Accuracy of Free Energies of Hydration for Organic Molecules from 6-31G*-Derived Partial Charges

Heather A. Carlson, Toan B. Nguyen, Modesto Orozco,* and William L. Jorgensen†

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

Absolute free energies of hydration have been computed for 13 diverse organic molecules using partial charges derived from *ab initio* 6-31G* wave functions. Both Mulliken charges and charges fit to the electrostatic potential surface (EPS) were considered in conjunction with OPLS Lennard–Jones parameters for the organic molecules and the TIP4P model of water. Monte Carlo simulations with statistical perturbation theory yielded relative free energies of hydration. These were converted to absolute quantities through perturbations to reference molecules for which absolute free energies of hydration had been obtained previously in TIP4P water. The average errors in the computed absolute free energies of hydration are 1.1 kcal/mol for the 6-31G* EPS charges and 4.0 kcal/mol for the Mulliken charges. For the EPS charges, the largest individual errors are under 2 kcal/mol except for acetamide, in which case the error is 3.7 kcal/mol. The hydrogen bonding between the organic solutes and water has also been characterized. © 1993 by John Wiley & Sons, Inc.

INTRODUCTION

Currently, the standard way of representing non-bonded interactions in molecular mechanics force fields utilizes Coulomb and Lennard–Jones interactions.[1] Beyond van der Waals contact, the Coulombic term is dominant for polar molecules and controls ion–molecule interactions and hydrogen bonding. Consequently, the choice of partial charges, which determine the Coulomb interactions, continues to receive much attention.[1-7] A particularly successful approach has been to fit the charges to reproduce structural and thermodynamic observations for pure liquids and aqueous solutions through iterative fluid simulations.[1c,8] However, this procedure is relatively time consuming and depends on the availability of experimental data. More direct approaches have been pursued as an alternative or supplement and have focused on deriving the charges from quantum mechanical calculations on isolated molecules.[2-7] Recent studies converged on partial charges obtained from fitting to the electrostatic potential surface (EPS) of *ab initio* 6-31G* calculations as a de facto standard.[2-5] Support for this choice has noted: (1) Dipole moments are over-estimated by 10–20% with 6-31G* calculations, which is desirable to compensate for the neglect of polarization effects with fixed-charge models[2,5]; (2) EPS charges are relatively insensitive to extension of the basis set beyond 6-31G* and inclusion of electron correlation[5,9]; and (3) 6-31G* EPS charges correlate well ($r = 0.93$) with OPLS charges for organic molecules, which have been derived to reproduce fluid properties.[10]

Though much effort has gone into comparing charges from alternative quantum mechanical approaches and obtaining appropriate scaling factors for converting charges from semiempirical molecular orbital calculations into charges similar to 6-31G* EPS ones, the viability of the 6-31G* EPS charges in fluid simulations has received little direct testing. This issue is addressed here by asking a key question: How accurate are computed free energies of hydration when 6-31G* charges are used for the solutes? The test has been carried out for 13 diverse organic molecules via Monte Carlo simulations with full representation of the solvent molecules using both EPS and Mulliken charges.

COMPUTATIONAL DETAILS

Molecular Structures and Charges

The 13 molecules are listed in the tables and their structures are illustrated in Figures 1 and 2. Each molecule incorporates a different type of functionality and both first- and second-row atoms are represented. Standard geometries have been adopted based on microwave structures.[11] The methyl groups have been standardized with C–H bond lengths of 1.09 Å and bond angles of 109.47°, and they have been oriented in the lowest-energy conformations.
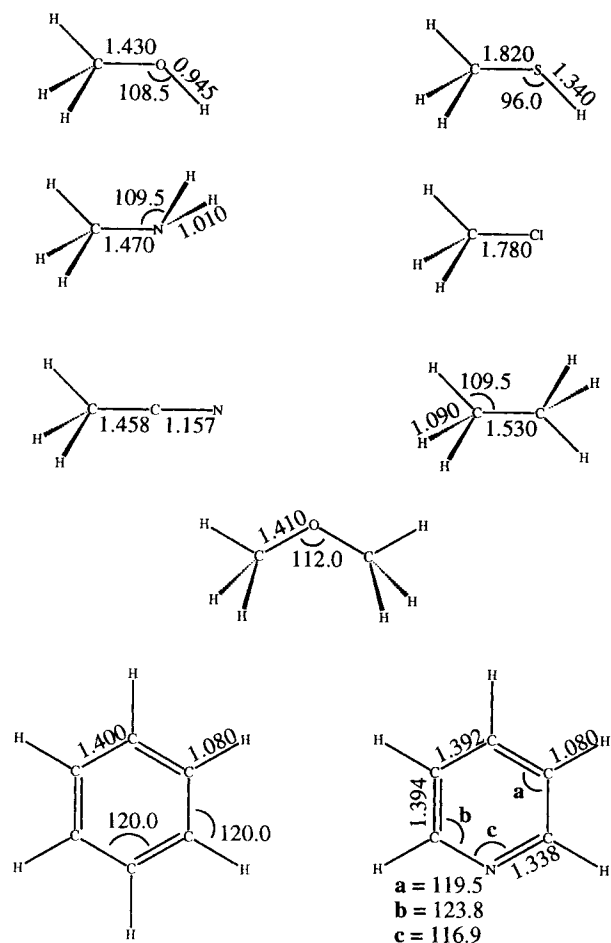

Figure 1. Standard geometries for solutes used in the calculations. All distances in Å and angles in degrees. For methyl groups, r(C—H) = 1.09 Å and $\angle$H—C—X = 109.47°.

Figure 2. Standard geometries for the carbonyl compounds used in the calculations. Other details as in Figure 1.

Monte Carlo Simulations

Relative and absolute free energies of hydration were then computed for the 13 molecules by series of charge and molecular interconversions. The computational procedure is now well established and features Monte Carlo simulations with statistical perturbation theory.[8a,14,15] 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 [eqs. (2) and (3)]. The TIP4P model of water[16] has been

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

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

$$\sigma_{ij} = (\sigma_i \sigma_j)^{1/2} \quad (3)$$

adopted for the present calculations and standard OPLS Lennard–Jones parameters were used for the solute atoms (Table I).[1c,8,15,17]

The free energy changes were computed with the BOSS program.[18] The systems consisted of 260 TIP4P water molecules plus the solute in a cubic cell ca. 20 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. Two types of perturbations were performed, as indicated in Table III. A solute A was gradually converted to a solute B in a series of four or five separate simulations with a coupling parameter, λ, that ranged from 0 (A) to 1 (B). This yielded 8 or 10 free energy increments for each mutation via "double-wide" sampling, i.e., by perturbing from λ_i to $\lambda_i + \Delta\lambda$ and

Single-point ab initio 6-31G* calculations were carried out with Gaussian 92,[12] which generated both Mulliken charges from a standard population analysis and EPS charges from CHELPG calculations.[13] The charges are listed in Table I and the resultant dipole moments are given in Table II. Normally, for fluid simulations the charges for hydrogens that can interconvert by rotations about single bonds are made equivalent to avoid complications if the interconversion occurs during a simulation.[1] Consistently, the charges for hydrogens in CH_3 and NH_2 groups in Table I have been averaged over the three or two hydrogens. The actual variations in charges are typically less than 0.05 e in CH_3 groups,[2,5,7] e.g., for acetamide the 6-31G* EPS charges are 0.114 and 0.104 e for the eclipsed and clinal hydrogens, respectively, while the corresponding Mulliken charges are 0.222 and 0.183 e, respectively. The largest concern for the present molecules would be for the hydrogens on nitrogen syn and anti to the oxygen in acetamide. For the syn H and anti H, the 6-31G* EPS charges are 0.469 and 0.441 e, respectively, and the Mulliken charges are 0.400 and 0.387 e, respectively. Potential errors from this source are considered quantitatively below.
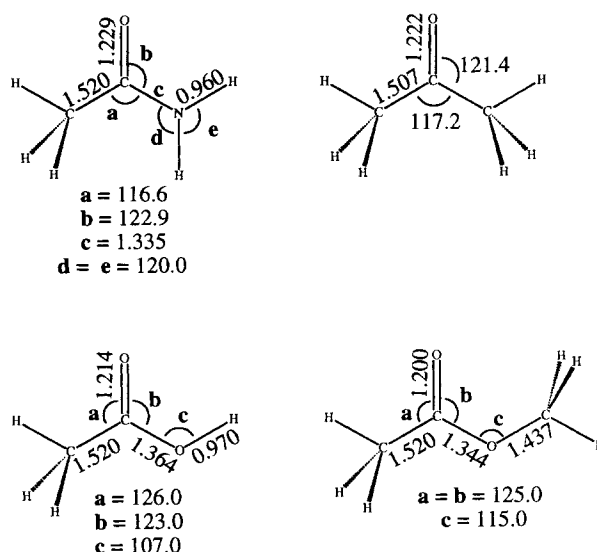
Table I. Computed 6-31G* atomic charges and OPLS Lennard–Jones parameters.

Molecule	Atom	q (Mulliken)	q (EPS)	σ (Å)	ε (kcal/mol)
CH_3OH	C	−0.167	0.301	3.50	0.066
	HC	0.156	−0.010	2.50	0.030
	O	−0.738	−0.697	3.07	0.170
	H	0.437	0.426	0.0	0.0
CH_3NH_2	C	−0.295	0.410	3.50	0.066
	HC	0.156	−0.046	2.50	0.030
	N	−0.837	−1.028	3.25	0.170
	H	0.332	0.378	0.0	0.0
CH_3CN	C	−0.535	−0.209	3.50	0.066
	HC	0.229	0.104	2.50	0.030
	C	0.309	0.400	3.65	0.150
	N	−0.461	−0.503	3.20	0.170
CH_3OCH_3	C	−0.160	0.095	3.50	0.066
	HC	0.154	0.038	2.50	0.030
	O	−0.604	−0.418	3.00	0.170
CH_3SH	C	−0.628	0.013	3.50	0.066
	HC	0.199	0.049	2.50	0.030
	S	−0.058	−0.360	3.55	0.250
	H	0.089	0.200	0.0	0.0
CH_3Cl	C	−0.536	−0.091	3.50	0.066
	HC	0.212	0.098	2.50	0.030
	Cl	−0.100	−0.203	3.40	0.300
C_2H_6	C	−0.471	0.027	3.50	0.066
	H	0.157	−0.009	2.50	0.030
CH_3CONH_2	C	−0.587	−0.404	3.50	0.066
	HC	0.196	0.107	2.50	0.030
	C	0.714	0.964	3.75	0.105
	O	−0.620	−0.680	2.96	0.210
	N	−0.881	−1.113	3.25	0.170
	H	0.393	0.456	0.0	0.0
CH_3COOH	C	−0.563	−0.311	3.50	0.066
	HC	0.208	0.102	2.50	0.030
	C	0.738	0.859	3.75	0.105
	O=	−0.562	−0.623	2.96	0.210
	O—	−0.704	−0.680	3.00	0.170
	H	0.467	0.449	0.0	0.0
$(CH_3)_2CO$	C	−0.571	−0.377	3.50	0.066
	HC	0.192	0.097	2.50	0.030
	CO	0.525	0.773	3.75	0.105
	O	−0.535	−0.601	2.96	0.210
CH_3COOCH_3	C	−0.558	−0.497	3.50	0.066
	HC	0.202	0.134	2.50	0.030
	C	0.762	0.959	3.75	0.105
	O=	−0.567	−0.633	2.96	0.210
	O—	−0.609	−0.508	3.00	0.170
	C	−0.189	0.130	3.50	0.066
	HC	0.185	0.049	2.50	0.030
C_6H_6	C	−0.200	−0.103	3.55	0.070
	H	0.200	0.103	2.42	0.030
C_5H_5N	N	−0.521	−0.679	3.25	0.170
	C1	0.069	0.471	3.55	0.070
	H1	0.202	0.015	2.42	0.030
	C2	−0.256	−0.448	3.55	0.070
	H2	0.210	0.158	2.42	0.030
	C3	−0.144	0.218	3.55	0.070
	H3	0.215	0.069	2.42	0.030

Charges for hydrogens in CH_3 and NH_2 groups are taken as the average of the computed values.

to $\lambda_i - \Delta\lambda$ for the simulation at λ_i.[14] The potential function parameters (q, σ, ϵ) and geometric variables (bond lengths and angles) were all scaled linearly with λ. Bond lengths to disappearing atoms were reduced to 0.1 Å for hydrogens and 0.3 Å for other atoms. The free energy change between λ_i and λ_j,

$G_j - G_i$, is given by eq. (4), where the average involves the total energy difference, $E_j - E_i$, obtained

$$G_j - G_i = -k_B T \ln \langle \exp[-(E_j - E_i)/k_B T]\rangle_i \quad (4)$$

by sampling configurations selected for λ_i.[19] No internal degrees of freedom were sampled for the sol-

Table II. Computed and experimental dipole moments (D).

Molecule	6-31G*	Mulliken	EPS	Exptl.[a]
MeOH	1.96	2.94	2.19	1.70
MeNH$_2$	1.56	1.80	1.79	1.31
MeCN	4.07	4.82	4.06	3.92
MeOMe	1.58	3.19	1.81	1.30
MeSH	1.79	0.91	1.99	1.52
MeCl	2.20	1.96	2.25	1.87
AcNH$_2$	4.32	4.84	4.49	3.76
AcOH	1.82	2.12	1.76	1.74
MeCOMe	3.27	4.23	3.41	2.88
AcOMe	1.95	1.54	1.83	1.72
Pyridine	2.39	2.95	2.37	2.19

Calculated values using the standard geometries in Figures 1 and 2.

[a]Gas-phase values: R.D. Nelson, D.R. Lide, and A.A. Maryott, *Nat. Std. Ref. Data Ser. Nat. Bur. Std. US*, **10**, 1 (1967).

utes or water, so the free energy changes resulted solely from changes in solute–water interactions. 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.[8a,14] The selection of λ values is partially guided by insisting that the statistical uncertainty (1 SD) for a computed free energy increment not exceed 0.1 kcal/mol as obtained from the fluctuations in separate averages over 1×10^5 configurations. The second type of perturbation was simpler and only involved coverting the charge set from EPS to Mulliken for a fixed solute; this could be done typically with high precision in one or two simulations, yielding two or four free energy increments.

Each simulation entailed an equilibration period of 10^6 configurations followed by averaging over an additional 2×10^6 configurations. Attempted moves of the solute were performed every 60 configurations and volume changes were tried every 1625 configurations. The intermolecular interactions were truncated at 8.5 Å for water–water interactions based on

the O—O distance and at 10.0 Å for solute–water interactions based roughly on the distance of the center of mass of the solute 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 selected to give ca. 40% acceptance rates for new configurations.

RESULTS AND DISCUSSION

Relative Free Energies of Hydration

Relative free energies of hydration were computed for three series of solutes using methanol, acetic acid, and benzene as references (Table III). The perturbations between molecules were performed with the EPS charges, and for each molecule the change in free energy for perturbing between the EPS and Mulliken charges was also determined. These quantities can then be combined to give the change in free energies of hydration for interconverting the solutes with the Mulliken charges according to eq. (5). The results are summarized in Table III along

$$A \text{ (Mulliken)} \to A \text{ (EPS)} \to$$
$$B \text{ (EPS)} \to B \text{ (Mulliken)} \qquad (5)$$

with the experimental data.[20–22] The statistical uncertainties for the computed quantities are $\pm 1\sigma$ and were obtained from the fluctuations in the averages over subsets of 10^5 configurations. This procedure may underestimate the true uncertainties; however, the discrepancy is not large. For example, second independent mutations with modified protocols were carried out for methanol to ethane and methanol to acetonitrile with the EPS charges. The two free energy changes were within 0.2 kcal/mol for the former mutation and 0.4 kcal/mol for the latter, while the uncertainties reported in Table III are both 0.2 kcal/mol.

Table III. Computed free energy changes (kcal/mol) in TIP4P water at 25° C.

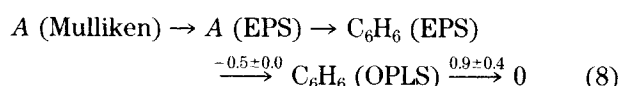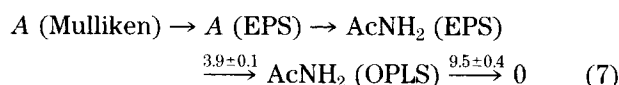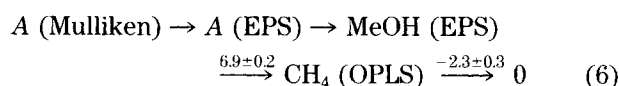|			$\Delta G(B)$	$\Delta G(A \to B)$		
A	$\longrightarrow$	B	EPS → Mulliken	Mulliken	EPS	Exptl.[a]
MeOH		MeOH	−4.6 ± 0.1			
MeOH		MeNH$_2$	0.1 ± 0.1	5.0 ± 0.2	0.3 ± 0.1	0.5
MeOH		MeCN	−2.1 ± 0.1	2.4 ± 0.2	−0.1 ± 0.2	1.2
MeOH		MeOMe	−9.9 ± 0.2	−2.1 ± 0.3	3.2 ± 0.2	3.2
MeOH		MeSH	1.3 ± 0.1	10.5 ± 0.2	4.6 ± 0.1	3.8
MeOH		MeCl	0.4 ± 0.1	9.6 ± 0.2	4.7 ± 0.1	4.6
MeOH		Ethane	−0.3 ± 0.1	12.2 ± 0.2	7.9 ± 0.2	6.9
AcOH		AcOH	−1.8 ± 0.1			
AcOH		AcNH$_2$	−0.1 ± 0.1	−3.1 ± 0.2	−4.9 ± 0.2	−3.0
AcOH		MeCOMe	−2.1 ± 0.1	4.7 ± 0.2	4.9 ± 0.2	2.9
AcOH		AcOMe	−5.4 ± 0.1	−0.4 ± 0.3	3.2 ± 0.2	3.4
Benzene		Benzene	−7.8 ± 0.1			
Benzene		Pyridine	−5.8 ± 0.1	−2.5 ± 0.2	−4.5 ± 0.1	−3.9

[a]Refs. 20–22.

A scan of Table III shows that the EPS results parallel the experimental values without any discrepancies above 2 kcal/mol; the average error for these interconversions is 0.8 kcal/mol. The largest errors are for the series acetamide → acetic acid → acetone. The experimental free energy increment is about 3 kcal/mol for both steps, while the EPS calculations give 4.9 kcal/mol. In comparison to the charges of -0.50 e and $+0.50$ e for the carbonyl O and C of amides in the OPLS potentials,[1c] it may be noted that the polarization for the C=O bond in acetamide (-0.680 and $+0.964$ e) is much greater with the EPS charges in Table I. There is nothing unusual about the pattern of dipole moments for acetamide in Table II, so the dipole moments are not a reliable predictor of success in the free energy calculations. The individual charges on hydrogen-bonding atoms and their covalently bonded neighbors are the more critical issue.

The relative free energies of hydration with the Mulliken charges show much greater deviations from experiment than with the EPS charges; the average error is 3.5 kcal/mol. The ordering of the free energies of hydration within the methanol and acetic acid series is also much deteriorated from the EPS results. The most serious problems are for the interconversions of methanol to methanethiol and methyl chloride. The Mulliken charges on second-row heteroatoms and their attached hydrogens are small in magnitude (Table I) in comparison to the EPS or OPLS charges. The dipole moment for methanethiol is also too small with the Mulliken charges in Table II, though the dipole moment of methyl chloride is reasonable. However, it is not clear if the source of the problem is with these solutes or with methanol, for which the Mulliken charges yield an excessive dipole moment (2.94 D). Particularly poor results are also obtained for the methanol to dimethylether conversion; the Mulliken charge on oxygen (-0.60 e) is too negative compared to the OPLS (-0.50 e)[23] and EPS (-0.42 e) values and contributes to the too large dipole moment (3.19 D) in Table II. Further analyses are facilitated by placing the results on an absolute scale in the next section.

Absolute Free Energies of Hydration

It is highly desirable to obtain the absolute free energies of hydration, ΔG_{hyd}, for transfer of the solutes from the gas phase into aqueous solution because the relative values may mask compensating errors. The absolute quantities are obtained for rigid solutes by computing the free energy change for making the solutes shrink away to nothing in water.[8a,24–27] For large solutes, such calculations are lengthy because the shrinking must be done slowly over many simulations. However, the calculations have been carried out with OPLS potential functions for several key solutes in TIP4P water including methane,[25] ben-

zene,[26] and acetamide.[27] The present problem then reduces to converting the 13 solutes here to one of these reference compounds. In view of the interconversions in Table III, all that is needed is to compute the free energy change for converting the EPS model of methanol to OPLS methane, EPS acetamide to OPLS acetamide, and EPS benzene to OPLS benzene. Then, the free energies for annihilating the EPS or Mulliken models for all the solutes are obtained from the conversions in eqs. (6), (7), or (8).

$$A \text{ (Mulliken)} \rightarrow A \text{ (EPS)} \rightarrow \text{MeOH (EPS)}$$

$$\xrightarrow{6.9\pm0.2} \text{CH}_4 \text{ (OPLS)} \xrightarrow{-2.3\pm0.3} 0 \qquad (6)$$

$$A \text{ (Mulliken)} \rightarrow A \text{ (EPS)} \rightarrow \text{AcNH}_2 \text{ (EPS)}$$

$$\xrightarrow{3.9\pm0.1} \text{AcNH}_2 \text{ (OPLS)} \xrightarrow{9.5\pm0.4} 0 \qquad (7)$$

$$A \text{ (Mulliken)} \rightarrow A \text{ (EPS)} \rightarrow \text{C}_6\text{H}_6 \text{ (EPS)}$$

$$\xrightarrow{-0.5\pm0.0} \text{C}_6\text{H}_6 \text{ (OPLS)} \xrightarrow{0.9\pm0.4} 0 \qquad (8)$$

It may be noted that both the OPLS acetamide and OPLS benzene (12-site) annihilations involved conversion to the OPLS united-atom methane. Consequently, the result for the annihilation of methane is particularly important. This calculation has been repeated; the new result, -2.47 ± 0.36 kcal/mol,[27] is in good accord with the original finding, -2.3 ± 0.3,[25] so the original value still stands. It corresponds to a ΔG_{hyd} of $+2.3 \pm 0.3$ kcal/mol, which compares well with the experimental result of 2.005 kcal/mol.[21] The computed ΔG_{hyd} values for OPLS acetamide (-9.5 ± 0.4 kcal/mol) and benzene (-0.9 ± 0.4 kcal/mol) are also in accord with the experimental results of -9.7[22] and -0.8 kcal/mol.[20,21]

The conversion of the EPS model of methanol to OPLS methane involved collapsing the hydroxyl group and methyl hydrogens onto the carbon with removal of the charges and conversion of the Lennard–Jones parameters. This was carried with good precision in 6 Monte Carlo simulations, yielding 12 free energy increments. All details of the simulations were the same as described above for the other mutations. The total free energy change was 6.90 ± 0.20 kcal/mol. The mutation from EPS to OPLS acetamide[28] was simpler; it only required contraction of the methyl hydrogens, charge adjustments, and conversion of the Lennard–Jones parameters for the methyl group to the united-atom form. This was carried out in four simulations, yielding eight free energy increments and a free energy change of 3.85 ± 0.10 kcal/mol. The final connective calculation for EPS to OPLS all-atom benzene required only a trivial charge change; one simulation sufficed (two free energy increments), yielding a free energy change of -0.48 ± 0.03 kcal/mol.

These results could then be combined via eqs. (6)–(8) with those in Table III to give the absolute free energies of hydration shown in Table IV. The merits of the 6-31G* Mulliken and EPS approaches now

Table IV. Absolute free energies of hydration (kcal/mol) at 25 C[a]

Molecule	ΔG Mulliken	ΔG EPS	ΔG Exptl.
MeOH	-9.2 ± 0.5	-4.6 ± 0.4	-5.1^a
MeNH$_2$	-4.2 ± 0.5	-4.3 ± 0.5	-4.6^b
MeCN	-6.8 ± 0.5	-4.7 ± 0.5	-3.9^a
MeOMe	-11.3 ± 0.5	-1.4 ± 0.5	-1.9^a
MeSH	1.3 ± 0.5	0.0 ± 0.5	-1.2^a
MeCl	0.5 ± 0.5	0.1 ± 0.5	-0.5^a
Ethane	3.0 ± 0.5	3.3 ± 0.5	1.8^a
AcNH$_2$	-13.5 ± 0.4	-13.4 ± 0.4	-9.7^c
AcOH	-10.3 ± 0.4	-8.5 ± 0.4	-6.7^a
MeCOMe	-5.7 ± 0.5	-3.6 ± 0.5	-3.8^a
AcOMe	-10.7 ± 0.5	-5.3 ± 0.5	-3.3^a
Benzene	-8.2 ± 0.4	-0.4 ± 0.4	$-0.8^{a,b}$
Pyridine	-10.7 ± 0.4	-4.9 ± 0.4	-4.7^a

For transfer of the solute from the ideal gas to aqueous solution with 1 M standard states.
[a]Ref. 20.
[b]Ref. 21.
[c]Ref. 22.

clearly emerge. The Mulliken charges yield an average error of 4.0 kcal/mol for ΔG_{hyd}. In fact, for the methanol series the serious problems are with methanol itself and also dimethyl ether. For the carbonyl compounds, the Mulliken results are again too exoergic, which correlates with the presence of oxygen atoms. And, for the two aromatic molecules the Mulliken charges on the hydrogens are much larger ($+0.2$ e) than with the OPLS and EPS alternatives (ca. $+0.1$ e) and again cause ΔG_{hyd} to be far too exoergic.

The computed absolute free energies of hydration with the EPS charges are generally impressive with an average error of 1.1 kcal/mol. The results are also displayed in Figure 3. A good linear relationship with
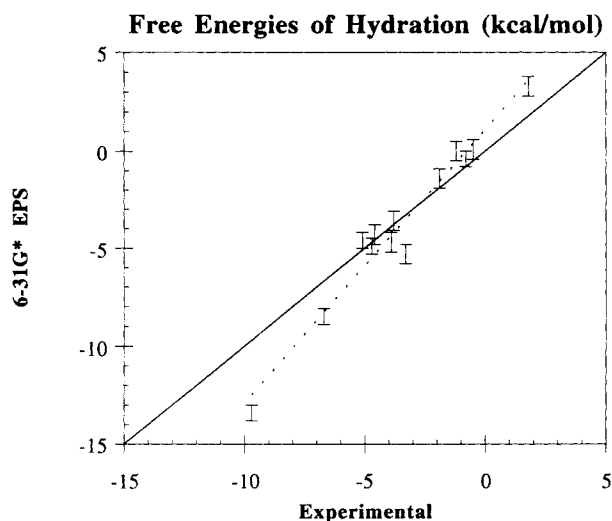
Free Energies of Hydration (kcal/mol)

Figure 3. Correlation of absolute free energies (kcal/mol) of hydration from experimental measurements and Monte Carlo simulations with the 6-31G* EPS charges. The solid line has unit slope, while the broken line shows the best linear fit.

a correlation coefficient, r, of 0.98 exists between the experimental and EPS results, $\Delta G(EPS) = 1.40 \Delta G(exptl) + 1.11$. In view of the statistical uncertainties in the computed values, which are probably in reality between the values of 0.5 kcal/mol in Table IV and 1 kcal/mol, it is difficult to do much better. Unfortunately, the largest individual error (3.7 kcal/mol) is for acetamide; it is too hydrophilic, while another important biochemical model compound, ethane, is too hydrophobic by 1.5 kcal/mol. These findings are a potentially serious problem for intended use of the 6-31G* EPS approach for simulations of peptides and proteins. It is advisable to consider hybrid models in which OPLS charges are used for amide and alkyl units because these do give the correct free energies of hydration.

Connection between the Charge Distributions and ΔG_{hyd}

The present results provide insights into the connection between point-charge distributions and errors in ΔG_{hyd}. Dipole moments alone are not a useful indicator of potential success or failure. For example, with the Mulliken charges for the carbonyl compounds the smallest (methyl acetate) and largest (acetone) errors in dipole moments correspond to the largest and smallest errors in ΔG_{hyd} (Tables II and IV).

Looking at individual charges at hydrogen-bonding sites is also not a good predictor. For example, the charges on oxygen and hydroxyl hydrogen in methanol are similar in the Mulliken (-0.74, $+0.44$ e), EPS (-0.70, 0.43 e), and OPLS (-0.700, 0.435 e) models; however, the Mulliken charges lead to a 4-kcal/mol error in ΔG_{hyd}. The problem must lie in the charge distribution for the methyl group. The Mulliken approach places too much positive charge on the hydrogens, so that the carbon still ends up -0.17 e, while it is $+0.30$ with the EPS charges. Consequently, a hydrogen in a hydrogen bond at the oxygen also experiences strong additional attraction with the proximal methyl carbon using the Mulliken charges but strong repulsion with the EPS carbon. In the OPLS united-atom model, the methyl group has a charge of $+0.265$, which also reduces the hydrogen bonding at oxygen. If the Mulliken charges for the methyl group were combined into an united-atom form, a reasonable model would result. In fact, a better correlation exists between 6-31G* Mulliken charges ($r = 0.99$) than 6-31G* EPS charges ($r = 0.93$) in a united-atom format with OPLS united-atom charges for first-row atoms, though the Mulliken charges for second-row atoms are not useful.[10]

The gross error (9.4 kcal/mol) for ΔG_{hyd} of dimethyl ether with the Mulliken charges has a similar origin. The oxygen is more negative with the Mulliken charges (-0.60 e) than with the EPS (-0.42 e) or OPLS (-0.50 e) models; however, the problem is

aggravated by the two methyl groups in which the carbons are -0.16 e (Mulliken) v. $+0.10$ (EPS) and $+0.25$ e for the united-atom methyl group in the OPLS model.

The error in the EPS ΔG_{hyd} for acetamide also warrants further analysis. The Mulliken and EPS charge distributions in this case are fairly similar (Table I); the methyl hydrogens are still more positive with the Mulliken model, though the methyl group is no longer adjacent to a hydrogen-bonding site, i.e., in a position to do much damage. Consequently, the two computed ΔG_{hyd} values in Table IV are close but too exoergic by nearly 4 kcal/mol. For comparison, the OPLS charges for acetamide, which give the correct ΔG_{hyd},[27] are CH_3 (0.0), C($+0.50$ e), O(-0.50 e), N(-0.85 e), and H_N($+0.425$ e).[28] The dipole moment with the OPLS model is 4.3 D, marginally smaller than the EPS value of 4.5 D (Table II). The difference is again in the details of the charge distributions; in particular, the polarization of the

N—C=O fragment with the EPS charges (-1.11, $+0.96$, -0.68 e) is too great and makes hydrogen bonds to the oxygen too strong. This is evident through optimizations for a single TIP4P water molecule with the EPS and OPLS models of acetamide. The results in Figure 4 were obtained for hydrogen bonding of water to the carbonyl oxygen both syn and anti to the nitrogen. The syn structures are the lowest-energy forms and show stronger attraction with the EPS charges by 0.9 kcal/mol. The difference increases to 2.1 kcal/mol for the anti forms. Six intermolecular degrees of freedom were optimized for each of these structures except for the anti form with the OPLS model, which had to be constrained to coplanarity to avoid conversion to the syn structure. Because primary amides form three to four hydrogen bonds with water in aqueous solution,[28] the accumulation of enhanced attractions with the EPS charges can account for the 3.7-kcal/mol overly exoergic ΔG_{hyd}.
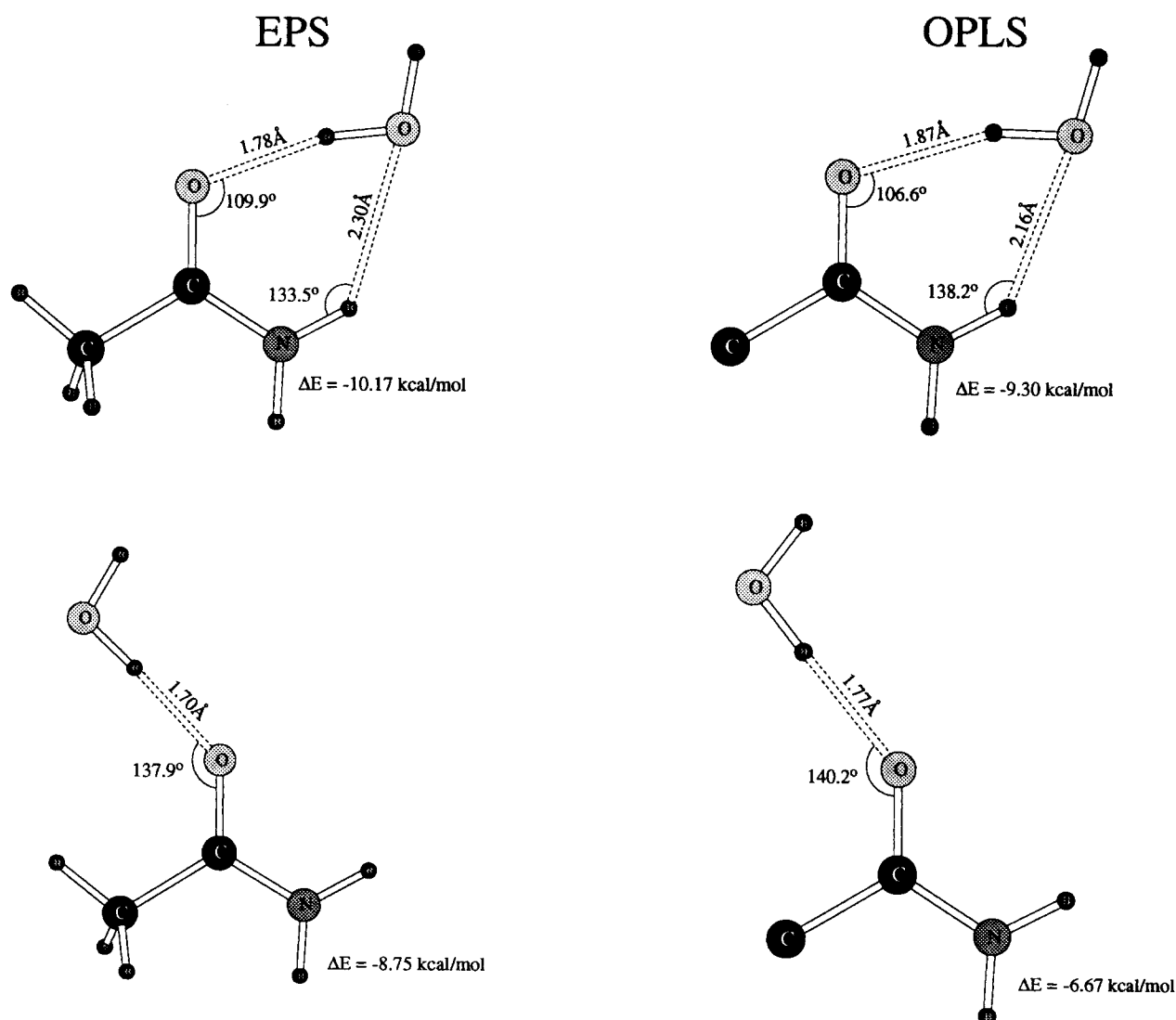

Figure 4. Results of gas-phase optimizations for acetamide–water complexes. Results with the 6-31G* EPS charges for acetamide are given on the left and with the OPLS charges on the right. The water model is TIP4P.

Two additional points can be considered concerning the error in the ΔG_{hyd} for acetamide. Perhaps a change in water model could improve the results. Similar free energy calculations have been performed for methanol, methylamine, acetic acid, and acetone with the TIP3P model of water,[16] frequently used in molecular dynamics simulations of biochemical systems.[1] As described in detail elsewhere, the results for ΔG_{hyd} are not improved by this change.[29] In addition, the optimizations in Figure 4 were repeated with TIP3P water. The computed interaction energies for the syn and anti forms are -10.4 and -8.33 kcal/mol, respectively, with the EPS charges, and -9.34 and -6.40 kcal/mol, respectively, with OPLS acetamide. Considering the similarity to the results in Figure 4, it is doubtful that the error in ΔG_{hyd} for acetamide with the EPS charges would diminish significantly by switching to the TIP3P water model. The second point is the potential error mentioned above from the averaging of the EPS charges for hydrogens in CH_3 and NH_2 groups. This was tested by performing Monte Carlo simulations to perturb from the averaged charges to the original unaveraged values for acetamide. Just perturbing the NH_2 group yielded a free energy change of $+0.09 \pm 0.01$ kcal/mol, while perturbation of the CH_3 and NH_2 groups simultaneously gave a free energy change of $+0.25 \pm 0.01$ kcal/mol. This means that the unaveraged charge distribution for acetamide is less well hydrated than for the averaged distribution. Thus, if the charge averaging was not done for acetamide the computed ΔG_{hyd} with the EPS charges would be improved slightly to -13.1 kcal/mol.

Solute–Water Hydrogen Bonding

The present results provide an opportunity to characterize the hydrogen bonding between the solutes and water. Hydrogen bonding analyses were performed on 100 configurations saved at an interval of 20,000 configurations during the Monte Carlo simulations. For this purpose, a hydrogen bond is defined by a solute–water interaction that is attractive by at least 2.25 kcal/mol and by an $N\cdots H$ or $O\cdots H$ distance of no more than 2.5 Å. The distance restriction is supported by the limits of the first peaks in radial distribution functions for hydrogens in hydrogen bonds with N and O atoms;[8a] the first peaks for hydrogen bonding with chlorides and thiols are less distinct, so only the energetic criterion was enforced for them. The energetic cutoff is an empirical one that has been used previously, e.g., for water and amides, and is supported by the often bimodal nature of energy pair distributions for hydrogen-bonded systems.[16,30] The average numbers of hydrogen bonds with water are given in Table V and have uncertainties of ca. ± 0.1 hydrogen bond. Each hydrogen bond has also been assigned to an atom of

Table V. Average numbers of hydrogen bonds with water

Molecule	Atom	No. hydrogen bonds	
		Mulliken	EPS
MeOH	O	2.01	1.71
	H	0.71	1.05
MeNH$_2$	N	0.94	1.02
	H[a]	0.29	0.78
MeCN	N	1.62	1.42
MeOMe	O	2.22	0.99
MeSH	S	0.00	0.24
	H	0.01	0.27
MeCl	Cl	0.02	0.02
AcNH$_2$	O	2.21	2.41
	H syn	0.23	0.42
	H anti	0.65	0.80
AcOH	O=	1.49	1.54
	O—	1.02	0.63
	H	1.03	0.97
MeCOMe	O	2.00	1.64
AcOMe	O=	1.78	1.38
	O—	1.80	0.57
Pyridine	N	1.30	1.16

A hydrogen bond is defined by an interaction energy ≤ -2.25 kcal/mol and for O and N an $O\cdots H$ and $N\cdots H$ distance ≤ 2.5 Å.

[a] Total for both hydrogens.

the solute based on the shortest distance to the water molecule.

In comparison to prior results for many of the present functional groups with OPLS potentials, the results in Table V with the EPS charges are reasonable. Alcohols including methanol have been found to participate in two to three hydrogen bonds with water, always one as a donor and one to two as an acceptor.[8a,31] A previous study with TIP4P water also found 2 hydrogen bonds for methylamine, 1 as donor and 1 as acceptor, and 1.3–1.4 for acetonitrile.[32] Similarly, 1.1 water molecules were computed to be hydrogen bonded to the nitrogen in benzonitrile.[8a] For ethers, one hydrogen bonds to water is the rule,[8a,33,34] and is reflected again here in the EPS results. The maximal attractions are less for methanethiol and methyl chloride with water, i.e., -3.9 and -3.0 kcal/mol, respectively, with the EPS charges, and -2.9 and -3.2 kcal/mol, respectively, with the Mulliken charges. Consequently, these solutes do not compete effectively against the stronger water–water interactions and virtually no solute–water hydrogen bonding occurs.

For carbonyl compounds, two hydrogen bonds to the carbonyl oxygen are typical.[28,34–37] The present results with the EPS charges are somewhat less than this except for acetamide. The overall number of hydrogen bonds, 3.6, for acetamide agrees with the previous finding for formamide; the only difference is a shift in assignment of 0.4 hydrogen bonds from the syn H to the carbonyl oxygen.[28] The one hydrogen bond to the acidic hydrogen in acetic acid and the limited hydrogen bonding to the alkoxy oxygen is acetic acid and methyl acetate are also reasonable

based on previous studies of the interactions of these solutes with water.[38] Finally, the roughly one hydrogen bond for pyridine is consistent with the results for methylamine and acetonitrile.

The results with the Mulliken charges in Table V do reflect some of the problems from the free energies of hydration. The principal error is too much hydrogen bonding to singly bonded oxygens. Thus, dimethyl ether is predicted to have 2.2 hydrogen bonds rather than 1, and the alkoxy oxygens in acetic acid and methyl acetate have 1–2 hydrogen bonds v. 0.6 with the EPS charges.

CONCLUSION

An important test of the viability of partial charges from 6-31G* wave functions for use in intermolecular potential functions has been made. In an all-atom format, charges obtained from fitting to the electrostatic potential surface reproduce experimental free energies of hydration for organic molecules with an average error of 1.1 kcal/mol. Further, the EPS charges were found to yield reasonable predictions for the number of hydrogen bonds between the organic solutes and water. It may also be noted that results from a self-consistent reaction field approach with 6-31G* wave functions perform well for predicting free energies of hydration.[29] The results with 6-31G* Mulliken charges are much more erratic. A principal problem was traced to the charge distribution for methyl groups attached to hydrogen-bonding functionality, which could be largely relieved in a united-atom format. Aromatic C—H bonds are also found to be too polarized with the Mulliken charges. Though the results with the EPS charge are impressive, restraint is warranted for biochemical applications in view of the individual errors of 3–4 kcal/mol for acetamide and 1–2 kcal/mol for ethane. A combination of, for example, OPLS and 6-31G* EPS charges might be a viable alternative for rapid derivation of partial charges for new applications. Scaling of the 6-31G* EPS charges could also be investigated, though uniform scaling is unlikely to be fruitful in view of the good accord that already exists between many of the predicted free energies of hydration and the experimental values. In any event, further work in this area needs to stress the performance of proposed charge models for key observables in solution rather than correlations between alternative models.

References

1. See, for example: (a) S.J. Weiner, P.A. Kollman, 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) W.L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
2. B.H. Besler, K.M. Merz Jr., and P.A. Kollman, *J. Comp. Chem.*, **11**, 431 (1990).
3. M. Orozco and F.J. Luque, *J. Comp. Chem.*, **11**, 909 (1990).
4. P. Cieplak and P.A. Kollman, *J. Comp. Chem.*, **12**, 1232 (1991).
5. K.M. Merz Jr., *J. Comp. Chem.*, **13**, 749 (1992).
6. G. Alagona and C. Ghio, *J. Mol. Struct. (THEOCHEM)*, **256**, 187 (1992).
7. K.B. Wiberg and P.R. Rablen, *J. Comp. Chem.*, to appear.
8. For recent examples, see: (a) W.L. Jorgensen and T.B. Nguyen, *J. Comp. Chem.*, **14**, 195 (1993); (b) W.L. Jorgensen, E.R. Laird, T.B. Nguyen and J. Tirado-Rives, *J. Comp. Chem.*, **14**, 206 (1993).
9. F.J. Luque, M. Orozco, F. Illas, and J. Rubio, *J. Am. Chem. Soc.*, **113**, 5203 (1991).
10. T.B. Nguyen, PhD thesis, Yale University, New Haven, CT, 1993.
11. 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).
12. M.J. Frisch, G.W. Trucks, M. Head-Gordon, P.M.W. Gill, M.W. Wong, J.B. Foresman, B.G. Johnson, H.B. Schlegel, M.A. Robb, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, J.S. Binkley, C. Gonzalez, R.L. Martin, D.J. Fox, D.J. Defrees, J. Baker, J.J.P. Stewart, and J.A. Pople, Gaussian 92, revision A, Gaussian, Inc., Pittsburgh, PA, 1992.
13. C.M. Breneman and K.B. Wiberg, *J. Comp. Chem.*, **11**, 361 (1990).
14. W.L. Jorgensen and C. Ravimohan, *J. Chem. Phys.*, **83**, 3050 (1985).
15. W.L. Jorgensen, J.M. Briggs, and L. Contreras, *J. Phys. Chem.*, **94**, 1683 (1990).
16. W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
17. OPLS all-atom hydrocarbon parameters: E.M. Duffy, G. Kaminskii, T. Matsui, and W.L. Jorgensen, to be published.
18. W.L. Jorgensen, BOSS, v. 3.2, Yale University, New Haven, CT, 1992.
19. R.W. Zwanzig, *J. Chem. Phys.*, **22**, 1420 (1954).
20. J. Hine and P.K. Mookerjee, *J. Org. Chem.*, **40**, 292 (1975).
21. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81**, 2016 (1984).
22. R. Wolfenden, Y.-L. Liang, M. Matthews, and R. Williams, *J. Am. Chem. Soc.*, **109**, 463 (1987).
23. J.M. Briggs, T. Matsui, and W.L. Jorgensen, *J. Comp. Chem.*, **11**, 958 (1991).
24. T.P. Straatsma, H.J.C. Berendsen, and J.P.M. Postma, *J. Chem. Phys.*, **85**, 6720 (1986).
25. W.L. Jorgensen, J.F. Blake, and J.K. Buckner, *Chem. Phys.*, **129**, 193 (1989).
26. W.L. Jorgensen and D.L. Severance, *J. Am. Chem. Soc.*, **112**, 4768 (1990).
27. E.M. Duffy, D.L. Severance, and W.L. Jorgensen, *Isr. J. Chem.*, to appear.
28. W.L. Jorgensen and C.J. Swenson, *J. Am. Chem. Soc.*, **107**, 569 (1985).
29. M. Orozco, W.L. Jorgensen, and F.J. Luque, *J. Comp. Chem.*, to appear.
30. E.M. Duffy, D.L. Severance, and W.L. Jorgensen, *J. Am. Chem. Soc.*, **114**, 7535 (1992).
31. W.L. Jorgensen and J.D. Madura, *J. Am. Chem. Soc.*, **105**, 1407 (1983).
32. W.J. Dunn III and P.I. Nagy, *J. Phys. Chem.*, **94**, 2099 (1990).

33. W.J. Dunn III and P.I. Nagy, *J. Comp. Chem.*, **13**, 468 (1992).

34. D.L. Severance and W.L. Jorgensen, *J. Am. Chem. Soc.*, **114**, 10966 (1992).

35. P.J. Rossky and M. Karplus, *J. Am. Chem. Soc.*, **101**, 1913 (1979).

36. P.K. Mehrotra and D.L. Beveridge, *J. Am. Chem. Soc.*, **101**, 4287 (1980).

37. J.F. Blake and W.L. Jorgensen, *J. Am. Chem. Soc.*, **113**, 7430 (1991).

38. J.M. Briggs, T.B. Nguyen, and W.L. Jorgensen, *J. Phys. Chem.*, **95**, 3315 (1991).