Monte Carlo Simulations of the Hydration of Substituted Benzenes with OPLS Potential Functions

William L. Jorgensen* and Toan B. Nguyen

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Received 15 June 1992; accepted 18 August 1992

Intermolecular potential functions have been developed for use in computer simulations of substituted benzenes. Previously reported optimized potentials for liquid simulations (OPLS) for benzene and organic functional groups were merged and tested by computing free energies of hydration for toluene, *p*-xylene, phenol, anisole, benzonitrile, *p*-cresol, hydroquinone, and *p*-dicyanobenzene. The calculations featured Monte Carlo simulations at 25°C and 1 atm with statistical perturbation theory. The average difference between the computed results and experimental data for the absolute free energies of hydration is 0.5 kcal/mol. The AM1-SM2 method is also found to perform well in predicting the free energies of hydration for the substituted benzenes. In addition, the Monte Carlo simulations provided details on the hydration of the substituted benzenes, in particular for the solute–water hydrogen bonding. © 1993 by John Wiley & Sons, Inc.

INTRODUCTION

A variety of force fields has evolved for classical computer simulations of organic and biomolecular systems.[1] Testing the functions in the prediction of well-characterized experimental observables can never be overdone; however, it may receive inadequate attention in the rush toward the latest challenging application on complex molecular systems. The development of techniques for the precise computation of free energies in solution has greatly enhanced the opportunities for meaningful comparisons of experiment and computation.[2] It has also provided exciting opportunities for computation to play a role in drug design through the prediction of relative binding affinities.[3] Success is medicinally important, important to the credibility of computational chemistry, and only possible if the underlying force fields for the computations provide an accurate description of the energetics of the molecular systems.

With this in mind, we continue to expand the development and testing of the optimized potentials for liquid simulations (OPLS) functions, which describe nonbonded interactions. Coverage now includes standard peptide residues,[1d] nucleotide bases,[4] many small ions, and numerous organic functional groups.[1d,5] The extension reported here treats a series of substituted benzenes, a common component in drugs and natural products.[6] They are also an important element in organic host–guest systems, especially cyclophane–arene complexes, for which we recently reported binding results.[7] This included predictions of free energies of binding for two guests, benzene and *p*-cresol, relative to *p*-xylene in water that were subsequently determined to agree well with values measured by nuclear magnetic resonance (NMR).[7b] Because these host–guest complexes feature benzene rings substituted by methyl, hydroxy, methoxy, and cyano groups, the calculations reported here were carried out prior to the binding studies to validate the ability of the potential functions to reproduce the observed free energies of hydration of the substituted benzenes. Errors in this area could undermine the binding studies because cyclophane–arene complexation has a substantial hydrophobic component that reflects the water solubility of the guests.[7]

The present work also extends the previous effort, in which an all-atom model for benzene was developed and shown to describe successfully many thermodynamic and structural properties for pure liquid benzene and benzene in water.[8] The current focus is on the free energies of hydration of the substituted benzenes and on their accommodation into the water network, especially the solute–water hydrogen bonding. A subsequent article details the results of pure liquid simulations for toluene, anisole, aniline, benzonitrile, and *m*-cresol that demonstrate the potential functions' success in providing liquid densities and heats of vaporization with less than 2% average errors.[9]

COMPUTATIONAL DETAILS

Intermolecular Potential Functions

The solutes considered here are benzene, toluene, *p*-xylene (*p*-dimethylbenzene), phenol, *p*-cresol

*Author to whom all correspondence should be addressed.

© 1993 by John Wiley & Sons, Inc.
CCC 0192-8651/93/020195-11

Table I. Standard geometric parameters for substituted benzenes.

C=C	1.400
C—H	1.080
C—O	1.360
C—CN	1.451
C—CH_3	1.510
C≡N	1.158
O—H	0.960
O—CH_3	1.430
<CCC	120.0
<CCH	120.0
<CCO	120.0
<COH	109.0
<$COCH_3$	117.0
<CC≡N	180.0

Bond lengths in Å, bond angles in degrees.

(p-methylphenol), hydroquinone (p-dihydroxybenzene), anisole, benzonitrile, and p-dicyanobenzene. Standard geometries based upon experimental data have been adopted, as summarized in Table I.[10] An all-atom representation is used except for methyl groups, which are treated as united atoms centered on carbon.

The potential energy between two molecules, ΔE_{ab}, consists of Coulomb and Lennard–Jones interactions between the atoms i on a and the atoms j on b, which are separated by a distance r_{ij} [eq. (1)].[1] Geometric combining rules are used for the Lennard–Jones parameters, ϵ and σ [eq. (2)]. The Lennard–Jones parameters are in general transferable

$$\Delta E_{ab} = \sum_i \sum_j \{q_i q_j e^2/r_{ij} + 4\epsilon_{ij}[(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6]\} \quad (1)$$

$$\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2} \qquad \sigma_{ij} = (\sigma_i \sigma_j)^{1/2} \quad (2)$$

to larger molecules from their components,[1] while the choice of partial charges for a new molecule is the main problem. The OPLS partial charges for neutral molecules have been optimized typically through iterative fluid simulations to reproduce experimental results for liquid structure and properties, especially densities and heats of vaporization.[1d,5] However, in the present case a remarkably simple approach was taken and then supported by the subsequent simulations of aqueous systems and pure liquids.

OPLS parameters have previously been reported for benzene,[8] alkanes,[11] alphatic alcohols,[12] ethers,[13] and acetonitrile.[14] For the substituted benzenes, the parameters for benzene and the substituents were simply merged. The Lennard–Jones parameters were taken without change, keeping the σ and ϵ for the *ipso*-carbon (the point of attachment of a substituent) the same as in benzene. The charge for the *ipso*-carbon was then taken as the charge in benzene (−0.115 e) plus the charge at the corresponding position in the functional group. And, the first atom in the functional group has to absorb the +0.115 e charge for the replaced hydrogen in benzene, e.g., for aliphatic alcohols, the charges on the hydroxy CH_n group and O are +0.265 and −0.700 e,[12] respectively, which give charges of −0.115 + 0.265 = 0.150 e for the *ipso*-carbon and +0.115 − 0.700 = −0.585 e for the O in phenols, while the hydroxy H remains unchanged at +0.435 e. This pattern of charge shifts is consistent with the greater acidity and weaker basicity characteristic of phenols relative to aliphatic alcohols. The OPLS parameters obtained in this way for substituted benzenes are summarized in Table II. The same approach may be taken for other substituted benzenes such as aniline, benzophenone, benzoates, phenyl esters, benzoic acid, *N*-phenyl amides, aryl sulfides, etc. using the reported OPLS parameters for the functional groups and benzene.[1d,5,8] Naturally, testing of the resultant parameters along the lines described here and for the pure liquids[9] is advised.

The OPLS charge distributions (Table II) and the standard geometries (Table I) can be combined to yield calculated dipole moments. The results in Debyes are as follows with experimental gas-phase values[15] in parentheses: toluene 0.24 (0.36), anisole 1.68 (1.38), phenol 2.11 (1.45), p-cresol 2.22 (1.71), and benzonitrile 3.23 (4.18). The level of accord is typical for the OPLS potentials and continues patterns that exist for aliphatic compounds. The OPLS dipole moments for aliphatic ethers and alcohols are also greater than experimental values for isolated molecules by 0.5–0.6 D.[12,13] Further, the OPLS dipole moment for acetonitrile is 3.44 D, while the experimental value is 3.92 D.[14] Coincidence between experimental and OPLS dipole moments is not sought in view of the limitations of the partial point-charge model and the desire for the OPLS potentials to focus on reproducing liquid-state properties.

An important aspect of the charge distributions that should be noted is that there is no charge alternation within the benzene ring, that is, for a mon-

Table II. OPLS parameters for substituted benzenes.

Atom	Example	q (e)	σ (Å)	ϵ (kcal/mol)
C	Benzene	−0.115	3.550	0.070
H	Benzene	0.115	2.420	0.030
CH_3	Toluene	0.115	3.800	0.170
O (H)	Phenol	−0.585	3.070	0.170
H(O)	Phenol	0.435	0.000	0.000
C(OH)	Phenol	0.150	3.550	0.070
O(R)	Anisole	−0.385	3.000	0.170
CH_3(O)	Anisole	0.250	3.800	0.170
C(OR)	Anisole	0.135	3.550	0.070
C(≡N)	Benzonitrile	0.395	3.650	0.150
N(≡C)	Benzonitrile	−0.430	3.200	0.170
C(CN)	Benzonitrile	0.035	3.550	0.070

osubstituted benzene all benzene carbons and hydrogens except the *ipso*-carbon have charges of −0.115 and +0.115 e. Resonance theory suggests that the charges at the *ortho*- and *para*-carbons should be more negative than at the *meta*-carbons when the substituent is π-donating as in phenol and anisole, while the opposite trends would apply for a π-accepting substituent such as a cyano group. Results of *ab initio* calculations support this idea and indicate that the magnitude of the charge variations for anizole is 0.1–0.2 e.[16] Potential functions for anisole were derived using the partial charges obtained from fitting to the electrostatic potential surface and tested in molecular dynamics simulations for the free energy of hydration of anisole and 1,2,3-trimethoxybenzene relative to benzene.[16] The extent of charge alternation is basis set dependent; however, this appears to be a less significant issue than the general magnitudes of the charges. The quality of the present results without charge alternation is comparable to the best results in that study, which used 6-31G(*d*)-derived charges, and far superior to the results with charges obtained from smaller basis sets. This observation along with the results for the larger number of solutes reported here and the results in the companion article[9] do not provide evidence that the simple model needs to be abandoned. The simple approach has also been taken in a previous molecular dynamics study of the hydration of trimethoprim and derivatives.[17]

Monte Carlo Simulations

Monte Carlo statistical mechanics simulations were carried out to determine the relative free energies of hydration of the benzene derivatives by interconverting them in TIP4P water[18] using statistical perturbation theory.[19,20] The interconversions shown in Figure 1 were performed. The absolute free energies of hydration from the gas phase were also established through the conversions to benzene because its absolute free energy of hydration had been determined by annihilating it in TIP4P water.[8]

The free energy changes were computed with the BOSS program[21] in the same manner as in the past.[5,20] The systems consisted of 500 TIP4P water molecules plus the solute in a cubic cell ca. 25 Å on a side with periodic boundary conditions. Metropolis and preferential sampling were used in the isothermal–isobaric ensemble at 25°C and 1 atm. A solute A was gradually converted to a solute B in a series of 5–10 separate simulations with a coupling parameter, λ, that ranged from 0 (A) to 1 (B). This yielded 10–20 free energy increments for each mutation via "double-wide" sampling, i.e., by perturbing from λ_i to $\lambda_i + \Delta\lambda$ and to $\lambda_i - \Delta\lambda$ for the simulation at λ_i.[20] The potential function parameters (q, σ, ϵ) and geometric variables (bond lengths and angles) were all

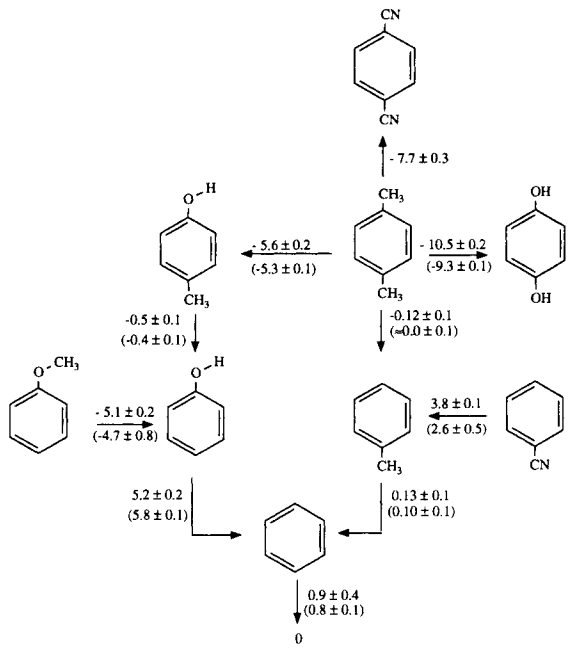

Figure 1. Computed and experimental (in parentheses) relative free energies of hydration in kcal/mol.

scaled linearly with λ. The free energy change between λ_i and λ_j, $G_j - G_i$, is given by eq. (3), where the average involves the

$$G_j - G_i = -k_B T \ln < \exp[-(E_j - E_i)/k_B T] >_i \quad (3)$$

total energy difference, $E_j - E_i$, obtained by sampling configurations selected for λ_i.[19] More simulations were in general used for larger energetic perturbations such as CH_3 or $OCH_3 \rightarrow$ OH and $CH_3 \rightarrow$ CN. Also, the spacing of the λ values was not uniform; they were concentrated in the regions where the free energy is changing most rapidly, e.g., near the more polar solute.[20] The selection of λ values is partially guided by insisting that the statistical uncertainty (1 σ) for a computed free energy increment not exceed 0.1 kcal/mol as obtained from the fluctuations in separate averages over 1×10^5 or 2×10^5 configurations.[22]

Each simulation entailed an equilibration period of at least 10^6 configurations followed by averaging over an additional 2×10^6 to 4×10^6 configurations. The longer periods are typically used for larger free energy changes. Attempted moves of the solute were performed every 120 configurations and volume changes were tried every ca. 3000 configurations. The intermolecular interactions were truncated at 8.5 Å for water–water interactions based upon the O—O distance and at 10.5 Å for solute–water interactions based upon the distances of a well-separated subset of solute atoms to O of water. The potential functions were quadratically feathered to zero over the last 0.5 Å. The ranges for translations and rotations of the solvent and solute molecules were se-

lected to give ca. 40% acceptance rates for new configurations.

The geometries of the individual solute and water molecules were not altered during the simulations except the torsional motion for the methoxy methyl group and hydroxy hydrogen in anisole and the phenols was included. A twofold torsional energy term [eq. (4)] was used with V_2 values of 2.2, 2.6,

$$V(\phi) = (V_2/2)(1 - \cos 2\phi) \qquad (4)$$

and 3.5 kcal/mol for anisole, hydroquinone, and phenol, respectively, which come from IR measurements and *ab initio* calculations.[23] The change in torsional energy is included in $E_j - E_i$ [eq. (3)] along with the change in solute–water energy. The net difference in free energy of hydration for cases where there is a torsional energy change then requires a correction for the change in the gas-phase free energy. This is obtained by mutating A to B in the absence of water and only involves sampling the dihedral angle. Specifically, from the thermodynamic cycle below, $\Delta\Delta G_{hyd} = \Delta G_{hyd}(B) - \Delta G_{hyd}(A) = \Delta G_{aq} - \Delta G_{gas}$. The gas-phase simulation was done for the mutation of anisole to phenol and provided

$$\begin{array}{ccc} A(g) & \xrightarrow{\Delta G_{gas}} & B(g) \\ \downarrow \Delta G_{hyd}(A) & & \downarrow \Delta G_{hyd}(B) \\ A(aq) & \xrightarrow{\Delta G_{aq}} & B(aq) \end{array}$$

0.168 ± 0.002 for ΔG_{gas}. The other cases are for the mutation of OH to united CH_3 and can be handled in a simpler way. The hydroxyl hydrogen becomes a dummy atom with $q = \sigma = \epsilon = 0$; however, the full torsional potential is retained throughout the mutation in water and the dihedral angle continues to be varied. The corresponding mutation in the gas phase yields $\Delta G_{gas} = 0$ because the torsional potential is the only contributor and is the same for reactant and product. ΔG_{aq} then contains the correct solvent effect, if any, from the torsion to the difference in free energies of hydration.

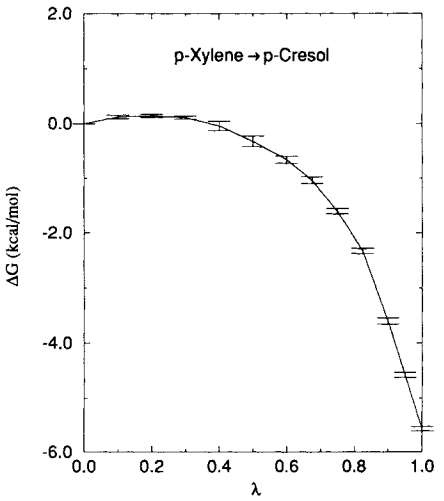

Figure 2. Change in free energy as a function of the coupling parameter λ for the mutation of *p*-xylene to *p*-cresol in the TIP4P model of water.

A typical plot of ΔG_{aq} vs. λ is shown in Figure 2 for *p*-xylene going to *p*-cresol. In this case, seven simulations were performed with $\Delta\lambda$s of 0.1 near *p*-xylene and of 0.05 near the more hydrophilic *p*-cresol; 71% of the free energy change occurs between $\lambda = 0.75$ and 1.0. The smoothness of the curve and the small statistical uncertainties for the individual free energy increments support the precision of the results.

RESULTS AND DISCUSSION

Free Eneregies of Hydration

The computed changes in free energies of hydration are compared with experimental data from solubility measurements[24-27] in Figure 1. We previously determined the absolute free energy of hydration of benzene from the ideal gas phase into TIP4P water as -0.9 ± 0.4 kcal/mol.[8] This value anchors the results in Figure 1 and allows computation of the absolute free energies of hydration for all the substituted benzenes, as summarized in Table III. The computed statistical uncertainties are $\pm 1\sigma$ and were obtained from the fluctuations in separate averages over blocks of 1×10^5 or 2×10^5 configurations.[22] The uncertainties obtained in this way are 0.1–0.2 kcal/mol for the mutations in Figure 1; however, a more realistic estimate of the uncertainties is ca. 0.4 kcal/mol. For example, the hysteresis for the cycle relating the five compounds in the center of Figure 1 is 0.8 kcal/mol, which could arise from five mutations with an uncertainty of $(0.8^2/5)^{1/2} = 0.36$ kcal/mol each. With this value, the uncertainty for the computed absolute free energies of hydration in Table III from the Monte Carlo simulations rises to ca. 0.6 kcal/mol.

There is also variation in the experimental results from different sources. The range is in general 0.1–0.2 kcal/mol; however, a recent measurement of the free energy of hydraton of benzene from the gas phase yielded -1.3 kcal/mol,[16] while earlier values are -0.767 and -0.887 kcal/mol.[24,25] The discrepancy between the recent result for anisole (-2.5 kcal/mol)[16] and the value from the extensive list of Hine and Mookerjee (-1.0 kcal/mol)[25] is particularly large. The present prediction is either -1.0 or -1.8 depending upon the route around the central cycle in Figure 1; the average has been listed in Table III. We actually favor the longer route, yielding -1.8 kcal/mol, because the phenol to benzene result is the most out of line with the experimental data. If the computed value for phenol to benzene was more

Table III. Absolute free energies of hydration.

Solute	$-\Delta G_{hyd}$ (kcal/mol)		
	Monte Carlo	Experimental	AM1-SM2
Benzene	0.9 ± 0.4	0.9,[a] 0.8,[b] 1.3[c]	0.5[d]
Toluene	1.0 ± 0.4	0.8,[a] 0.9[b]	0.3[d]
p-Xylene	0.9 ± 0.4	0.8[a]	0.0
Phenol	6.1 ± 0.4	6.5,[a] 6.4[e]	5.8[d]
p-Cresol	5.7 ± 0.4	6.1[a]	5.5
Hydroquinone	11.4 ± 0.5	10.1[e]	11.0
Anisole	1.4 ± 0.5	1.0,[a] 2.5[c]	2.3[d]
Benzonitrile	4.8 ± 0.5	(3.5)[f]	4.3
p-Dicyanobenzene	8.6 ± 0.5		6.9

For transfer of the solute from the 1*M* ideal gas state to 1*M* aqueous solution at 25°C and 1 atm.
[a]Ref. 25.
[b]Ref. 24.
[c]Ref. 16.
[d]Ref 28.
[e]Ref. 26.
[f]Estimated—see the text and ref. 27.

positive, similar to the *p*-cresol to *p*-xylene result, it would also reduce the hysteresis for the cycle. For benzonitrile, hydroquinone, and *p*-dicyanobenzene, the computed values in Table III correspond to the shorter path through *p*-xylene and toluene in Figure 1.

The experimental data in Table III come mostly from the compilations of Ben-Naim and Marcus[24] and Hine and Mookerjee[25] except for hydroquinone and benzonitrile, which they did not treat. Wolfenden et al. performed the measurement for hydroquinone,[26] and for benzonitrile, we estimated the experimental free energy of hydration from aqueous solubility[27a] and vapor pressure[27b] data following Ben-Naim's and Marcus' procedure.[24] Consequently, the uncertainty in the ΔG_{hyd} for benzonitrile is large, at least 0.5 kcal/mol. A more direct determination is needed along with a measurement for dicyanobenzene.

Table III also includes predictions from AM1-SM2 calculations,[28] which were carried out with the AMSOL 3.0 program.[29] This method combines semiempirical quantum mechanical calculations (AM1) with an extended continuum model including solute polarization for the solvation that is parameterized explicitly for water at 25°C. The results in Table III are from fully optimized geometries in the presence of the solvent field (AM1-SM2//AM1-SM2). The results for these solutes are changed no more than 0.1 kcal/mol if the gas-phase geometries are used (AM1-SM2//AM1). However, the computation time decreases dramatically, e.g., from 4.6 h to 0.5 min for benzene on a Silicon Graphics 4D/35 computer. Of course, the Monte Carlo simulations are still more time consuming, ca. 48 h for a complete mutation. Some significant advantages of the latter approach are the ability to switch trivially solvents, temperature, and pressure, the wealth of solute–solvent structural information obtained, the ability to sample internal degrees of freedom of the solutes during the simulations, and the classical correspondence of the calculations to classic statistical mechanics in well-defined ensembles.

In view of the uncertainties in the data, the accord between the Monte Carlo and measured absolute free energies of hydration in Table III is nearly perfect. The accord is in particular notable given the simple origin of the OPLS parameters for the substituted benzenes and the significant range of free energies of hydration. The worst discrepancies are for hydroquinone (1.3 ± 0.5 kcal/mol) and possibly benzonitrile (1.3 ± ca. 1 kcal/mol). The computed ΔG_{hyd} of −11.4 kcal/mol for hydroquinone reflects reasonable additivity of substituent solvation effects, namely, the computed $\Delta\Delta G_{hyd}$ is 5.2 ± 0.6 kcal/mol from benzene to phenol and 5.3 ± 0.6 kcal/mol from phenol to hydroquinone. The computed additivity is still good in comparing *p*-xylene to *p*-cresol (4.8 ± 0.6) and *p*-cresol to hydroquinone (5.7 ± 0.6). The additivity for the AM1-SM2 results is even better. The experimental data agree with the ca. 5 kcal/mol difference for introducing the first hydroxyl group. However, the experimental results show smaller differences for introduction of the second hydroxyl group, 3.7 kcal/mol for phenol to hydroquinone and 4.0 kcal/mol for *p*-cresol to hydroquinone, than the computations, 5–6 kcal/mol. The lack of additivity in the experimental results was attributed to "anticooperativity" in the opposing hydration of the *para* hydroxyl groups.[27] Such an effect, if it is real, should be in the computations. Because it is not observed in the computed results, a possible explanation is that the nonadditivity stems from an electronic reorganization in hydroquinone, that is, the approximation of using the same charge distributions for the mono- and disubstituted benzenes is somewhat flawed. Nevertheless, the AM1-SM2 results, which allow for the electronic reorganization, are still additive for the hydroxybenzenes.

The AM1-SM2 results are also in close agreement with the experimental data. The average error is 0.7 vs. 0.5 kcal/mol from the Monte Carlo simulations. In view of the uncertainties in the Monte Carlo and experimental results, the two computational approaches should be considered to have comparable reliability.

Among other interesting items in Figure 1 and Table III, the nearly identical free energies of hydration for benzene, toluene, and *p*-xylene are especially notable. It might be thought that changing an aromatic hydrogen for a methyl group would increase hydrophobicity. However, the results suggest that the penality for creating the larger cavity in water is matched by the enhanced Lennard–Jones (van der Waals) interactions between the larger group and water. This compensation is also observed for the *n*-alkane series.[24,25] In contrast, replacing a hydrogen by methyl in going from phenol to anisole makes the free energy of hydration less favorable by 4–5 kcal/mol. The more polar phenol participates in two to three strong hydrogen bonds with water, while the ether oxygen for anisole accepts one weaker hydrogen bond from water (*vide infra*). The additivity issue arises again for the cyanobenzenes. The Monte Carlo results are still additive, e.g., for the *p*-xylene to *p*-dicyanobenzene transformation (−7.7 kcal/mol) vs. twice the toluene to benzonitrile difference (−7.6). The AM1-SM2 results are less additive (−6.9 and −8.0 kcal/mol) and yield a less negative predicted ΔG_{hyd} for *p*-dicyanobenzene.

Solute–Water Structure

Besides the thermodynamic results, the present simulations also provided much information on the hydration of the substituted benzenes. The hydration of benzene itself has been addressed in several earlier studies.[8,30] There is agreement that the first solvation shell contains about 23 water molecules including 2 on either side of the ring with a hydrogen directed toward the ring center. The latter feature is also apparent in crystal structures of hydrated substituted benzenes.[31] Among the compounds studied here, the focus will be on the representative systems—*p*-cresol, anisole, and benzonitrile—with emphasis on characterizing the hydrogen bonding with the solute's functional group. Two types of distributions will be considered, radial distribution functions (rdfs) and energy pair distributions. The rdf $g_{xy}(r)$ gives the probability of occurrence of an atom of type y at a distance r from an atom of type x, normalized to account for the bulk density of the fluid. Consequently, peaks in rdfs are associated with solvation shells or specific neighbors and can be integrated to yield coordination numbers. For the present purposes, y is always the O or H of water and x is a site in the substituted benzene. The energy

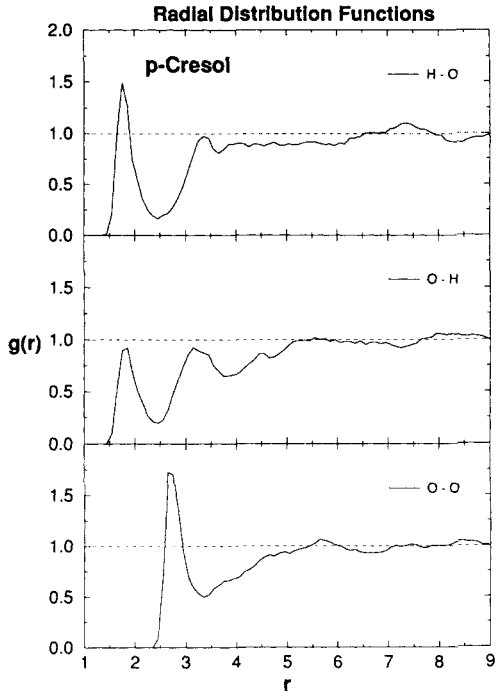

Figure 3. The O—O, O—H, and H—O radial distribution functions computed for *p*-cresol in water.

pair distributions considered here are for the solute–water interactions and show the average number of water molecules that have a given individual interaction energy with the sustituted benzene.

Hydration of p-Cresol

The key *rdfs* for *p*-cresol are shown in Figures 3–5. The hydrogen bonding is immediately addressed by the O—O, O—H, and H—O *rdfs* in Figure 3, where the first atom is the hydroxyl O or H in *p*-cresol and the second atom is the O or H of water. The first peaks in these distributions unequivocally reflect the hydrogen bonding in view of the distances involved. The clearest picture of the hydrogen bonding including the breakdown into acceptance and donation is provided by the O—H and H—O rdfs, respectively. Both have sharp first peaks that integrate to 1.5 for the O—H rdf and 1.0 for the H—O rdf out to the minima at 2.5 Å.* As usual,[12,18] these sum to a little less than the first peak of the O—O rdf, which contains 3.0 water neighbors to the minimum at 3.4 Å.

The overall impression is reinforced by the energy pair distribution in Figure 6. The low-energy band is assigned to the hydrogen-bonded water molecules and extends to −7.0 kcal/mol for the most attractive individual *p*-cresol–water interactions. A reasonable

*The first peak for the O—H rdf appears smaller than for the H—O rdf; however, the integral of the former is larger because the number density of water hydrogens is twice that for water oxygens.

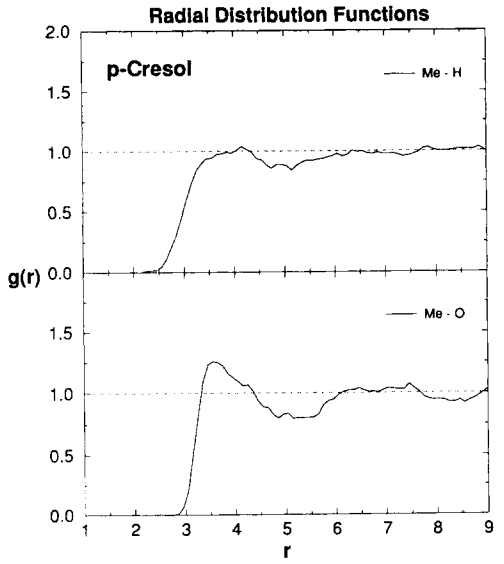

Figure 4. The CH_3—O and CH_3—H radial distribution functions computed for *p*-cresol.

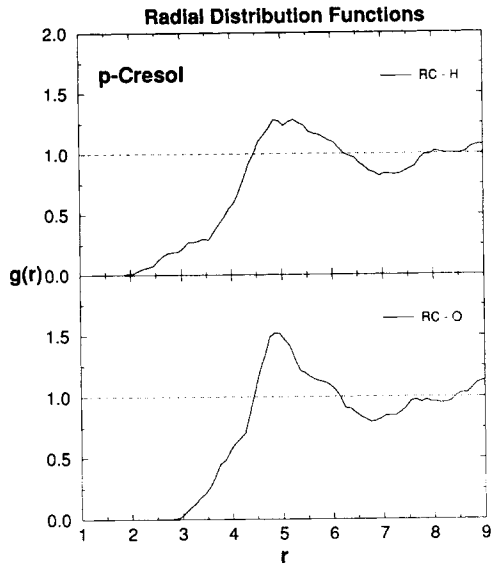

Figure 5. The ring center—O and ring center—H radial distribution functions computed for *p*-cresol.

energetic cutoff for hydrogen bonds between water and *p*-cresol is about −4.0 kcal/mol because the water–ring center interactions can be as low as −3.8 kcal/mol.[8*] The −4.0-kcal/mol limit is also close to the position of the minimum in the energy pair distribution that separates the low-energy group of waters from the many weaker interactions that cluster in the spike centered at 0 kcal/mol. Integration of the energy pair distribution to −4.0 kcal/mol gives 2.6 water molecules. Thus, the picture that emerges has the phenolic hydrogen always hydrogen bonded to a water molecule and the phenolic oxygen hydrogen bonded to an average of ca. 1.5 water molecules. This view is supported by displaying numerous instantaneous configurations of the system; Figure 7 provides an example with three hydrogen bonds to the hydroxyl group of *p*-cresol. This situation alternates about an equal amount with one having 2 hydrogen bonds to give the average of 2.5 hydrogen bonds. The switch from two to three hydrogen bonds is clear in following the total solute–water interaction energy during the simulation. It oscillates near −26 kcal/mol when there are two hydrogen bonds but drops as low as −32 kcal/mol when the third hydrogen bond appears. Forming the third hydrogen bond brings an energetic penalty for the water–water interactions; also, the present simulations show a rise in the torsional energy of the solute by ca. 0.3 kcal/mol in the triply hydrogen-bonded state. Apparently, the phenolic hydrogen has to rotate 40–60° out of plane to accommodate the three hydrogen bonds, as in Figure 7, where the dihedral angle is 51°. On the technical side, it is important that the free energy calculations adequately sample both states. For the *p*-xylene to *p*-cresol mutation from $\lambda = 0.95$ to 1.0 (*p*-cresol), the free energy change averages −1.0 kcal/mol, but it is near −1.3 kcal/mol while the system is in the triply hydrogen-bonded state. If only one state is sampled, it would lead to an error of about 0.2 kcal/mol just for this increment. The same effect was apparent for the *p*-xylene to hydroquinone mutation. In both cases, the averaging was extended to 4×10^6 configurations for the last window.

There is little structure in the rdfs beyond the first peaks in Figure 3. The second peak in the O—H *rdf* at 3.2 Å arises from the more remote hydrogen of

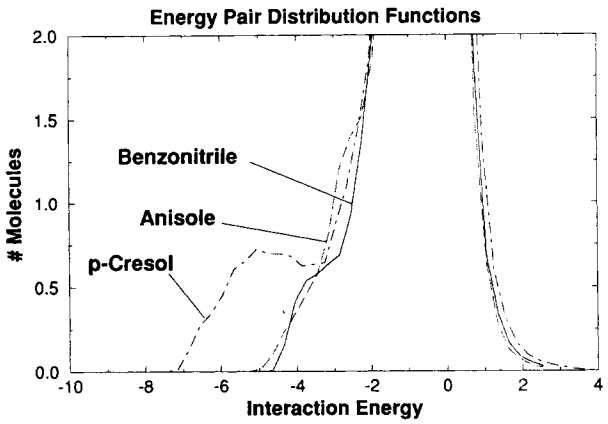

Figure 6. The solute–water energy pair distribution functions computed for *p*-cresol, anisole, and benzonitrile. The ordinate gives the number of water molecules that interact with the solute with the interaction energy given by the abscissa. Units for the ordinate are number of molecules per kcal/mol.

*We consider hydrogen bonds to feature specific *X*—H···Y interactions, where *X* and *Y* are normally atoms more electronegative than carbon. Consequently, the interaction between water and the ring center is not considered in this context.

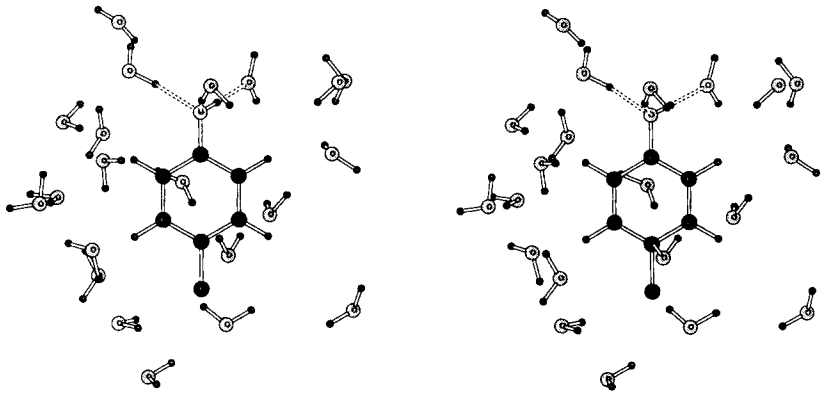

Figure 7. Stereoplot of a configuration taken from the simulation of *p*-cresol in water. Only water molecules with any atom within 3.5 Å of a solute atom are shown.

the linearly hydrogen-bonded water molecules. There is also a lack of striking features in the other rdfs for *p*-cresol in water including those in Figures 4 and 5. The CH_3—O and CH_3—H rdfs in Figure 4 are typical for hydrocarbons in water and have broad first peaks with summits near 4.0 Å.[32] The ring center—O and ring center—H rdfs are shown in Figure 5. The presence of water hydrogens in the favorable orientation near the ring centers does not create separate bands in these distributions; however, the small band out to 3.5 Å in the ring center—H rdf can be attributed to this arrangement. The integral to this point is 1.8 water hydrogens. Similarly, the shoulder in the ring center—O rdf out to 3.9 Å has the same origin and encompasses three water oxygens. The constant presence of water molecules poised over the ring center is clear in plots of configurations including Figure 7.

Hydration of Anisole

The rdfs for anisole in water that have the strongest features are the O—O and O—H rdfs in Figure 8 and the CH_3—O and CH_3—H rdfs in Figure 9. The first peaks in the O—O and O—H rdfs are not as striking as for *p*-cresol. The integral of the first peak in the O—H rdf to the minimum at 2.5 Å is now 1.2, and the integral of the O—O rdf to 3.4 Å is 1.7. The hydrogen bonds are also weaker for anisole, as reflected in Figure 6; the lowest energy interaction is −4.8 kcal/mol vs. −7.0 for *p*-cresol. Further, the hydrogen bonding interactions no longer form a distinct energetic band separate from the interactions between water and the ring. Displays of configurations also typically show just one hydrogen bond between water and the oxygen of anisole, as in Figure 10.

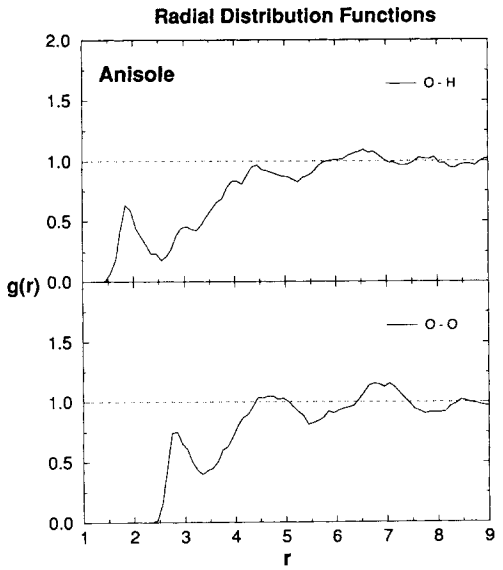

Figure 8. The O—O and O—H radial distribution functions computed for anisole in water.

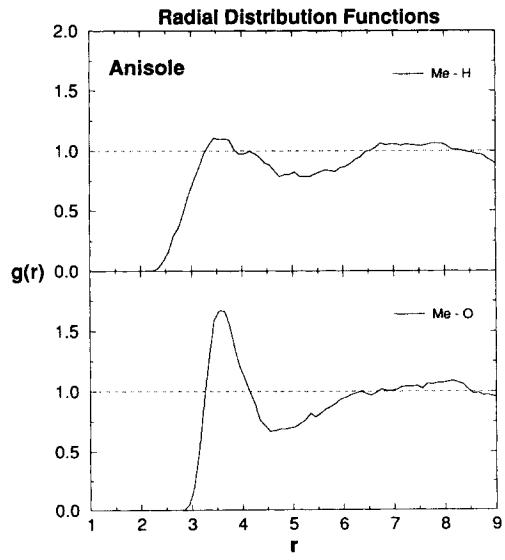

Figure 9. The CH_3—O and CH_3—H radial distribution functions computed for anisole in water.

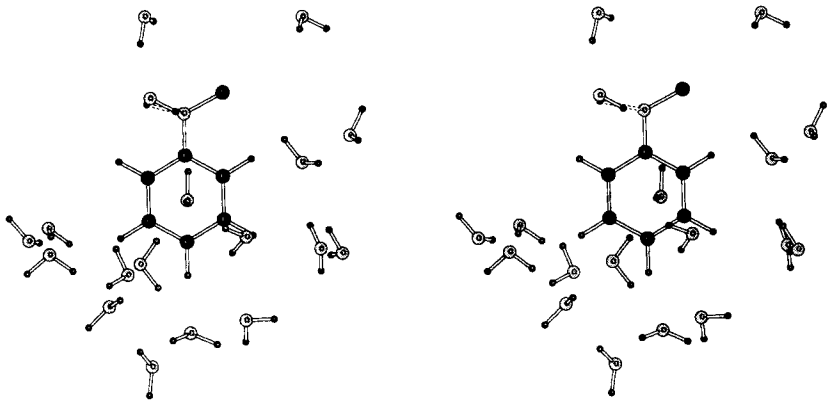

Figure 10. Stereoplot of a configuration taken from the simulation of anisole in water. Water molecules selected as in Figure 7.

The polarity of the ether functionality is also reflected in the CH_3—O rdf in Figure 9, which can be contrasted to the CH_3—O rdf for *p*-cresol in Figure 4. A significantly higher first peak and deeper first minimum are obtained for the ether. The integrals to 4.75 Å are 11 in both cases; however, the more positively charged ether methyl group influences water oxygens to be relatively closer to it. At the same time, there is little change in the average disposition of the water hydrogens, as evidenced by the similarity of the CH_3—H *rdfs* in Figures 4 and 9. Thus, water molecules in the first shell around the ether methyl group arrange in an electrostatically reasonable fashion.

Hydration of Benzonitrile

The quantity and energetics of the solute–water hydrogen bonding are similar for benzonitrile and anisole. The lowest individual interaction energy for benzonitrile and water is −4.5 kcal/mol in Figure 6. Again, the low-energy region for benzonitrile just merges into the central spike, rather than showing a distinct band as for *p*-cresol. Thus, the strength of any hydrogen bonds at the nitrogen does not differ significantly from the most attractive interactions between water and the aromatic ring, ca. −4 kcal/mol. The average total solute–water interaction energies are also less negative for benzonitrile (−21.6 ± 0.5 kcal/mol) and anisole (−21.1 ± 0.5 kcal/mol) than for *p*-cresol(−27.5 ± 0.5 kcal/mol).

Consistently, the rdfs for benzonitrile in water have little structure in comparison to those for *p*-cresol (Figs. 11 and 12). The N—H rdf in Figure 11 does show a well-resolved, small first band out to 2.5 Å, which integrates to 1.1 water molecules. The N—O rdf has a corresponding feature that is contained in the leading shoulder out to the flat area at 3.0–3.5 Å. The integral to this point encompasses 0.95 water neighbors. Thus, there is one water mol-

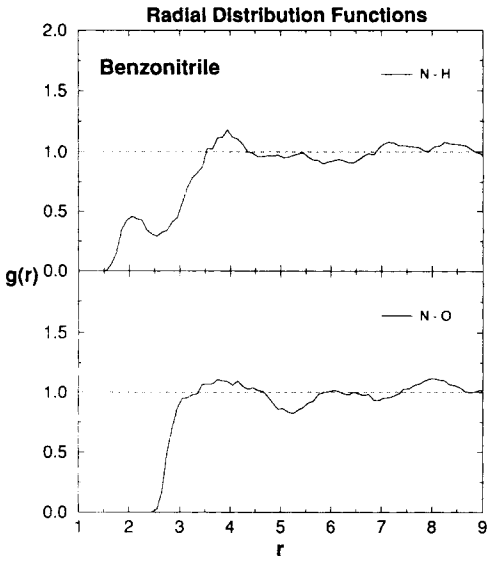

Figure 11. The N—O and N—H radial distribution functions computed for benzonitrile in water.

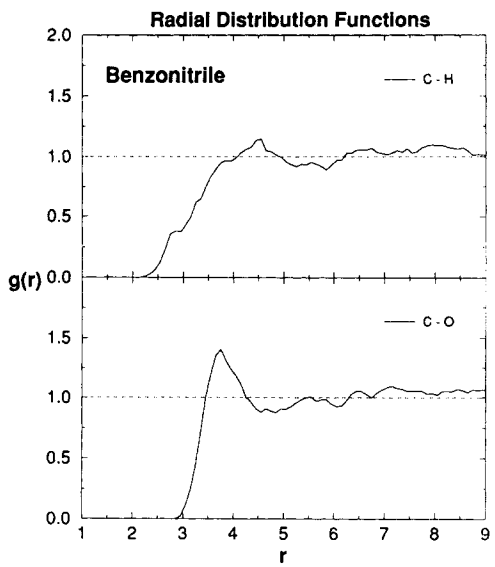

Figure 12. The C_N—O and C_N—H radial distribution functions computed for benzonitrile in water.

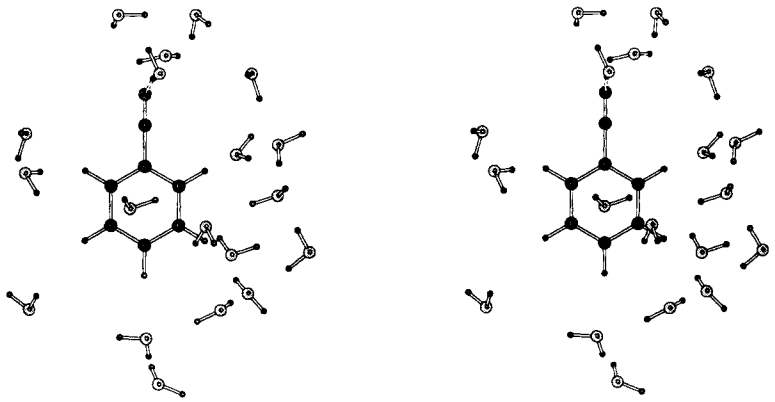

Figure 13. Stereoplot of a configuration taken from the simulation of benzonitrile in water. Water molecules selected as in Figure 7.

ecule acting as a hydrogen bond donor at the nitrogen. This is also clear from the display of configurations. For example, the hydrogen bond is apparent in Figure 13 along with the usual water molecules near the ring center. The C_N—O *rdf* in Figure 12 has a broad first peak out to 4.9 Å with the maximum at 3.8 Å and integrates to 12 waters. The peak is again stronger than for a less highly charged carbon such as the methyl group of *p*-cresol (Fig. 4).

CONCLUSION

In summary, OPLS potential functions for substituted benzenes have been developed by combining previously reported parameters for benzene and the corresponding aliphatic systems. Calculations of relative and absolute free energies of hydration for a series of aromatic solutes demonstrated generally excellent accord with experimental data. Further support for the model comes from simulations of the pure liquids for several substituted benzenes representative of the four substituents considered here; methyl, hydroxy, methoxy, and cyano.[9] The hydration of prototypical molecules was also characterized with emphasis on the solute–water hydrogen bonding. It was found that phenol has an average of 2.5 hydrogen bonds with water molecules, 1.0 as donor and 1.5 as acceptor, while anisole and benzonitrile accept only 1 hydrogen bond from water. In view of the variety of substituents considered so far and the quality of the results, extension of the present approach to other substituted benzenes is likely to be fruitful, although it should always be accompanied by testing on observable properties.

Gratitude is expressed to the National Science Foundation and National Institutes of Health for support of this research and to Professor Donald G. Truhlar for providing the AMSOL 3.0 program.

References

1. See, for example: (a) S.J. Weiner, P.A. Kolman, D.T. Nguyen, and D.A. Case, *J. Comp. Chem.*, **7,** 230 (1986); (b) B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, **4,** 187 (1983); (c) M. Levitt, *J. Mol. Biol.*, **168,** 595 (1983); (d) W.L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110,** 1657 (1988).
2. (a) D.L. Beveridge and F.M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18,** 431 (1989); (b) W.L. Jorgensen, *Acc. Chem. Res.*, **22,** 184 (1989).
3. (a) T.P. Straatsma and J.A. McCammon, *Annu. Rev. Phys. Chem.*, **43,** 407 (1992); (b) P.A. Kollman and K.M. Merz, Jr., *Acc. Chem. Res.*, **23,** 246 (1990); (c) W.L. Jorgensen, *Chemtracts Org. Chem.*, **4,** 91 (1991).
4. J. Pranata, S.G. Wierschke, and W.L. Jorgensen, *J. Am. Chem. Soc.*, **113,** 2810 (1991).
5. W.L. Jorgensen, J.M. Briggs, and L. Contreras, *J. Phys. Chem.*, **94,** 1683 (1990).
6. R.S. Silverman, *The Organic Chemistry of Drug Design and Drug Action*, Academic Press, New York, 1992.
7. (a) F. Diederich, *Angew. Chem. Int. Ed. Engl.*, **27,** 362 (1988); (b) W.L. Jorgensen, T.B. Nguyen, E.M. Sanford, I. Chao, K.N. Houk, and F. Diederich, *J. Am. Chem. Soc.*, **114,** 4003 (1992).
8. W.L. Jorgensen and D.L. Severance, *J. Am. Chem. Soc.*, **112,** 4768 (1990).
9. W.L. Jorgensen, E.R. Laird, T.B. Nguyen, and J. Tirado-Rives, *J. Comp. Chem.*, **14,** 206 (1993).
10. M.D. Harmony, V.W. Laurie, R.L. Kuczkowski, R.H. Schwendeman, D.A. Ramsay, F.J. Lovas, W.J. Lafferty, and A.G. Maki, *J. Phys. Chem. Ref. Data*, **8,** 619 (1979); G.M. Anderson, P.A. Kolman, L.N. Domelsmith, and K.N. Houk, *J. Am. Chem. Soc.*, **101,** 2344 (1979).
11. W.L. Jorgensen, J.D. Madura, and C.J. Swenson, *J. Am. Chem. Soc.*, **106,** 6638 (1984).
12. W.L. Jorgensen, *J. Phys. Chem.*, **90,** 1276 (1986).
13. J.M. Briggs, T. Matsui, and W.L. Jorgensen, *J. Comp. Chem.*, **11,** 958 (1990).
14. W.L. Jorgensen and J.M. Briggs, *Mol. Phys.*, **63,** 547 (1988).
15. R.D. Nelson, D.R. Lide, and A.A. Maryott, *Natl. Std. Ref. Data Serv. Natl. Bur. Std. US*, **10,** 1 (1967).
16. L.F. Kuyper, T.N. Hunter, D. Ashton, K.M. Merz, Jr., and P.A. Kollman, *J. Phys. Chem.*, **95,** 6661 (1991).
17. C.L. Brooks and S.H. Fleischman, *J. Am. Chem. Soc.*, **112,** 3307 (1990).

18. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *J. Chem. Phys.*, **79,** 926 (1983).
19. R.W. Zwanzig, *J. Chem. Phys.*, **22,** 1420 (1954).
20. W.L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, **83,** 3050 (1985).
21. W.L. Jorgensen, *BOSS*, v. 3.1, Yale University, New Haven, CT, 1991.
22. W.L. Jorgensen, *Chem. Phys. Lett.*, **92,** 405 (1982).
23. L. Radom, W.J. Hehre, J.A. Pople, G.L. Carlson, and W.G. Fateley, *J. Chem. Soc. Chem. Comm.*, 308 (1972); D.C. Spellmeyer, P.D.J. Grootenhuis, M.D. Miller, L.F. Kuyper, and P.A. Kollman, *J. Phys. Chem.*, **94,** 4483 (1990).
24. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81,** 2016 (1984).
25. J. Hine and P.K. Mookerjee, *J. Org. Chem.*, **40,** 292 (1975).
26. R. Wolfenden, Y.-L. Liang, M. Matthews, and R. Williams, *J. Am. Chem. Soc.*, **109,** 463 (1987).
27. (a) J.A. Riddick, W.B. Bunger, and T.K. Sakano, *Organic Solvents: Physical Properties and Methods of Purification*, 4th Ed., John Wiley & Sons, New York, 1986; (b) D.R. Stull, *Ind. Eng. Chem.*, **39,** 517 (1947).
28. C.J. Cramer and D.G. Truhlar, *Science*, **256,** 213 (1992).
29. C.J. Cramer, G.C. Lynch, and D.G. Truhlar, AMSOL 3.0, University of Minnesota, Minneapolis, MN, 1992.
30. P. Linse, G. Karlstrom, and B. Jonsson, *J. Am. Chem. Soc.*, **106,** 4096 (1984); G. Ravishanker, P.K. Mehrotra, M. Mezei, and D.L. Beveridge, *J. Am. Chem.Soc.*, **106,** 4102 (1984).
31. J.L. Atwood, F. Hamada, K.D. Robinson, G.W. Orr, and R.L. Vincent, *Nature* **349,** 683 (1991).
32. W.L. Jorgensen, *J. Chem. Phys.*, **77,** 5757 (1982).