J. Phys. Chem. A **1999,** *103,* 1853−1856

Prediction of Aqueous Solvation Free Energies from Properties of Solute Molecular Surface Electrostatic Potentials

Jane S. Murray, Fakhr Abu-Awwad, and Peter Politzer*

Department of Chemistry, University of New Orleans, New Orleans, Louisiana 70148

Received: October 28, 1998; In Final Form: January 4, 1999

It is shown that aqueous solvation free energies, ΔG_{sol}, can be expressed quantitatively in terms of properties of the molecular surface electrostatic potentials of the solutes. The latter are obtained computationally by the B3P86/6-31+G** density functional procedure. Regression analyses and an experimental database encompassing 50 solutes of various types are used to obtain an analytical representation of ΔG_{sol} which reproduces the experimental values with a standard deviation of 1.57 kJ/mol and an average absolute deviation of 1.14 kJ/mol.

Introduction

Considerable progress has been made in modeling and quantifying the interactions between a medium and solute molecules, and in predicting free energies of solvation.[1−10] One general approach is to account explicitly for the solute and implicitly for the solvent, treating the latter as a dielectric continuum; a disadvantage is that this may not adequately reflect specific features such as solute−solvent hydrogen bonding. An alternative is to include the solvent molecules explicitly; however, the resulting need to appropriately deal with the very large number of interactions required for a reasonable representation of the system imposes major demands upon computer resources. The Langevin dipole model, in which the solvent is described as a fixed cubic lattice of point polarizabilities,[10] is another option, intermediate between fully explicit and implicit treatments of the solvent. Very recently, a screening concept that has been effective in this lattice model was implemented as a “dielectric screening” approximation within the continuum description.[8] In all of this methodology, a primary objective is to adequately account for electrostatic interactions.[1−10]

In this paper, we demonstrate that aqueous solvation energies can be predicted using only computed molecular properties of the solute. Our procedure is to calculate the electrostatic potentials $V(\mathbf{r})$ on the molecular surfaces of the solutes and to use statistical quantities defined in terms of these potentials, together with an experimentally-determined data base, to develop an analytical formulation for aqueous solvation energies. This type of approach has proven to be very fruitful in the representation and prediction of a variety of condensed phase properties,[11−13] including normal boiling points and critical constants,[14] heats of vaporization[11] and sublimation,[15] solubilities in supercritical solvents,[16,17] diffusion constants,[18] C_{60} solubilities,[19] partition coefficients,[20,21] surface tensions,[22] liquid and solid densities,[22] heats of fusion,[22] and lattice energies.[23]

Methods

We have used the density functional B3P86/6-31+G** option of Gaussian 94[24] to optimize the molecular geometries of 50 solutes and to compute the electrostatic potential $V(\mathbf{r})$ on the molecular surfaces defined by the 0.001 au contour of the electronic density $\rho(\mathbf{r})$.[25]

The electrostatic potential $V(\mathbf{r})$ created in the space around a molecule by its nuclei and electrons is well established as a guide to the interactive behavior of molecules.[26−30] It is defined by eq 1, in which the molecule is treated as a collection of stationary positive point charges, the nuclei, surrounded by a continuous but static distribution of electrons.

$$V(\mathbf{r}) = \sum_{A} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')\,d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \qquad (1)$$

Z_A is the charge on nucleus A, located at $\mathbf{R}_A$. The first term on the right side of eq 1 is the nuclear contribution to $V(\mathbf{r})$ and is positive; the second term is due to the electrons and is accordingly negative.

The electrostatic potential has been used extensively as a means of identifying molecular regions most susceptible to electrophilic and, more recently, nucleophilic attack and for inferring general patterns of positive and negative potential that may promote or inhibit particular molecular interactions, such as those that occur between drugs and receptors.[26−30] In the past, the quantitative analysis of $V(\mathbf{r})$ often focused upon locating and ranking the most negative potentials, V_{min}, in the space surrounding a molecule; the more negative V_{min} were viewed as indicating the sites initially most reactive toward electrophiles. These V_{min} are usually associated with (a) the lone pairs of the more electronegative atoms, such as N, O, F, Cl, S, and Br, and (b) unsaturated, aromatic and strained carbon−carbon bonds.

In recent years, the electrostatic potential computed on the molecular surface, $V_S(\mathbf{r})$, has been shown to provide an effective basis for correlating and predicting properties that reflect molecular interactions.[11−13] The GIPF approach (general interaction properties function) involves characterizing $V_S(\mathbf{r})$ in terms of a group of both site-specific and global statistical quantities; the former include the surface maxima and minima ($V_{S,max}$ and $V_{S,min}$), and the latter are the average deviation and the positive and negative variances of $V_S(\mathbf{r})$ over the entire surface. As mentioned earlier, we have used various subsets of these quantities to obtain multivariable analytical expressions for a number of macroscopic condensed phase properties.[11−23] In the present work, we have been able to extend this approach to the

10.1021/jp984271w CCC: $18.00

TABLE 1: Experimental[a] and Calculated Aqueous Solvation Free Energies, and Solute Molecular Surface Properties[b]

molecule	ΔG_{sol}(exp)[a] (kJ/mol)	ΔG_{sol}(calc) (kJ/mol)	negatively-charged surface area A_S^-	total surface area A_S	$\bar{V}_S^-$	$\bar{V}_S^+$	$V_{S,max}$ (kcal/mol)	$V_{S,min}$ (kcal/mol)	V_{min}[c] (kcal/mol)
N-methylformamide	−41.80	−38.41	34.5	100.2	−23.68	15.20	48.5	−43.8	−59.2
acetamide	−40.63	−41.40	37.3	100.6	−23.60	16.62	49.0	−44.1	−59.3
propionamide	−39.38	−40.08	47.9	120.6	−18.29	14.05	48.2	−43.6	−57.7
1,2-ethanediol	−32.02	−31.10	40.2	101.9	−17.20	14.03	51.0	−37.8	−55.2
1,2-ethanediamine	−31.77	−32.44	35.8	112.5	−17.82	12.70	28.0	−37.6	−139.0[c]
piperazine	−30.85	−29.91	38.5	132.1	−14.46	8.34	28.5	−37.8	−134.2[c]
acetic acid	−28.01	−26.91	37.3	95.6	−17.45	14.15	52.2	−34.3	−46.0
phenol	−27.67	−26.45	62.4	134.7	−11.60	11.52	49.9	−32.3	−44.3
propionic acid	−27.04	−26.38	40.2	115.8	−16.14	11.07	51.5	−34.2	−46.0
butyric acid	−26.58	−27.58	37.9	137.3	−17.63	9.80	51.2	−35.2	−47.3
p-cresol	−25.62	−27.87	63.2	154.9	−11.89	9.56	49.7	−33.0	−45.6
2-methylpyrazine	−23.07	−22.68	48.8	136.4	−12.68	9.10	19.6	−32.5	−106.4[c]
aniline	−22.97	−22.80	67.3	138.8	−12.12	12.10	38.1	−27.8	−45.5
pyrrolidine	−22.91	−18.30	5.9	121.7	−14.39	6.55	25.7	−39.6	−72.3
cyclohexanol	−22.86	−20.13	31.4	149.8	−15.43	6.43	41.2	−36.8	−53.7
piperidine	−21.36	−20.02	34.9	137.2	−8.12	4.68	26.0	−35.5	−68.0
methanol	−21.23	−22.04	24.7	71.1	−19.22	12.65	45.2	−36.7	−53.7
1,4-dioxane	−21.13	−19.28	38.5	125.3	−16.07	8.98	15.3	−30.6	−91.6
ethanol	−20.48	−21.09	24.8	93.1	−18.93	9.86	43.2	−36.9	−53.9
1-propanol	−20.19	−21.17	31.8	114.0	−15.12	7.66	43.9	−36.4	−53.2
2-propanol	−19.90	−19.86	25.0	112.3	−18.42	8.18	40.9	−36.9	−53.8
1-butanol	−19.73	−20.82	26.1	134.3	−18.27	7.64	42.6	−36.9	−53.8
pyridine	−19.65	−16.77	52.0	119.9	−11.66	10.85	21.8	−37.9	−61.8
methylamine	−19.10	−19.14	19.8	76.6	−18.56	8.87	26.5	−40.1	−74.0
ethylamine	−18.81	−18.61	30.8	98.4	−12.02	7.81	26.4	−39.3	−73.0
butylamine	−18.31	−18.25	39.2	139.6	−9.80	6.20	26.1	−39.3	−72.9
1-hexanol	−18.22	−20.77	33.7	175.9	−14.30	6.80	42.4	−36.9	−53.8
ammonia	−18.02	−20.40	18.7	52.2	−22.34	15.51	28.1	−42.5	−76.1
dimethylamine	−17.93	−19.29	20.2	99.0	−15.30	6.02	26.6	−36.8	−69.3
anthracene	−17.68	−18.01	108.2	217.0	−9.21	9.38	16.7	−15.0	−15.0
pentaneamine	−17.14	−18.33	33.6	160.1	−11.41	6.32	25.3	−40.2	−73.1
diethylamine	−17.01	−18.18	39.6	140.8	−7.76	5.01	24.5	−35.4	−68.5
acetone	−16.09	−16.93	29.8	107.3	−22.87	11.93	21.2	−37.8	−51.3
2-butanone	−15.22	−16.36	35.8	126.9	−18.74	9.95	20.2	−37.8	−51.5
acetaldehyde	−14.63	−14.05	27.5	86.4	−22.28	14.03	22.5	−35.3	−47.8
propionaldehyde	−14.38	−13.48	30.3	106.8	−20.15	10.98	20.1	−35.5	−48.1
3-methyl-2-butanone	−13.54	−16.57	37.0	143.9	−18.16	8.88	20.4	−38.0	−51.9
butanal	−13.29	−12.69	37.4	125.3	−15.63	9.65	20.0	−35.1	−47.9
2,4-dimethyl-3-pentanone	−11.45	−13.71	46.4	178.0	−13.51	6.90	16.5	−37.0	−49.9
naphthalene	−9.99	−8.87	82.0	172.0	−9.29	8.93	15.7	−15.4	−16.0
dimethylsulfide	−6.44	−5.66	33.3	107.8	−13.50	8.85	15.1	−24.8	−34.3
methylethyl sulfide	−6.23	−5.88	36.7	129.0	−12.85	7.28	14.7	−25.1	−34.9
diethyl sulfide	−5.98	−5.97	40.5	149.5	−12.20	6.21	12.6	−25.3	−35.5
ethane thiol	−5.43	−5.58	33.0	108.1	−12.61	9.19	19.1	−23.4	−31.6
methane thiol	−5.18	−5.70	32.1	87.6	−12.80	11.08	19.8	−13.3	−31.6
o-xylene	−3.76	−3.60	57.6	161.6	−9.55	6.46	12.7	−19.5	−20.1
benzene	−3.64	−2.17	55.1	124.2	−8.98	8.34	14.4	−16.7	−16.8
ethylbenzene	−3.34	−3.00	61.2	164.4	−8.65	6.37	13.6	−18.3	−18.5
toluene	−3.18	−2.89	56.3	144.8	−9.27	7.03	13.5	−18.4	−18.7
butylbenzene	−1.67	−2.91	77.1	202.2	−6.80	5.76	13.6	−18.3	−18.3

[a] Reference 8. [b] Computed at the B3P86/6-31+G* level using Gaussian 94. [c] For 1,2-ethanediamine, piperazine, 2-methylpyrazine, and 1,4-dioxane, the number given as V_{min} is actually the sum of the two most negative electrostatic potential minima. See discussion in text.

representation of aqueous solvation energies. We develop our correlations through regression analyses with the SAS statistical code.[31]

Results

In Table 1 are listed the experimentally-determined aqueous solvation energies, ΔG_{sol}(exp), for 50 solute molecules, of a wide variety of chemical types. The solvation energies range from −41.80 kJ/mol for *N*-methylformamide to −1.67 kJ/mol for butylbenzene. Also presented in Table 1 are some key features of the molecular surface electrostatic potentials, which were found to be relevant to aqueous solvation energies. These include the following:

(a) A_S^-, the surface area over which $V_S(\mathbf{r})$ is negative, and for comparison, A_S, the total area;

(b) $\bar{V}_S^-$, the average negative potential on the surface, defined by $\bar{V}_S = (1/n)\sum_{i=1}^{n} V_S^-(\mathbf{r}_i)$, where $V_S^-(\mathbf{r}_i)$ represents the negative values of $V_S(\mathbf{r})$;

(c) for comparison, the analogously defined average positive potential on the surface, $\bar{V}_S^+$;

(d) $V_{S,max}$ and $V_{S,min}$, the most positive and most negative values of $V_S(\mathbf{r})$; and

(e) V_{min}, the most negative $V(\mathbf{r})$ anywhere in the three-dimensional space around the molecule. For four of the molecules in Table 1, it was found that V_{min} must be taken to be the sum of two contributions. This shall be discussed later in this paper.

The data in Table 1 show that negative potentials generally cover 25−40% of the total molecular surface; only for naphthalene and anthracene are the positive and negative areas

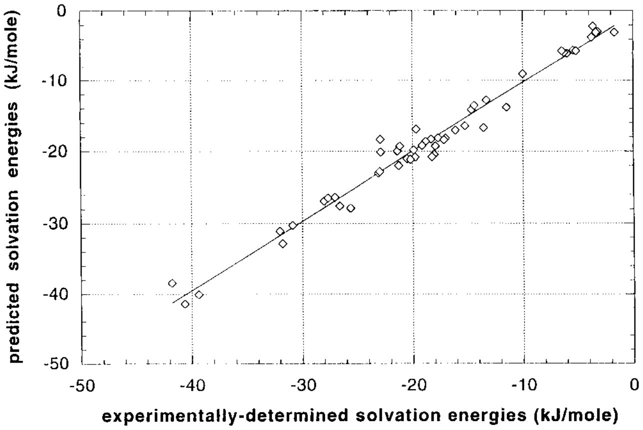

Figure 1. Plot of predicted versus experimentally-determined aqueous solvation energies. The linear correlation coefficient is 0.988, the standard deviation is 1.57 kJ/mol, and the average absolute deviation is 1.14 kJ/mol.

approximately the same. On the other hand, the negative potential is typically stronger than the positive; i.e. usually $|\bar{V}_S^-| > \bar{V}_S^+$. These general tendencies are in accord with earlier observations.[32]

In seeking an analytical representation of the experimental solvation energies in Table 1, we tested a number of quantities related to $V_S(\mathbf{r})$, including some not given in Table 1 (the average deviation and the positive and negative variances); the selection of such quantities is guided by past experience, as was recently discussed,[13] as well as factors specific to the present objective that become apparent in the course of the regression analyses (see next section). Our best correlation is given by eq 2 and is shown in Figure 1, in which ΔG_{sol}(calc) is plotted against (exp).

$$\Delta G_{\text{sol}}\ (\text{kJ/mol}) = 0.17201 V_{\min} - 2.6412 \times 10^{-5}(V_{S,\max} - V_{S,\min})^3 + 5.1892 \times 10^{-2} A_S^- \bar{V}_S^- + 9.7042 \times 10^3 (A_S^- \bar{V}_S^-)^{-1} + 46.827 \quad (2)$$

The linear correlation coefficient is 0.988, the standard deviation is 1.57 kJ/mol, and the average absolute deviation is 1.14 kJ/mol.

In four of the molecules in Table 1, there are two relatively widely-separated (two intervening atoms) $V_{\min}$ of essentially equal magnitudes. They are associated with the nitrogens in 1,2-ethanediamine, piperazine, and 2-methylpyrazine and the oxygens in 1,4-dioxane. Since these $V_{\min}$ indicate two equivalent reactive sites, it seemed reasonable to combine their values in carrying out the regression analyses (Table 1). It might be anticipated that 1,2-ethanediol would also fit into this category; however, one of its $V_{\min}$ is considerably less negative than the other (−43.2 vs −55.2 kcal/mol), because it is associated with an oxygen that is involved in intramolecular hydrogen bonding.

Discussion

In order to better understand the functions of the variable terms in eq 2, we will examine their contributions separately. The experimental data correlate roughly with $(V_{S,\max} - V_{S,\min})^3$ alone, which can be viewed as a measure of the range of $V(\mathbf{r})$ on the surface. This indicates the significance of interactions with both the hydrogens and oxygens of the water molecules. The relationship is considerably improved by introducing $V_{\min}$ as a variable, suggesting that the interactions with the hydrogens are the more important. Thus, when ΔG_{sol} is expressed in terms of both $(V_{S,\max} - V_{S,\min})^3$ and $V_{\min}$, the linear correlation coefficient is 0.931 and the standard deviation is 3.71 kJ/mol. However, a very large error, nearly 14 kJ/mol, is obtained for anthracene, which also has the most extensive negative surface area, A_S^- (Table 1). Another of the larger errors, more than 6 kJ/mol, is for naphthalene, which is second in negative surface area. It seems evident, therefore, that an additional, size-dependent term is required. The product $A_S^- \bar{V}_S^-$ (a modified version of a quantity introduced by Grigoras[33]) satisfies this need. The correlation coefficient and standard deviation improve to 0.969 and 2.48, respectively. Yet further improvement is obtained (perhaps surprisingly) by including also the reciprocal of $A_S^- \bar{V}_S^-$, which finally produces eq 2.

The contributions of all of the variable terms in eq 2 are negative, so that they all tend to increase the strength of the solvation interaction. The role of the $(A_S^- \bar{V}_S^-)^{-1}$ term appears to be to somewhat level the larger effect of the $A_S^- \bar{V}_S^-$ term. Whereas the contributions of the latter range roughly from −20 to −50 kJ/mol, the sum of the two is generally between −45 and −52 kJ/mol; thus it typically balances the positive constant term. It is interesting to note that for the seven aromatic hydrocarbons in Table 1, ΔG_{sol}(exp) correlates well with $A_S^- \bar{V}_S^-$ alone.

In many of our earlier expressions for various condensed-phase properties, the total molecular surface area was a useful variable, reflecting the molecule's size. It is accordingly interesting that only the negative (not the total) area was found to play a significant role in representing aqueous solvation free energies, eq 2. This is particularly notable because theoretical treatments of solute-solvent interactions frequently model the solute molecule as forming and occupying a size-related cavity in the solvent.[4−9,34−36] Indeed the total surface areas in Table 1 correlate extremely well with free energies of cavity formation given by Luo *et al*;[8] the linear correlation coefficient is 0.996.

Conclusions

We have shown that the aqueous solvation free energies of 50 solutes of various chemical types can be represented at a good level of accuracy (standard deviation = 1.57 kJ/mol, average absolute deviation = 1.14 kJ/mol) in terms of properties of the electrostatic potentials on their molecular surfaces. It is anticipated that the same general approach can be used for other solvents, although other variables may be found to be of primary significance. For nonpolar solvents, for example, the quantities $A^-\bar{V}^-$ and/or $A^+\bar{V}^+$ may play dominant roles.

Acknowledgment. We greatly appreciate the financial support of the Office of Naval Research, through contract N00014-97-1-0066 and Program Officer Dr. Judah Goldwasser.

References and Notes

(1) Beveridge, D. L.; Mazei, M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

(2) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.

(3) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.

(4) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(5) Davis, M. E. *J. Chem. Phys.* **1994**, *100*, 5149.

(6) Cramer, C. J.; Truhlar, D. G. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; Chapter 2.

(7) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461.

(8) Luo, R.; Moult, J.; Gilson, M. K. *J. Phys. Chem. B* **1997**, *101*, 11226.

(9) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.

(10) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.

(11) Murray, J. S.; Politzer, P. In *Quantitative Treatments of Solute/Solvent Interactions*; Murray, J. S., Politzer, P., Eds.; Elsevier: Amsterdam, 1994; Chapter 8.

(12) Murray, J. S.; Brinck, T.; Lane, P.; Paulsen, K.; Politzer, P. *J. Mol. Struct. (Theochem)* **1994**, *307*, 55.

(13) Murray, J. S.; Politzer, P. *J. Mol. Struct. (Theochem)* **1998**, *425*, 107.

(14) Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M. E.; Politzer, P. *J. Phys. Chem.* **1993**, *97*, 9369.

(15) Politzer, P.; Murray, J. S.; Grice, M. E.; DeSalvo, M.; Miller, E. *Mol. Phys.* **1997**, *91*, 923.

(16) Politzer, P.; Lane, P.; Murray, J. S.; Brinck, T. *J. Phys. Chem.* **1992**, *96*, 7938.

(17) Politzer, P.; Murray, J. S.; Lane, P.; Brinck, T. *J. Phys. Chem.* **1993**, *97*, 729.

(18) Politzer, P.; Murray, J. S.; Flodmark, P. *J. Phys. Chem.* **1996**, *100*, 5538.

(19) Murray, J. S.; Gagarin, S. G.; Politzer, P. *J. Phys. Chem.* **1995**, *99*, 12081.

(20) Brinck, T.; Murray, J. S.; Politzer, P. *J. Org. Chem.* **1993**, *58*, 7070.

(21) Murray, J. S.; Brinck, T.; Politzer, P. *J. Phys. Chem.* **1993**, *97*, 13807.

(22) Murray, J. S.; Brinck, T.; Politzer, P. *Chem. Phys.* **1996**, *204*, 289.

(23) Politzer, P.; Murray, J. S. *J. Phys. Chem. A* **1998**, *102*, 1018.

(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1995.

(25) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. *J. Am. Chem. Soc.* **1987**, *109*, 7968.

(26) Scrocco, E.; Tomasi, J. *Topics in Current Chemistry*; Springer-Verlag: Berlin, 1973; No. 42; p 95.

(27) Politzer, P.; Daiker, K. C. In *The Force Concept in Chemistry*; Deb, B. M., Ed.; Van Nostrand Reinhold: New York, 1981; Chapter 6.

(28) *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P.; Truhlar, D. G., Eds.; Plenum Press: New York, 1981.

(29) Politzer, P.; Murray, J. S. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, Chapter 7.

(30) Náray-Szabó, G.; Ferenzy, G. G. *Chem. Rev.* **1995**, *95*, 829.

(31) *SAS*; SAS Institute Inc.: Cary, NC 27511.

(32) Murray, J. S.; Lane, P.; Politzer, P. *Mol. Phys.* **1998**, *93*, 187.

(33) Grigoras, S. *J. Comput. Chem.* **1990**, *11*, 493.

(34) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(35) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.

(36) Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. *J. Comput. Chem.* **1993**, *14*, 1240.