum perpendicular to the z-axis. Simulations were carried out for roughly 10 ns. Density calculations were performed with NPT simulations averaged over roughly a 5 ns simulation time. Finally, all MD simulations were performed at a temperature of 303 K, unless otherwise specified.

It is important to point out that the CG potentials developed here are effective potentials of mean force. The potentials subsume some of the entropy that is removed as a result of the reduction in the number of degrees of freedom. The implication here is that the virial, and hence the surface tension formula, is not rigorous. As such, the potentials are nonadditive and valid only for a specific state point. This limitation is not confined to the potentials presented here but is a common artifact of the effective pairwise potential including many AA and most other CG models.[32−34]

D. Partitioning. The solubility of molecules in various solvents can be used as a measure of relative hydrophobicity. Indeed, much work has been done to relate hydrophobicity scales to the partitioning between various solvents and their vapor phases such as air−water, water−hexane, and so forth.[35−37] This data also makes an excellent target when developing a model or a good source of validation for an existing model,

particularly for molecules in heterogeneous environments. Although partitioning can be directly evaluated in MD simulations, it is often more efficient to use enhanced sampling techniques to calculate a free energy for transferring a solute molecule from one phase to another. Thermodynamics then allows the direct comparison of partition coefficients to transfer free energies. Many values for the partitioning (or free energy) of phenyl-based molecules from water to hexane (and other alkanes including cyclic hydrocarbons) exist in the literature. Sharp et al. performed a careful evaluation of experimental partitioning data and considered the often overlooked entropic contributions which enter into the equations as volume corrections for conversion of the partitioning data to free-energy values.[35−39]

E. Free-Energy Methods. For the CG models, the potential of mean force (PMF) calculations (solvation free energy and dimerization energy) were performed using steered MD simulations implemented in LAMMPS.[31] The forces calculated from the nonequilibrium steered MD trajectories are related to the equilibrium free-energy difference via Jarzynski's equality.[40] Solvation free-energy calculations were performed starting with a box of equilibrated CG water and then extending the *z*-dimension of the box, thus forming a vacuum region with the interface perpendicular to the *z*-axis. The target molecule (benzene, toluene, etc.) was inserted and pulled at a constant velocity in the *z*-direction through the box from water to vacuum.

Transfer free energies were calculated using a system containing a mixture of hexane and water phases with the interface perpendicular to the *z*-axis. The target molecule was pulled from one solvent to the other (hexane to water) at a constant velocity. For solvation and transfer free-energy calculations, the reaction coordinate was taken as the *z*-axis, which was perpendicular to the air−water and hexane−water interfaces, respectively. Our model, being designed to reproduce the experimental surface/interfacial tension, provides well-defined boundaries at air−water and hexane−water interfaces. The reaction coordinates were extended in both directions until asymptotic behavior was displayed in the PMF, indicating bulk phase solvation.

The dimerization free-energy calculations were performed on systems which contained two identical molecules in a box of CG water. The PMF was calculated by pulling the molecules together from a distance large enough to ensure monomeric solvation. The reaction coordinate was taken as the center-of-mass (COM) distance between the two molecules and extended from 3 to 12 Å separation. For all of these calculations, a pulling speed of 1 Å/ns and force constant of 50 kcal/mol/Å^2 was used. Since these calculations are relatively inexpensive in CG MD, highly converged PMFs were calculated using roughly 50 repetitions along the full reaction coordinate.

For the AA systems, dimerization free-energy calculations (PMF between two identical molecules) were calculated with NAMD using the CHARMM PARAM27 force field and the adaptive biasing force (ABF) method.[26,28,41,42] The reaction coordinate used here was the distance between the COM of the two molecules (as in the corresponding CG calculations). Samples were taken for COM distances of 3 to 12 Å with bins 0.2 Å in width. Simulation times of 30 ns were used, resulting in accumulation of at least 250 000 samples for each bin.

III. Parametrization

A. Coarse-Grain Mapping. In order to develop a CG model that can be used for MD simulation of functionalized benzene moieties such as toluene and *p*-cresol, the mapping level was carefully examined. It is well-understood that the act of coarse-graining necessarily requires the loss of some level of detail in the representation of the system of interest. Taken into consideration were the consequences of representing a highly planar structure such as benzene with a small number of spherical sites and how this would affect the effective size and shape of the CG model. For modeling phenyl-based amino acids (PHE, TYR, TRP), this could have serious implications for side-chain packing, as shown in our previous work.[12] Further, for TYR (*p*-cresol), the need to construct the para configuration was a factor. Considering these points, a four-site model was selected.

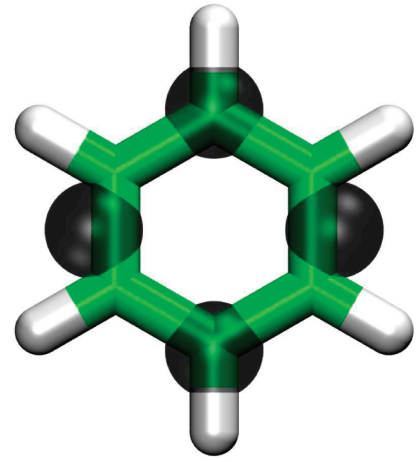

Figure 1. The CG mapping of benzene is shown here with the CG BER sites in black and the AA benzene model shown with green carbons and white hydrogens.

This level of CG mapping introduces a smaller reduction in the number of interaction sites than is typical with our CG models but is necessary to capture the highly planar geometry of the phenyl-based molecules and allow a more accurate representation of para configurations for functionalized phenyls. Nevertheless, significant computational savings are still gained with the use of this CG representation, especially if one considers that long-range electrostatic interactions have been removed and that typically most of the simulation box contains solvent, for example, water, which is significantly more computationally efficient in the CG approach.

B. Benzene. The starting point for parametrization of phenyl-based molecules is benzene. In the CG model presented here, benzene is represented by four equivalent sites, leading to a mapping ratio of 4 CG sites to 12 atoms in the AA representation. The CG type for benzene is identified as BER, and the AA to CG mapping is shown in Figure 1. The intramolecular potential was parametrized to reproduce the bond distances and angles taken from a geometry-optimized structure based on Hartree−Fock ab initio theory at the 6-311G* level. AA MD simulations were performed, and the trajectories were mapped to the CG representation. On the basis of this mapping, it was observed that CG benzene is essentially a rigid ring structure. To reproduce this characteristic, the force constants in the CG intramolecular force field were set to produce an essentially rigid ring structure. A similar method was used for parametrization of the intramolecular potentials for all molecules shown here. Parameters for eqs 3 and 4 are given in Tables 1 and 2, respectively. Finally, a dihedral angle was used to maintain a planar ring geometry. The LJ parameters for the BER−BER interaction were adjusted to reproduce the experimental surface tension and density of pure benzene at 303 K, and the values

TABLE 1: Bond Parameters r_0 and k_b for Equation 3[a]

types	k_b (kcal/mol/Å^2)	r_0 (Å)
BER−BER	40.0	2.00
BER−PHR	40.0	2.60
BER−XYR	40.0	2.60
CT−CT	6.955	3.71
CT2−CM	9.000	3.13

[a] Bonded parameters for CT−CT were previously developed and reported.[10]

TABLE 2: Angle Parameters θ_0 and k_a for Equation 4

types	k_a (kcal/mol/radian^2)	θ_0 (degrees)
BER−BER−BER	30.0	90.0
BER−XYR−BER	30.0	58.0
BER−PHR−BER	30.0	58.0
XYR−BER−BER	30.0	106.0
PHR−BER−BER	30.0	106.0
XYR−BER−XYR	30.0	122.0
XYR−BER−PHR	30.0	122.0
CT2−CM−CT2	1.70	173.0

are reported in Table 4. For the interaction with water (BER−W), the LJ parameters between BER and W were fit to reproduce the experimental interfacial tension between benzene and water. This is a viable option for liquids that are not miscible. The results for the interfacial surface tension are given in Table 4.

C. Toluene and *p*-Xylene. The toluene CG model is composed of three BER-type CG sites and an additional site, XYR, to account for the additional methyl group attached to the benzene ring. The left side of Figure 2 shows the AA representation of toluene with the three BER sites (black) and the XYR site (blue) superimposed over the AA structure of toluene. These parameters are given in Tables 1 and 2. For parametrization of nonbonded interactions between XYR−XYR and XYR−BER, an iterative procedure was used. Along with the BER−BER parameters developed above, an initial guess was made for XYR−XYR, and then BER−XYR was fixed using eq 2. This parameter set was tested and then the procedure repeated until the properties of pure toluene were in acceptable agreement with experimental values as shown in Table 4. For the interaction between water and toluene, the BER−BER, XYR−XYR, BER−XYR, and BER−W nonbonded interaction parameters developed above were used, leaving only the XYR−W interaction to be fixed. This interaction was adjusted in order to reproduce experimental interfacial tension data between toluene and water (see Table 4). The mapping for *p*-xylene is shown on the right side of Figure 2. Here, two XYR sites (blue) and two BER sites (black) are used to model *p*-xylene. No further parametrization was needed for the *p*-xylene model. The parameters developed to this point show a high level of transferability (with regard to being applicable to molecules other than the one used in the parametrization, for example, using the BEN CG site in CG toluene) and accurately reproduced the bulk behavior of *p*-xylene, as shown in Table 4.

TABLE 4: Density (DENS), Surface Tension (ST), and Interfacial Tension (IT) Measurements of Various Molecules from CG MD Simulations Using the Current Parameter Set Compared to Experiment[a]

	DENS (g/cm^3)		ST (mN/m)		IT (mN/m)	
species	exp	sim	exp	sim	exp	sim
benzene	0.87	0.87	28	28	35	38
toluene	0.86	0.89	27	26	36	37
p-xylene	0.85	0.89	27	27	37	38
phenol	1.04	1.04	39	38	NA	NA
p-cresol	1.03	1.04	35	35	NA	NA

[a] The surface tension is between the species and air interface, while the interfacial tension is between the species and water.

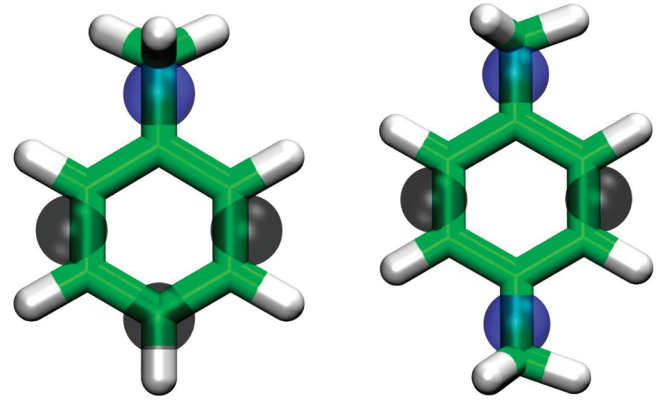

Figure 2. The CG mapping of toluene (left) and *p*-xylene (right) are shown here with spheres representing the CG sites BER (black) and XYR (blue) superimposed on the AA models.

D. Phenol and *p*-Cresol. The phenol and *p*-cresol models require a new CG site labeled PHR as well as the existing BER and XYR (for phenol) sites (see Figure 3). Parametrization of species that have melting points above room temperature is, in general, an issue for the current approach. As discussed in section II, this results from the fact that the parameters developed here are effective potentials valid at the temperature that they were parametrized for, which for the current model is 303 K. However, phenol and *p*-cresol have melting temperatures of 313.65 and 308.65 K, respectively. Fortunately, these temperatures are not significantly higher than the target temperature, and this is not expected to cause significant problems. For example, with the current water model, a temperature shift of 20 K causes less than 0.1% deviation in the density compared

TABLE 3: Nonbonded Parameters, ϵ (kcal/mol) and σ (Angstroms), Based on the Functional Forms (Denoted with Subscripts) Given in Equation 1[a]

(ϵ/σ)	W	CM	CT	CT2	BER	PHR	XYR
W	$0.895/4.371_{12-4}$	$0.340/4.4385_{12-4}$	$0.360/4.478_{12-4}$	$0.290/4.2960_{12-4}$	$0.320/4.085_{9-6}$	$0.680/3.935_{9-6}$	$0.480/4.185_{9-6}$
CM	$0.340/4.4385_{12-4}$	$0.420/4.5060_{9-6}$	$0.444/4.5455_{9-6}$	$0.3620/4.3635_{9-6}$	$0.3100/4.1530_{9-6}$	−	−
CT	$0.360/4.478_{12-4}$	$0.444/4.5455_{9-6}$	$0.469/4.585_{9-6}$	$0.383/4.403_{9-6}$	$0.3000/4.193_{9-6}$	$0.2000/4.043_{9-6}$	$0.5400/4.293_{9-6}$
CT2	$0.290/4.2960_{12-4}$	$0.362/4.3635_{9-6}$	$0.383/4.403_{9-6}$	$0.312/4.2210_{9-6}$	$0.2370/4.0105_{9-6}$	−	−
BER	$0.320/4.085_{9-6}$	$0.310/4.1530_{9-6}$	$0.300/4.193_{9-6}$	$0.2370/4.0105_{9-6}$	$0.1750/3.800_{9-6}$	$0.3102/3.650^{*}_{9-6}$	$0.2370/3.900^{*}_{9-6}$
PHR	$0.680/3.935_{9-6}$	−	$0.200/4.043_{9-6}$	−	$0.3102/3.650_{9-6}$	$0.5500/3.500_{9-6}$	$0.4195/3.750^{*}_{9-6}$
XYR	$0.480/4.185_{9-6}$	−	$0.540/4.293_{9-6}$	−	$0.2370/3.900^{*}_{9-6}$	$0.4195/3.750^{*}_{9-6}$	$0.3200/4.000_{9-6}$

[a] The * indicates that combination rules given from eq 2 were used, and − are for interaction parameters that were not yet developed. The parameters for self-interactions and cross interactions between CT, CM, CT2, and W were previously developed and reported.[10]

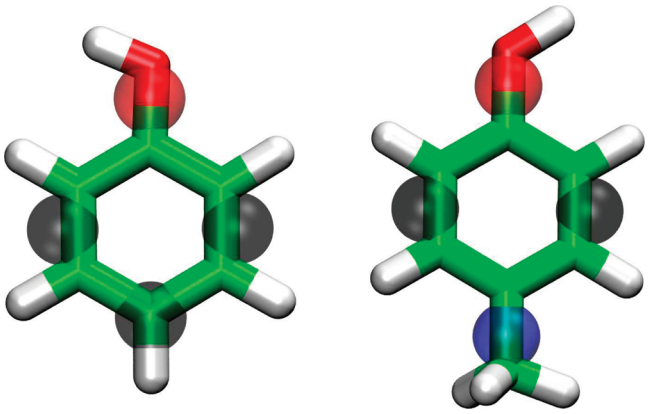

Figure 3. The CG mappings for phenol (left) and *p*-cresol (right) are shown here with spheres representing the CG sites BER (black), XYR (blue), and PHR (red) superimposed on the atomistic models.

TABLE 5: Solvation and Transfer (Water to Hexane) Free Energies (kcal/mol) from CG MD Simulations Compared to Experiment

	solvation		water/hexane	
species	exp	sim	exp	sim
benzene	−1	−1	−5	−5
toluene	−1	−1	−7	−7
p-cresol	−3	−3	−3	−3

to the experimental value and no deviation for surface tension. In order to ensure we are well within the liquid range for these species, temperatures slightly above the melting points, 318.15 and 313.15 K, were used for parametrization of the phenol and *p*-cresol molecules, respectively. The same iterative procedure used for the XYR site was used here for the PHR site. An initial guess was made for PHR−PHR; using these parameters, the BER−PHR and XYR−PHR interactions were set using eq 2. This parameter set was tested and the procedure repeated until acceptable agreement with experimental data was achieved. Table 1 shows that the PHR CG site performs very well for both phenol and *p*-cresol. For solutes miscible in water, interfacial tension with water is not a viable target to use for fitting the water interaction (PHR−W). Instead, we can take advantage of solvation free-energy data. Using the BER−W and XYR−W interactions from above, the PHR−W interaction was fit to reproduce the experimental hydration free energy of *p*-cresol (see Table 5). This calculation was performed at 303 K.

IV. Results

The thermodynamic data for various molecules calculated from CG MD simulations are compared to the experimental values in Table 4. The CG interaction parameters developed here are summarized in Tables 1−3. Here, we show the results for comparison of the CG model with AA MD simulations and experiment.

A. Radial Distribution Functions. Bulk simulations of benzene, toluene, and *p*-cresol were performed using both CG and AA representations. The RDFs, $g(r)$, from these simulations were calculated and compared to evaluate the structural properties of the CG model. In the AA simulations, the RDF was calculated by mapping the center of mass of all of the atoms to a single CG site and then calculating the RDF under this mapping. The same approach was used for the CG model. It should be noted that AA MD simulations were not used in the parametrization of the nonbonded interactions of the CG MD model.

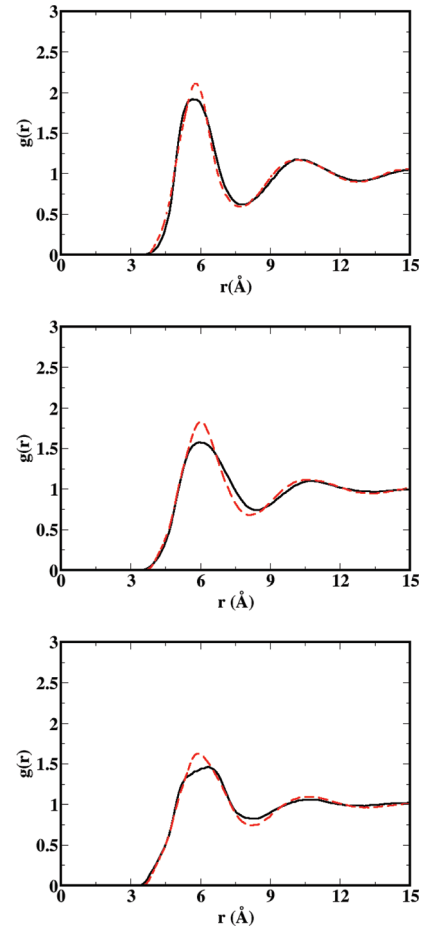

Figure 4. Shown here are the radial distribution functions, $g(r)$, based on the center-of-mass distances for the molecules. The results for bulk solutions of benzene (upper panel), toluene (middle panel), and cresol (lower panel) are plotted with AAMD results shown in black and CGMD results shown in red.

Figure 4 shows the comparison of the RDF calculations from AA MD and CG MD for benzene, toluene, and *p*-cresol. The upper panel of Figure 4 compares the benzene RDFs from the CG simulation (red dashed line) and the AA simulations (black solid line). The line shapes are in excellent agreement with the CG model, displaying a slightly larger magnitude in the first neighbor peak. To evaluate the effect of the difference in magnitude, the number of CG molecules, N, within the distance of the first minimum was calculated via eq 7.

$$N = \int_0^R \rho g(r) 4\pi r^2 \, dr \tag{7}$$

In eq 7, R is the first minimum in the RDF, taken as 7.5 Å for benzene. Calculation of N for AA MD and CG MD gave, respectively, values of $N = 11.9$ and 12.0, which indicates that the slight differences in magnitude of the first neighbor peaks do not lead to a large difference in the local density. Comparison of the toluene RDFs from AA MD (black solid line) and CG MD (red dashed line) is shown in the center panel of Figure 4. Again, a high level of agreement is displayed, with the line shape from AA and CG showing little differences. The CG MD result displays a slightly larger magnitude in the first neighbor peak, as in the case of benzene. Further, a slight shift in the position of the first minimum is noted, albeit very small (less than 0.5 Å). Integration of the RDF using eq 7 yields values of

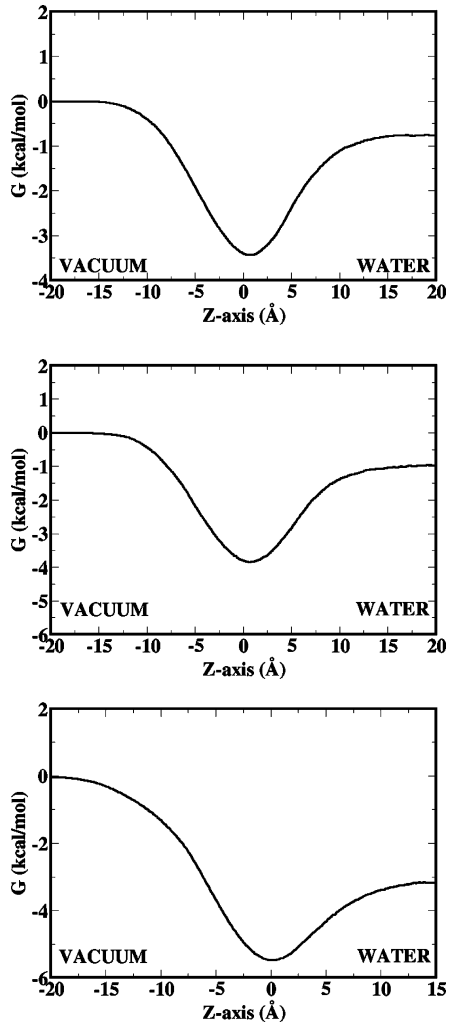

Figure 5. The free-energy profile for transferring benzene (upper panel), toluene (center panel), and *p*-cresol (lower panel) from vacuum to bulk water. The labels indicate the orientation of the system with respect to the reaction coordinate (*Z*-axis). The target values used for benzene, toluene, and *p*-cresol were −1, −1, and −3 kcal/mol, respectively.

$N = 13.5$ and 13.7 for AA MD and CG MD, respectively, with $R = 8.5$ Å. The lower panel of Figure 4 compares the RDFs for *p*-cresol from AA MD (black solid line) and CG MD (red dashed line). The overall line shape from the two methods shows excellent agreement in this case as well. Again, a slightly larger magnitude in the first neighbor peak is demonstrated in the CG MD result, as in the case of benzene and toluene. Further, the first minimum is slightly lower in magnitude for the CG model compared to that for AA MD. These slight deviations in the RDF do not lead to a large difference in the number of nearest neighbors with $N = 12.0$ and 12.0 for AA MD and CG MD, respectively. Overall, the CGMD demonstrates a high level of agreement with the AAMD simulations, with near-quantitative agreement for all three systems.

B. Solvation Free Energy. The BER−W and XYR−W interactions were set using the interfacial tension data for benzene, toluene, and *p*-xylene. To test the effectiveness of these interactions and the validity of creating molecules from individual CG sites, solvation free-energy calculations were performed. The solvation free-energy profiles for benzene and toluene are shown in Figure 5. The labels indicate the vacuum and water regions, and the interface is positioned at 0 Å on the *x*-axis. The result for benzene, upper panel of Figure 5, indicates that the solvation free energy predicted by the CG MD model is −0.75 kcal/mol, which is in reasonable agreement with the experimental values, which range from −0.76 (uncorrected) to +2 kcal/mol (corrected).[35−37] The minimum located at the air−water interface indicates that the CG benzene model is surface-active in water/vacuum systems, with an adsorption free energy (transfer from bulk water to the interface) of roughly −2.7 kcal/mol, which is in reasonable agreement with previous AA MD simulations for benzene.[43] The center panel of Figure 5 shows the free-energy profile for transferring toluene from vacuum to water. The CG MD model gives a value of −0.97 kcal/mol, which again shows reasonable agreement with the experimental values ranging from −0.8 (uncorrected) to +2 kcal/mol (corrected).[35−37] As with benzene, the free-energy profile indicates that toluene is surface-active. The adsorption free energy, −2.9 kcal/mol, is slightly larger for toluene as compared to that for benzene, which is to be expected. Finally, the PHR−W interaction was set using the volume-corrected solvation free energy value for *p*-cresol, −3.2 kcal/mol.[35−37] The profile from the final fitting is shown in the lower panel of Figure 5. An adsorption free energy of −2.3 kcal/mol is shown in Figure 5, which is comparable with the experimental value of −4.4 kcal/mol.[44]

C. Transfer Free Energy. The CG models presented here were parametrized to interact with an alkane model previously developed in our group.[10] The resulting free-energy profiles for transferring benzene, toluene, and *p*-cresol from hexane to water are shown in Figure 6. In these plots, the water and alkane regions are labeled accordingly, and the interface is positioned at 0 Å on the *x*-axis of the plot. The level of detail obtained can be seen in Figure 6, where the water-to-hexane transfer free energy (black line) for benzene is −5.2 kcal/mol, which is in agreement with the volume-corrected experimental value of −5.4 kcal/mol.[35−39] Further, the interfacial activity of benzene at the hexane/water interface is predicted to be small as no significant minimum appears at the interface between the two solvents. Also shown in the upper panel of Figure 6 is the PMF for transferring benzene from water to heptane (red line with circles). Since the transfer free energy of a solute from water to alkane should demonstrate little dependence on the specific solvent alkane, it is expected that the water/heptane transfer free energy should be similar to that of water/hexane for the model. As demonstrated in Figure 6, this is the case. The transfer free-energy profile (water-to-hexane) for toluene is shown in Figure 6. The CG MD model was parametrized to give a value of −7 kcal/mol, which is in agreement with the experimental value of −6.9 kcal/mol.[35−39] Again, no interfacial activity is predicted for toluene at the interface existing between hexane and water. The volume-corrected experimental transfer free-energy data (water-to-hexane) for *p*-cresol give a value of −2.9 kcal/mol in going from water to hexane. As shown in the profile, the CG MD force field was parametrized to give a transfer free energy of −3 kcal/mol, which is within acceptable agreement with the experimental value. Unlike benzene and toluene, the CG MD *p*-cresol model demonstrates a substantial adsorption free energy (bulk hexane to interface) of approximately −1.8 kcal/mol, indicating interfacial activity which is consistent with atomistic calculations of partitioning in a lipid bilayer.[35−39,45]

D. Dimerization Free Energy. As indicated in the parametrization section, self-interactions between the phenyl-based CG sites are set using simulations of the pure target molecule. The water interactions are then set using interfacial tension data with water or water/vacuum transfer (solvation) free-energy data. This does not directly take into account the relative interaction

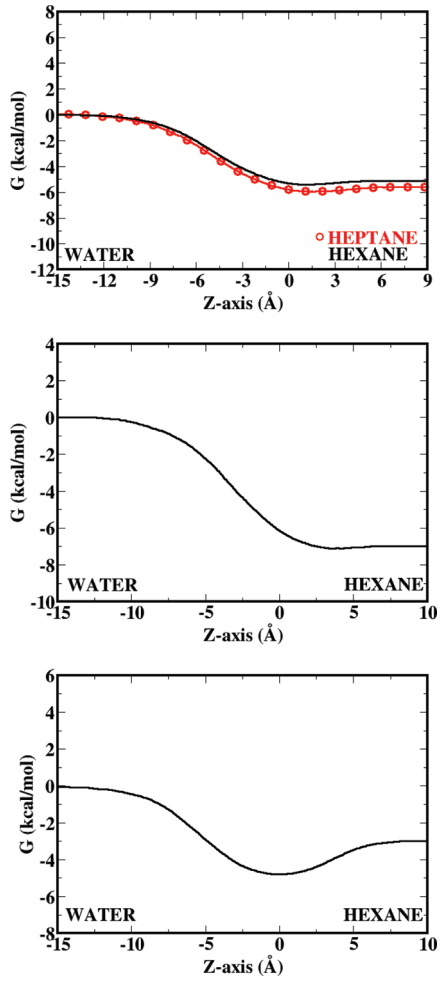

Figure 6. The water-to-hexane (black line) and water-to-heptane (red line with circles) transfer free-energy profiles for benzene are shown in the upper panel. The center and lower panels show the water-to-hexane transfer free-energy profiles for toluene and *p*-cresol, respectively. The water and alkane regions are labeled accordingly. The target values used for benzene, toluene, and *p*-cresol were −5, −7, and −3 kcal/mol, respectively.

between the target molecules in water. In order to test the balance of interaction strengths, PMFs were calculated for separating dimers in an aqueous environment via AA MD and CG MD. The calculations were also performed using the MARTINI CG model for comparison.[9] These calculations were performed on two analogues of the amino acid side chains, toluene and *p*-cresol, as applications of these models to protein simulations is one of the primary goals of our force field development. The PMF for separating two toluene molecules in water is shown in Figure 7 for AA MD (black solid line), CG MD (red dashed line), and the MARTINI model (green squares). The CG MD results are shown to demonstrate essentially quantitative agreement with the AA MD result. Both methods suggest a very shallow minima of approximately −0.8 kcal/mol at around 6 Å and a relatively flat profile beyond 8 Å. Further agreement in the repulsive region of the PMF (below 4 Å) demonstrates that the effective size of the toluene predicted by CG MD is equivalent to AA MD results. Compared to the AA MD results, the MARTINI model (green squares) demonstrates poor agreement with a deeper minimum (roughly −1.8 kcal/mol) located at a smaller separation distance (4 Å) as well as a slight barrier to dimerization (∼0.3 kcal/mol) at around 6.6 Å.

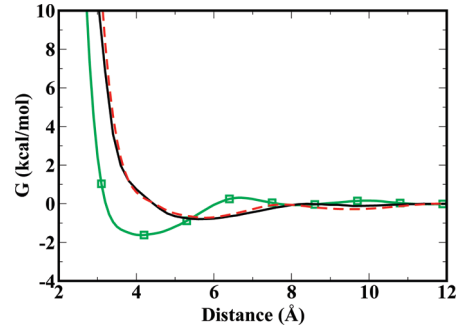

Figure 7. Dimerization free energy between two toluene (PHE side-chain analogue) molecules in an aqueous environment. AA MD results are shown with the solid (black) line, CG MD results are shown with the dashed (red) line, and MARTINI results are shown with the solid (green) line marked with squares. The PMF is plotted as a function of distance between the center of mass of the two toluene molecules.

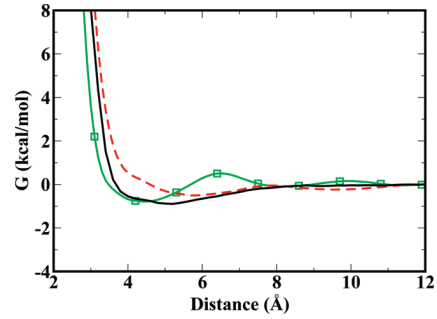

Figure 8. Dimerization free energy between two *p*-cresol (TYR side-chain analogue) molecules in an aqueous environment. AA MD results are shown with the solid (black) line, CG MD results are shown with the dashed (red) line, and MARTINI results are shown with the solid (green) line marked with squares. The PMF is plotted as a function of distance between the center of mass of the two *p*-cresol molecules.

The dimerization PMF data can be compared to experimental data via McMillan−Mayer theory, which relates the PMF, $w(r)$, to the osmotic second virial coefficient, B.[46,47] Although we were not able to locate data for toluene, an experimental value for benzene has been measured, $B = -1.2 \times 10^3$ Å at 308.15 K.[48] Using eq 8

$$B = -2\pi \int_0^\infty r^2[\exp(-w(r)/k_bT) - 1]dr \qquad (8)$$

where the PMF, $w(r)$, is written as a function of the distance, r, between the COM of the two molecules, a value of $B = -1.4 \times 10^3$ Å was calculated for toluene with the CG MD model at 303 K.[46,47,49] The PMF and B are also in good agreement with previous AA MD calculations performed on toluene.[49]

The PMF for separating two *p*-cresol molecules in water is shown in Figure 8 for AA MD (black solid line), CG MD (red dashed line), and the MARTINI model (green squares). Comparison between the CG MD and AA MD results here shows more qualitative agreement compared to the results for toluene. Both predict essentially no barrier to dimerization; however, the AA MD results suggest stronger dimerization of *p*-cresol with a minimum of ∼−0.9 kcal/mol while the CG MD model predicts a minimum of ∼−0.25 kcal/mol. Further, the location of the minimum in the CG MD model is shifted to a slightly higher separation distance of ∼6 Å compared to the AA MD result of ∼5 Å. The models show similar behavior at short range (less than ∼3 Å), with both models predicting similar behavior on the repulsive wall. As seen in Figure 8, the MARTINI model

(green squares) shows more reasonable agreement for *p*-cresol than for toluene. As a result of the redundant nature of the MARTINI potential parameter set, the results for toluene and *p*-cresol are similar. The MARTINI model demonstrates reasonable short-range (<5 Å) agreement with the AA MD results. Similar to toluene, it predicts a barrier to dimerization of ~0.5 kcal/mol. Nonetheless, the current CG MD model shows significant improvement of the thermodynamics, structure, and energetics over the MARTINI model for both toluene and *p*-cresol.

V. Conclusions

A systematic approach has been used to parametrize CG MD models for phenyl-based molecules. This class of molecules is important for applications including substituents of the amino acids PHE, TYR, and TRP, although the latter was not discussed here, as well as other applications such as carbon nanotubes, buckyballs, and larger carbon nanospheres. The models were parametrized against experimental thermodynamic data including solvation free energy, hexane/water transfer free energy, density, and surface tension. Further, despite the fact that AA MD simulations were not used in the parametrization, the models demonstrate a capacity to qualitatively reproduce AA simulation structural data. The calculation of the free energy of dimerization was also compared to AA MD simulations and experiment. These results indicate that the model quantitatively predicts the toluene dimerization free energy while qualitatively reproducing the same for *p*-cresol. It should be pointed out that the experimental values for *p*-cresol are not well-established. The current CG potentials show improvement over the MARTINI model when compared with AA MD, which suggests that they will be useful to provide insights into problems that are currently challenging with AA MD simulations.

Acknowledgment. The authors thank NSF, NIH, and Procter & Gamble for generous support. W.S. is grateful for the support of CREST-JST and the Next Generation Super Computing Project, Nanoscience Program, MEXT, Japan. Also, the authors would like to thank Axel Kohlmeyer for his helpful discussions and technical support.

References and Notes

(1) Levitt, M.; Warshel, A. *Nature* **1975**, *253*, 694–698.
(2) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464–4470.
(3) Klein, M. L.; Shinoda, W. *Science* **2008**, *321*, 798–800.
(4) Voth, G. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC: Boca Raton, FL, 2008.
(5) Zhou, J.; Thorpe, I. F.; Izvekov, S.; Voth, G. A. *Biophys. J.* **2007**, *92*, 4289–4303.
(6) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
(7) Noid, W. G.; Chu, J. W.; Ayton, G. S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 4116–4127.
(8) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
(9) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.
(10) Shinoda, W.; DeVane, R.; Klein, M. L. *Mol. Simul.* **2007**, *33*, 27–36.
(11) Shinoda, W.; DeVane, R.; Klein, M. L. *Soft Matter* **2008**, *4*, 2454–2462.
(12) DeVane, R.; Shinoda, W.; Moore, P.; Klein, M. *J. Chem. Theory Comput.* **2009**, *5*, 2115–2124.
(13) Bhargava, B. L.; DeVane, R.; Klein, M. L.; Balasubramanian, S. *Soft Matter* **2007**, *3*, 1395–1400.
(14) Stevens, M. J. *J. Chem. Phys.* **2004**, *121*, 11942–11948.
(15) Stevens, M. J.; Hoh, J.; Woolf, T. *Biophys. J.* **2002**, *82*, 542A–542A.
(16) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. L. *J. Chem. Phys.* **2003**, *119*, 7043–7049.
(17) Khurana, E.; DeVane, R.; Kohlmeyer, A.; Klein, M. L. *Nano Lett.* **2008**, *8*, 3626.
(18) Wong-Ekkabut, J.; Baoukina, S.; Triampo, W.; Tang, I. M.; Tieleman, D. P.; Monticelli, L. *Nat. Nanotechnol.* **2008**, *3*, 363–368.
(19) Izvekov, S.; Violi, A.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 17019–17024.
(20) Zacharopoulos, N.; Vergadou, N.; Theodorou, D. N. *J. Chem. Phys.* **2005**, *122*, 244111.
(21) Golubkov, P. A.; Ren, P. *J. Chem. Phys.* **2006**, *125*, 064103.
(22) Abrams, C. F.; Delle Site, L.; Kremer, K. *Phys. Rev. E* **2003**, *67*, 021807.
(23) Srinivas, G.; Klein, M. L. *Nanotechnology* **2007**, *18*, 9.
(24) Lorenz, S.; Walsh, T. R.; Sutton, A. P. *J. Chem. Phys.* **2003**, *119*, 2903–2907.
(25) Chiu, C.; DeVane, R.; Klein, M.; Shinoda, W.; Moore, P.; Nielsen, S. *J. Phys. Chem. B* **2009**, doi: 10.1021/jp9117375.
(26) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
(27) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
(28) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(29) Allen, M.; Tildesley, D. *Computer Simulation of Liquids*; Oxford Science: New York, 1987.
(30) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117–1157.
(31) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.
(32) Louis, A. A. *J. Phys.: Condens. Matter* **2002**, *14*, 9187–9206.
(33) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. *J. Chem. Phys.* **2007**, *126*, 144509.
(34) Krishna, V.; Noid, W. G.; Voth, G. A. *J. Chem. Phys.* **2009**, *131*, 12.
(35) Sharp, K.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686–9697.
(36) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849–855.
(37) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664–1670.
(38) Schulte, J.; Durr, J.; Ritter, S.; Hauthal, W. H.; Quitzsch, K.; Maurer, G. *J. Chem. Eng. Data* **1998**, *43*, 69–73.
(39) Ruelle, P. *Chemosphere* **2000**, *40*, 457–512.
(40) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018–5035.
(41) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
(42) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904–2914.
(43) Dang, L. X.; Feller, D. *J. Phys. Chem. B* **2000**, *104*, 4403–4407.
(44) Castro, A.; Bhattacharyya, K.; Eisenthal, K. B. *J. Chem. Phys.* **1991**, *95*, 1310–1315.
(45) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. *Biophys. J.* **2008**, *94*, 3393–3404.
(46) Rossky, P. J.; Friedman, H. L. *J. Phys. Chem.* **1980**, *84*, 587–589.
(47) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.
(48) Tucker, E. E.; Christian, S. D. *J. Phys. Chem.* **1979**, *83*, 426–427.
(49) Chipot, C.; Jaffe, R.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 11217–11224.

JP9117369