Charge-Dependent Cavity Radii for an Accurate Dielectric Continuum Model of Solvation with Emphasis on Ions: Aqueous Solutes with Oxo, Hydroxo, Amino, Methyl, Chloro, Bromo, and Fluoro Functionalities

Bojana Ginovska, Donald M. Camaioni,* Michel Dupuis, Christine A. Schwerdtfeger,[§] and Quinn Gil[||]

Chemical and Materials Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

Received: May 8, 2008; Revised Manuscript Received: July 9, 2008

Dielectric continuum solvation models are widely used because they are a computationally efficacious way to simulate equilibrium properties of solutes. With advances that allow for molecular-shaped cavities, they have reached a high level of accuracy, in particular for neutral solutes. However, benchmark tests show that existing schemes for defining cavities are unable to consistently predict accurately the effects of solvation on ions, especially anions. This work involves the further development of a protocol put forth earlier for defining the cavities of aqueous solutes, with resulting advances that are most striking for anions. Molecular cavities are defined as interlocked spheres around atoms or groups of atoms in the solute, but the sphere radii are determined by simple empirically based expressions involving the effective atomic charges of the solute atoms (derived from molecular electrostatic potential) and base radii. Both of these terms are optimized for the different types of atoms or functional groups in a training set of neutral and charged solutes. Parameters in these expressions for radii were fitted by minimizing residuals between calculated and measured standard free energies of solvation (ΔG_s*), weighted by the uncertainty in the measured value. The calculations were performed using density functional theory with the B3LYP functional and the 6-311+G** basis set and the COnductor-like Screening MOdel (COSMO). The optimized radii definitions reproduce ΔG_s* of neutral solutes and singly charged ions in the training set to within experimental uncertainty and, more importantly, accurately predict ΔG_s* of compounds outside the training set, in particular anions (*J. Phys. Chem. A* **2003**, *107*, 5778). Inherent to this approach, the cavity definitions reflect the strength of specific solute-water interactions. We surmise that this feature underlies the success of the model, referred to as the CD-COSMO model for Charge-Dependent (also Camaioni-Dupuis) COSMO model. These findings offer encouragement that we can keep extending this scheme to other functional groups and obtain better accuracy in using continuum solvation models to predict equilibrium properties of aqueous ionic solutes. The approach is illustrated for a number of test cases, including the determination of acidities of an amine base, a study of the tautomerization equilibrium of a zwitterionic molecule (glycine), and calculating solvation energies of transition states toward a full characterization of reaction pathways in aqueous phase, here in S_N2 exchange reactions. The calculated reaction barriers in aqueous solution are in excellent agreement with experimental values.

Introduction

Aqueous solvation of polar molecules and ions profoundly affects thermochemistry and reaction kinetics. Electronic structure calculations of gas phase properties of molecules achieve experimental accuracy. However, solution phase chemistry involves complex interactions of a solute with a large number of solvent molecules, and currently solvation models do not predict properties to the same accuracy as is seen in the gas phase, in particular for anions and cations. The present work involves the further development of a protocol based on a continuum model of solvation put forth earlier[1] and that is already demonstrating striking successes for this difficult class of species.

The prediction of solute/solvent interactions with solvation models dates back to the work of Born with his realization that a spherical cavity of a given radius interacting with a water continuum could accurately predict solvation energies of some ions.[2] Continuum solvation models make use of a solvent's dielectric constant (ε) to represent an average electrostatic field that is created by a large number of solvent molecules surrounding a solute.[3,4] Continuum solvation models have been implemented that calculate a gas phase wave function for a solute which is then used to calculate the reaction field[5,6] and the solute wave function gets recalculated in the presence of this reaction field. This process, known as the Self-Consistent Reaction Field method, is iterated until the reaction field and the solute wave function converge. A solute's cavity interacts with the polarizable solvent continuum through the electrostatic potential due to the surface charges induced by the response of the continuum to the solute's charge distribution. The solvation energy predicted by the continuum model depends on the strength

* To whom correspondence should be addressed.
[§] Current address: Department of Chemistry, University of Chicago, Chicago, IL 60637.
[||] Current address: Goldstar Consulting, Renton, WA 98058.

10.1021/jp804092v CCC: $40.75
Published on Web 09/25/2008

of interactions between the polarizable continuum and the solute in the cavity. Generally, a smaller cavity results in more favorable solvation energy. Therefore, it is necessary to define a cavity shape and size that will accurately reproduce experimental values of solvation free energies.

Improvements in algorithms for continuum models have made the use of more complex cavities possible. Since simple spherical cavities do not yield accurate predictions of solvation free energies, there has been an effort to develop molecular-shaped cavities that better account for the charge distribution and polarization of the solute and better capture the short-range solvation effects not accounted for by a simple spherical cavity. One approach has been to define a solute's cavity by interlocked spheres centered on each atom of the solute. Defining the radius of each sphere remains a major challenge in accurately predicting solution properties with solvation models. Bondi and Pauling radii have been used to define sphere radii, but Barone et al.[7] showed that using these radii does not lead to accurate predictions, nor is there a systematic error to translate the calculated values to experimental values. As a result, there has been a significant amount of effort focused on finding schemes to define cavities that can consistently reproduce experimental values.

Further progress in defining molecular shaped cavities was reported in the work of Barone et al.[7] These authors define the cavity radii according to their United Atom Hartree−Fock (UAHF) protocol. A solute's cavity is defined by putting a sphere on each atom, excluding hydrogens in most cases. Hydrogens in the solute are defined within the radius of the heavy atom to which they are bonded, as a united atom group. The radius of each atom's sphere is defined through a scheme that is parametrized, among other corrections, in terms of the row in the periodic table the atom resides in, the hybridization of the atom, its formal charge, and the number of hydrogens bonded to the heavy atom. Additional corrections are applied for solutes that carry a net charge. Benchmark tests show that this method gives solvation energies for neutrals that approach chemical accuracy, but it is unable to consistently and accurately predict the solvation energies of ions, especially anions. This observation remains essentially true even with recent improvements.

Camaioni et al.[1] gave a preliminary report of a novel cavity definition protocol that yields accurate solvation energies for oxoanions (XO_m^{n-}), using effective atomic charges derived via the CHarges from ELectrostatic Potential Grid (CHELPG) method.[5] In this approach, cavities are also built from interlocking atom-centered spheres with radii that are defined primarily by simple empirically based expressions involving the effective atomic charges of the solute atoms that fit the solute molecular electrostatic potential. The empirically based expressions also include a bond-length dependent factor to account for atomic size and hybridization. The scheme shows substantial qualitative differences with the UAHF protocol and other previously proposed schemes, for example, by assigning a large radius to the central atom of the oxoanions. Although these expressions were empirically derived, the authors put forth a strong theoretical basis for this new cavity scheme, based on analyses of the molecular electrostatic potential of oxoanions and of their interactions with a "solvent" water molecule.[1]

The present work extends the approach of Camaioni et al.[1] by adding homologous compounds to the training set, expanding the range of definitions, and reducing the standard deviations of cavity definition parameters. It also extends the approach to description of tautomeric reactions, and determining solvation effects on transition states and reaction pathways.[8] The protocol extends here, beyond the species in ref 1, to solutes having methyl, amino, hydroxo, chloro, bromo, and fluoro groups, including CH_4, NH_4^+, H_3O^+, OH, OH^-, Cl and Cl^-, Br and Br^-, F^-, and CH_3F prototypes. Unique radii definitions were needed for different functional groups to accommodate the variation in sensitivity of ΔG_s^* to the radius of atoms and groups in different structural environments. Prediction of the differential solvation between tautomers, in particular for compounds that adopt a zwitterionic form in solution, in contrast to a charge neutral form in the gas phase, is another example where the scheme can be successfully applied. A prototypical example is the glycine molecule for which we expect dramatic changes in radii when going from the neutral conformation to the zwitterionic conformation. The protocol extends naturally to determination of solvation effects on transition states and reaction pathways. The differential solvation effects between reactant, transition state, and product are well understood to reflect the changes in the charge distribution among these species. Since the cavity definition has the unique feature of being dependent on atomic charges and the atomic charges of the solute vary continuously from reactant to product along the reaction path, the model allows the cavity to change gradually from the reactant to the product state without discontinuity. Thus we expect our approach to be quite accurate also in the region of the transition state. We demonstrate here this capability for modeling accurately reactions in solution, with an initial study of solvation effects along the reaction path for prototypical self-exchange S_N2 reactions $X^- + R\text{–}X \rightarrow X\text{–}R + X^-$ for X = Cl, Br, F, OH and R = CH_3, CH_3CH_2, and C_5H_{11}.

Before getting into the current extensions, it is worth discussing two recent issues and developments that have come up in the literature: (i) the use of charge normalization schemes to account for the solute electron density that leaks into the continuum; and (ii) ways to capture the strong specific hydrogen-bond interaction of the solute with the solvent.

Various polarization charge normalization schemes have been discussed in the literature.[9] Recently, suggestions have been made that favor the use of the Solvent Volume Polarization and Integral Equation formulation of continuum solvation models,[10,11] without the need to apply charge renormalization, in contrast to older thinking and also in contrast to the approach proposed here. Here we used the COnductor-like Screening Model (COSMO)[12] of solvent continuum polarization in connection with two charge renormalization schemes. For geometry optimizations and frequencies, we used the widely accepted scheme whereby the charge is scaled by a constant factor (ICOMP = 2 in the Gaussian98 code) to get a charge polarization in accord with Gauss' law. For single point energies, we used the ICOMP = 4 option.[9] This scheme accounts for the effect of outlying charge by including an additional effective charge distributed according to the solute electronic density to satisfy Gauss' law. Although, charge normalization has little effect on the solvation energies of cations and neutral solutes, the effects are significant for anions. While such normalization schemes have fallen from favor, we found them to be critically necessary for the successful application of continuum solvation models to small anions such as O^-, O_2^-, and OH^- and other anions where the charge is as strongly localized as in these prototypical species. As indicated earlier, this charge-dependent model of solvation is abbreviated and referred to as CD-COSMO.

With regard to capturing specific interactions between solvent and solute in a continuum solvation model, Cramer and Truhlar[13,14] have proposed an approach that shows promise to overcome the difficulties associated with this goal. These authors

advocate the use of explicit microsolvation of the solute before applying a continuum model of solvation. Their approach succeeds in yielding improved accuracy over the base SM6 model of solvation,[14] albeit the error for anions remains significantly larger than for neutrals. These authors advocate the practice especially when the solute negative charge is strongly localized on one atom or on one functional group. The stronger the localization of the anionic charge, the more beneficial is the use of the explicit water molecule. In our approach, the strength of the specific interactions is built into the assignment of the radii according to the charge they carry in the solvated state.

The present paper is organized as follows: in Section 2 we describe our CD-COSMO protocol; in Section 3 we present results and discussions of the data for the training set and for the derivation of the empirical expressions for the atomic radii; in Section 4 we report applications of the empirically derived expressions in predicting the atomic radii of acids and conjugated bases for the calculations of pK_a, for zwitterionic molecules (glycine here), and also for activation barriers for several self-exchange S_N2 reactions $X^- + RX \rightarrow XR + X^-$ in aqueous phase. Summary and conclusions of the work are presented in Section 5. Explanation about the derivation of the empirical parameters for the cavity radii definitions is available in Appendix A.

Methods

Solvation calculations and geometry optimizations were performed using the Gaussian98 package.[15] Gas phase geometries were optimized using density functional theory (DFT),[5,6] with the B3LYP functional[16] and the 6-311+G** basis set.[17] Solvation calculations were carried out using the self-consistent-reaction-field (SCRF) method in the COSMO formulation of Klamt and Schüürmann.[12] CPCM is the implementation of COSMO in the Polarizable Continuum Model (PCM) suite of modules in Gaussian98. CPCM includes models for estimating nonelectrostatic contributions (cavity,[18] dispersion,[19] and repulsion[19])[3] (CDR) to the free energy in solution, which are included in the optimization calculations.[20] In this work, we used the same models for nonelectrostatic contributions but with radii defined by eqs 1a−i below. Optimizing solute geometries in solution did not significantly change calculated solvation energies of solutes in the training set. However, in some applications described below, we obtained better agreement with experiment when geometries were optimized.

To perform the solvation calculation, a cavity is defined for the molecule consisting of a sphere around each atom or functional group. The sphere radii are defined by simple functions of the effective atomic charges of the atoms in the solute. The effective atomic charges are calculated using the CHELPG method of atomic charge analysis in Gaussian98. Developing the functions defining the sphere radii required a fitting process that correlated the best relationship between CHELPG charges and cavity radii for solutes in the training set, and minimized the residuals between experimentally derived ΔG_s* and calculated ΔG_s*$_{calc}$.[21] The details of the procedure used in defining the radii functions are described in Appendix A.

For a given structure, the procedure begins with gas phase CHELPG charges and iterates the CHELPG charges (as functions of cavity radii) for each atom or functional group in the system until they converge, giving solution phase CHELPG charges and corresponding radii. Once converged, the ΔG_s*$_{calc}$ is calculated. The general procedure to calculate solvation free energies is summarized in steps 1−4 below:

(1) Define the cavity radii using the CHELPG charges for the solute in the gas phase at the gas phase optimized geometry.
(2) Use the cavity radii, carry out a CPCM calculation and calculate the CHELPG charges for the solute in the solution model.
(3) Define the cavity radii using the solution CHELPG charges.
(4) Repeat steps (2) and (3) until radii converge to within ±0.001 Å.

When optimizing the geometry of the molecule in solution, steps (5)−(7) are also included.

(5) Optimize the geometry using the solution model and the radii from step 4.
(6) Repeat steps (2)−(5) until the solution geometry radii converge to within ±0.001 Å.
(7) Perform frequency calculations of gas and solution geometries to obtain total free energies.

The difference between the solution and gas-phase total free energies gives the standard free energy of solvation (ΔG_s*). Different opinions exist about the proper treatment of molecular vibrations for a solute in a solvent. In the present work we adopted the definition of solvation advanced by Ben-Naim and Marcus.[21] They say that if solvent affects the internal partition function of the solute then the solvation free energy should include the free energy associated with these affects. Our view is that the CD-COSMO solvation free energy model gives the interaction free energy of the solvent with the solute, but does not include internal solute free energy changes. Accordingly, this approach of explicitly including the internal free energy change is most important when a solute structure is significantly different from its gas phase structure.

Step (2) in the procedure above makes use of the radii expressions that arise from the fitting to the training set of molecules (Appendix A). For specific atoms and functional groups the expressions are as follows:

$$R_O\,(\text{Å}) = -0.243|Q_O| + 1.700 \quad \text{for terminal oxygen} \tag{1a}$$

$$R_X\,(\text{Å}) = 0.440|Q_X| + 1.370D_{X\text{-}O} \quad \text{for central atom, } XO_n \tag{1b}$$

$$R_X\,(\text{Å}) = 0.440|Q_X| + 1510D_{X\text{-}O} \quad \text{for central atom, } XO_n^- \tag{1c}$$

$$R_{O,N}\,(\text{Å}) = -0.447|Q_{O,N}| + 1.852 \quad \text{for internal } -O \text{ or } -N \quad (-OH \text{ or } -NH_2) \tag{1d}$$

$$R_H\,(\text{Å}) = -0.447(n_H \times Q_H + Q_{O,N}) + 1.202 \quad \text{for } -OH \text{ or } -NH_2 \tag{1e}$$

$$R_{UA}\,(\text{Å}) = -0.206Q_{UA} + D_{X\text{-}H} + 1.089 \quad \text{for } -NH_3^+, -CH_3, -CH_2- \quad \text{(united atom treatment)} \tag{1f}$$

$$R_F\,(\text{Å}) = -0.830|Q_F| + 2.273 \quad \text{for } -F \tag{1g}$$

$$R_{Cl}\,(\text{Å}) = -0.241|Q_{Cl}| + 2.239 \quad \text{for } -Cl \tag{1h}$$

$$R_{Br}\,(\text{Å}) = -0.153|Q_{Br}| + 2.306 \quad \text{for } -Br \tag{1i}$$

We emphasize that eqs 1a−c were reported previously,[1] and are only included here for ease of reference. Equation 1a is slightly different than eq 1 in ref 1, due to the fact that the training set was extended to include OH, OH^-, and O_2H. Because eqs 1b and 1c have not changed nor has their training set extended, the evaluation of their performance (for species in or out of the training set) is not repeated in this work. In eqs 1b and 1c, $D_{X\text{-}O}$ refers to the distance in Ångstroms between the central X atom and the oxygen atom to which it is linked.[1] In eq 1f, $D_{X\text{-}H}$ refers to the distance between the "united atom" (N or C) and the H atom. The Qs are the CHELPG charges of the atoms (or group in the case of a united atom treatment, Q_{UA}) to which the cavity spheres are assigned. In all cases except eqs 1e and 1f, the radii expressions involve the absolute magnitude of the atomic charges, whereas eqs 1e and 1f involve the algebraic values of the charges. In eqs 1a, 1d, 1g−i, the absolute magnitudes are multiplied by a negative coefficient, reflecting the fact that the stronger the charges, the smaller the radii. A special charge term enters the radius expression for a hydrogen atom treated explicitly as in eq 1e. The charge term ($n_H \times Q_H + Q_{O,N}$) reflects the charge donation from the protons to O or N, and therefore the acidity of the protons, yielding smaller radii for more acidic protons and larger radii for less acidic protons. For example, $n_H = 1$ for −OH groups in H−OH, CH_3−OH, HO−OH, •OH, and •O−OH; $n_H = 2$ for $-OH_2^+$ and $-NH_2$ groups in hydronium ($H-OH_2^+$), ammonia ($H-NH_2$), and hydrazine (NH_2-NH_2). Similarly the united atom charge in eq 1f can be negative or positive. The base radius ($D_{X\text{-}H} + 1.089$) is large, and the actual radius is not changed drastically by the charge of the united atom. The resulting large value of the radius of the united atom reflects the lack of strong specific interaction of the united atom with the aqueous solvent, in contrast to eq 1e that captures the specific interaction of the aqueous solvent with the lone pair of O or N. The group charge, Q_{UA}, is the sum of the central atom's CHELPG charge plus $n_H \times Q_H$ where n_H is the number of hydrogen atoms that are part of the group. For example, n_H takes the value 3 for $-CH_3$ and $-NH_3^+$ groups in CH_3−H, CH_3−OH, CH_3-NH_2, $CH_3-NH_3^+$, $NH_2-NH_3^+$, and $H-NH_3^+$. Note that the training set includes molecules with $-CH_3$ functional groups, not with $-CH_2-$ functional groups. However, preliminary applications suggest that the parametrization for the $-CH_3$ united group can also be used as a parametrization for $-CH_2-$ groups without affecting the accuracy of the overall model. Charges and radii for each of the molecules in the training set used to derive eqs 1a, 1(d−i) are presented in Tables 1 and 2.

TABLE 1: Charges Obtained with CD-COSMO for the Molecules in the Training Set for This Work

	Q(−O)[a]	Q(N,O)[b]	Q(H)	Q($-NH_3^+$)	Q($-CH_3$)	radii eqs 1
O^-	−1.00					a
O_2^-	−0.50					a
O_2	0.00					a
OH	−0.472		0.472			a, e
OH^-	−1.417		0.417			a, e ($n_H = 1$)
H_2O_2		−0.477	0.477			d, e ($n_H = 1$)
O_2H	−0.216	−0.293	0.509			a, d, e ($n_H = 1$)
H_2O		−0.910	0.455			d, e ($n_H = 1$)
H_3O^+		−0.657	0.552			d, e ($n_H = 2$)
NH_3		−1.260	0.420			d, e ($n_H = 2$)
CH_3CH_3					0.000	f ($n_H = 3$)
CH_3NH_2		−1.232	0.431		0.369	d, e ($n_H = 2$), f ($n_H = 3$)
CH_3OH		−0.804	0.475		0.329	d, e ($n_H = 1$), f ($n_H = 3$)
NH_4^+				0.556		f ($n_H = 3$)
CH_4					−0.092	f ($n_H = 3$)
$CH_3NH_3^+$				0.760	0.240	f ($n_H = 3$)

[a] Terminal O atom bonded to only 1 other atom. [b] N or O atom in amino or hydroxyl group.

TABLE 2: Radii Obtained with Equations 1a, d−i for the Molecules in the Training Set for This Work

	R(−O)	R(N,O)	R(H)	R($-NH_3^+$)	R($-CH_3$)	D_{C-H}	D_{N-H}
O^-	1.457						
O_2^-	1.578						
O_2	1.700						
OH	1.585		1.202				
OH^-	1.356		1.649				
H_2O_2		1.639	1.202				
O_2H	1.647	1.721	1.105				
H_2O		1.445	1.405				
H_3O^+		1.558	1.002				
NH_3		1.289	1.390				
CH_3CH_3					2.183	1.094	
CH_3NH_2		1.301	1.367		2.113	1.100	
CH_3OH		1.492	1.349		2.116	1.095	
NH_4^+				2.001			1.026
CH_4					2.199	1.091	
$CH_3NH_3^+$				1.957	2.128	1.088	1.028

Results and Analysis

The radii definitions given above are found to produce accurate results for ΔG_s^* when used with the ab initio COSMO solvation model. Table 1 shows the compounds currently in the training set used to derive equations (1a, 1d−i), the charges, and the radii equations and associated n_H value (when applicable). The corresponding CD-COSMO radii resulting from the data in Table 1 are listed in Table 2. Experimental and calculated ΔG_s^* values from this work and from UAHF,[7] SM6,[13] UAKS[22] and MST-HF[23] are given in Table 3. This training set includes oxides, hydroxy compounds, and some small amino- and methyl-containing compounds. In addition, Table 3 lists the uncertainty or error in the experimental ΔG_s^*. These errors were used to weight the contribution of the data points in determining the definition equations. The species with large experimental errors were given less significance in the fitting procedure (Appendix A) than the species whose experimental free energy of solvation is known with better accuracy. This was done so that the uncertainties in experimental ΔG_s^* values are not propagated to the radii definitions.

Unique radii definitions are needed for different functional groups to accommodate the variation in sensitivity of ΔG_s^* to the radii of atoms and groups in different structural environments. Compounds with the same central atoms, such as nitrogen or oxygen, without lone pairs are also found to interact differently with the water continuum compared to compounds with lone pairs; these tetrahedral central atoms are given a different radius definition to reproduce the experimental ΔG_s^*. Within each category of atoms, the radius definition can predict sphere sizes that correlate with the interaction of each atom with the solvent continuum. For example, hydrogens with larger CHELPG charges, like in H_3O^+, are predicted to have smaller spheres, showing a close and strong electrostatic interaction with solvent water.

As seen in Table 3, the radii definitions parametrized in terms of CHELPG charges continue to provide accurate ΔG_s^* upon expansion to a larger range of functional groups. They also accurately predict the ΔG_s^* of singly charged anionic and cationic species, a shortcoming seen in the results predicted by other cavity definition schemes.[13,14,22] Adding compounds to the original training set[1] also improved the standard deviations in the parameters of the fitted radii definitions. The mean unsigned error of ΔG_s^* is 0.4 kcal/mol. The mean unsigned error of the values predicted by the UAHF scheme for compounds included

TABLE 3: Comparison of Experimental ΔG_s* (kcal/mol) with Values Calculated by This Work eqs 1a,d−i, the UAHF Scheme, SM6,[14] UAKS-CPCM,[22] and MST-HF[23]

	expt	ref	CD-COSMO	UAHF	SM6[a]	UAKS[b]	MST-HF[c]
O^-	−100.0 ± 1.0	57	−100.2	−104.1	−114.9	−91.5	
O_2^-	−82.3 ± 1.2	57	−81.5	−78.7	−94.0		
O_2	2.2 ± 0.0	57	2.2	0.9	−3.2		
OH	−3.5 ± 1.4	58	−4.8	−5.6	−9.3		
OH^-	−104.6 ± 0.5	59	−103.9	−111.2	−115.2	−101.8	−104.0
H_2O_2	−8.6 ± 0.5	60	−8.5	−12.2	−7.5	−12.2	
O_2H	−7.0 ± 1.0	61	−7.0	−7.6	−7.6		
H_2O	−6.3 ± 0.0	21, 32	−6.1	−7.5	−9.1	−7.3	
H_3O^+	−110.2 ± 0.2	59	−109.6	−107.7	−102.3	−108.1	−105.9
NH_3	−4.3 ± 0.3	21	−3.9	−5.1	−4.6	−5.0	
CH_3CH_3	1.8 ± 0.1	21	1.8	2.0	1.3		
CH_3NH_2	−4.6 ± 0.3	13	−3.7	−5.3	−4.3	−5.4	
CH_3OH	−5.1 ± 0.2	21	−5.1	−6.3	−5.0	−5.8	
NH_4^+	−85.2 ± 0.2	33	−85.3	−80.0	−91.6	−80.4	−93.1
CH_4	1.9 ± 0.2	21	1.9	1.8	2.0		
$CH_3NH_3^+$	−76.5 ± 0.2	33	−77.6	−70.7	−78.5	−71.2	−80.2
muse[d]	0.5		0.4	2.5	4.4	3.0	4.1

[a] B3LYP/6-311+G**. [b] B3LYP/6-31+G*. [c] HF/6-31+G*. [d] Mean unsigned error.

in the training set used here is 2.5 kcal/mol and for SM6 it is 4.4 kcal/mol, which is largely due to the errors in the prediction of the solvation energy for small ions (O^-, O_2^-, OH^-, and H_3O^+). The mean unsigned errors for a subset of molecules from the training set with UAKS-CPCM and MST-HF are 3.0 and 4.1 kcal/mol. The values from the five schemes are compared with experimental values in Table 3.

To summarize, our findings certainly support the notion that continuum solvation models do well in modeling the interactions of the solute with the bulk solution. It is also known that incorporating effects of short-range solute−solvent interactions is problematic and a major source of error. The present protocol of parametrizing cavity radii as functions of CHELPG charges provides highly encouraging results. An excellent fit to the training set is obtained, but more importantly, as will be discussed below, the fitted cavity definitions trend with chemical properties and give physically meaningful solvation cavities. We surmise that the apparent success of this approach is due to empirically accounting for short-range interactions.

Before going over individual cases of the various chemical functionalities described here, we would like to note that including HOO^- in the training set always introduces a bigger error in the fit. The ΔG_s* predicted by the radii definitions yields −93.4 kcal/mol that is approximately 3 kcal/mol less negative than the experimental value, $\Delta G_s^* = -96.3$ kcal/mol.[24] Other schemes[22,23] have shown difficulty in fitting this ion as well, with errors ∼6 kcal/mol. This anomalous behavior suggests that a study of the nature of the solvation of this ion is needed. Also, a review of experimental determinations of ΔG_s* of HOO^- is warranted.[25]

Hydrogen Radii in Functional Groups. Hydrogens attached to N or O as in amino and hydroxyl groups are given spheres with radii defined by eq 1e because these groups can accept and donate H-bonds. A united atom treatment in which the radius is centered on the central N or O, as adopted in the UAHF scheme, does not allow the electron density of the hydrogens and the lone pairs to interact with the continuum independently. While the charge-dependence and base radius for hydrogen radii may differ for these functional groups, we did not discern a significant difference between hydrogens attached to N and hydrogens attached to O for solutes in the training set. Therefore a single equation was adopted to define H radii for these two functional groups. The hydrogen radius is defined by the sum of the charges of the atoms in the group. By definition, the base radius corresponds to a hypothetical group in which the sum of the group charge is zero. Using the charge of the group better represents the properties exhibited by the hydrogen substituent and translates the interactions of hydrogen with the water solvent into a representative radius. As a result, the cavity radii trend with pK_a of acidic hydrogens, for example, H_2O (pK_a = 15.7, r_H = 1.405), OH (pK_a = 11.9, r_H = 1.202), H_2O_2 (pK_a = 11.7, r_H = 1.202), O_2H (pK_a = 4.8, r_H = 1.105) and H_3O^+ (pK_a = −1.7, r_H = 1.002). Stronger acids show greater ionic character in the acidic hydrogen bond and increased separation of charge, leading to stronger electrostatic interactions with the solvent and smaller cavity radii. By taking into account the group charge, these differences in acidity and interaction strength are accounted for more accurately.

Functional Groups with No Lone Pairs (Tetrahedral sp^3 Nitrogen and Carbon). Compounds that have no lone pairs show different chemistry than their counterparts with lone pairs, and this difference affects the way they interact with a solvent. For example, there are lone electron pairs in oxygen compounds and neutral nitrogen compounds. Lone pairs allow oxygen and nitrogen to be acceptors of hydrogen bonds with solvent water. This means that the solvation spheres around nitrogen and oxygen will be smaller in order to reflect the stronger interactions with the solvent. Tetrahedral nitrogen and carbon compounds have no lone pairs and cannot act as hydrogen bond acceptors. As a result, there is a much weaker interaction between the central nitrogen and carbon atoms with solvent water. Since the weak interactions with solvent cause the radius of the central atom to be larger than atoms with lone pairs, these tetravalent compounds can be simply treated as united atom groups and the solvation of the entire group is captured with one spherical cavity centered at the C or N atom. Note that definitions of these tetrahedral sp^3 nitrogen and carbon groups (eq 1f) include the distance of the N−H or C−H bonds. N−H bonds are shorter than C−H bonds such that including this variable allows the same base radius and charge dependence to be used. Typically, the $-NH_3^+$ group carries more positive charge than a $-CH_3$ group (for example in $CH_3-NH_3^+$, the charges are 0.24 and 0.76 for $-CH_3$ and $-NH_3^+$, respectively), so it leads to a radius for the nitrogen tetrahedron that is smaller

than that of carbon. This is justified because the cation has a stronger interaction with water and therefore obtains a smaller cavity.

Terminal Oxygen. Terminal oxygen atoms were given radii definitions different from oxygen atoms attached to more than one atom. A good fit cannot be obtained if these two types of oxygens are fitted with the same radius definition. This could result from the terminal oxygens being more exposed to solvent molecules and interacting differently compared to the nonterminal oxygens. However, the difference is not well understood. Analyses of electrostatic potential maps and interaction energies with water may give more insight into the differences between these atom types.

Internal Oxygen and Nitrogen. Oxygen and nitrogen were fitted using the same radius definition (eq 1d). Although oxygen is more electronegative, the CHELPG charges represent differences in nitrogen's and oxygen's ability to stabilize negative and positive charges and portray these characteristics in radius sizes. Other evidence suggests that the oxygen and nitrogen radii may be closely related. Bond critical points were calculated with Gaussian98 for a gas phase water−water dimer and ammonia−water dimer. This calculation shows that the distance to the critical point (point of minimum electron density between two atoms) in the hydrogen bond between oxygen and hydrogen was 1.257 Å from the oxygen while the distance to the critical point in the hydrogen bond between nitrogen and hydrogen (from the ammonia−water dimer) was 1.305 Å from the nitrogen. The total distance between oxygen and hydrogen was 1.96 Å and between nitrogen and hydrogen was 1.99 Å. The $R_{N,O}$ radii in Table 2 show that in practice, due to the scaling of the CHELPG charges, the nitrogen radius is usually smaller than the oxygen radius. A molecular dynamics study of methylamine in water and water in water shows that the first solvation shell is slightly closer to methylamine than it is to water,[26] suggesting that the result predicted by the best fit radii definitions is reasonable.

To summarize, this work has advanced on our approach of using charge-dependent cavity radii for the COSMO solvation model by adding compounds to the training set and adjusting/developing new radii definitions when chemically justified and necessary. It also shows that parametrizing the solvation sphere radii in terms of CHELPG charges can account for a wide range of behavior in solutions for neutrals, anions, cations, and dications. Further work is warranted to develop this approach. The addition of more compounds to the training set is expected to improve the existing radii definitions as well as expand the applicability of the scheme to handle more atoms/groups, for example, alkenes, alkynes, arenes, sulfides, and metals, that are not represented in the current training set.

It is worth noting again that the errors in the experimental data were taken into account when optimizing the cavity definitions. Some of the experimental ΔG_s* values, especially for ions, have larger uncertainties (without taking into account the uncertainty in ΔG_s*(H^+) that affects the solvation energy of all ions).[14,27,28] Since the calculated ΔG_s* in some compounds is very sensitive to cavity size, fitting closely to uncertain data could skew the radii definitions and have adverse effects when applying the definitions to compounds outside the training set. Table 3 contains the list of experimental errors that was used in weighting the residuals in the fitting process in order to minimize these effects. A final comment is in order: the training set in this work is smaller than the one used in ref 7 but in many cases gives results closer to experimental values. This could result from fitting a smaller number of compounds and therefore having a smaller amount of error to distribute.

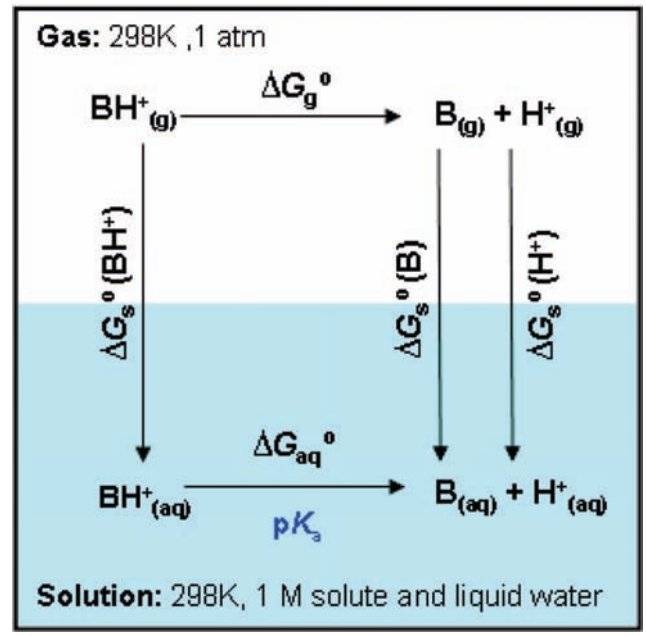

Figure 1. Thermodynamic cycle showing relationship of pK_a to gas phase acidity and free energies of solvation.

However, correlating the radii to the CHELPG charge, rather than the formal charge, allows for more flexibility in accommodating compounds within the same radii definition class and is likely to account for the improved results.

Applications

We carried out test applications of the proposed model of solvation. One application involved the prediction of acidity, a traditionally challenging test of continuum solvation models.[29−31] A second test involved the calculation of the tautomerization and solvation free energies for a zwitterionic compound, namely glycine. Such a test is also very popular and challenging for continuum models and discrete quantum mechanics/molecular mechanics (QM/MM) models of solvation. The change in the electronic charge distribution between the neutral and zwitterionic forms of glycine should be an appropriate test of our solvation model in which we assign a critical role to the solute atomic charges. Finally, we applied our model to the characterization of prototypical reaction energetics, namely for a set of S_N2 reactions. The differential solvation among reactants, transition states, and products is well understood to be associated with differences in the electronic charge distribution among these species. This is again a stringent test of our charge-dependent solvation model.

Acidity Prediction. The applicability of the current radii definitions were tested by calculating the pK_a of protonated forms of hydrazine, including doubly protonated hydrazine ($NH_3NH_3^{2+}$), ammonia, methylamine, and methanol. The pK_a values were calculated using eqs 2 and 3

$$pK_a(BH^+) = \frac{\Delta G^\circ_{aq}}{2.303RT} \quad (2)$$

$$\Delta G^\circ_{aq} = \Delta G^\circ_g + \Delta G^\circ_s(B) + \Delta G^\circ_s(H^+) - \Delta G^\circ_s(BH^+) \quad (3)$$

where ΔG°_g is the gas phase acidity (1 atm standard state) and ΔG°_s is the Gibb's standard free energy change for transfer from gas phase, 1 atm standard state, to aqueous solution, 1 M standard state (note that $\Delta G^\circ_s = \Delta G_s$* + 1.89 kcal/mol[21] and $\Delta G^\circ_s(H^+) = -264.0$ kcal/mol.[27,32,33]) Figure 1 shows the

TABLE 4: Free Energies (kcal/mol) of Solvation and Acidities for Protonated Amines and Alcohols

$BH^+ \rightleftharpoons B + H^+$	ΔG_s*(B)		ΔG_s* (BH^+)		gas phase acidity		pK_a	
BH^+	expt	CD-COSMO	expt	CD-COSMO	expt[a]	G3B3	expt	calc.
$NH_3NH_3^{2+}$	−87.5[b]	−89.6[c]	−305.0[b]	−304.5[c]		45.1	−1.05[d]	−3.0
$NH_2NH_3^+$	−9.3[b]	−10.6[c]	−87.5[b]	−89.6[c]	196.6	199.4	7.9[d]	8.8
NH_4^+	−4.3[e]	−3.9	−85.2[f]	−85.3	195.7	197.4	9.3[g]	9.6
$CH_3OH_2^+$	−5.1[e]	−5.1	−93.1[h]	−90.7	173.2	173.3	−2.0[i]	−3.8
$CH_3NH_3^+$	−4.6[j]	−3.7	−76.5[f]	−77.6	206.6	208.2	10.7[g]	12.1

[a] Reference 39. [b] References 37 and 38. [c] Solution optimized structures. [d] Reference 40. [e] Reference 21. [f] Reference 33. [g] Reference 41. [h] Reference 59. [i] Reference 42. [j] Reference 13.

thermodynamic cycle from which eq 3 is derived. The experimental and calculated pK_a values are shown in Table 4. The gas phase acidities were calculated with G3B3[34] using $G_g(H^+) = -0.00999$ hartree.[35] Comparisons with known experimental gas phase acidities (Table 4) show average discrepancy of 1.6 kcal/mol.[36] Since $NH_3NH_3^{2+}$ does not have an experimentally determined gas phase acidity, the G3B3 value of $\Delta G_{g2} = 45.1$ kcal/mol is used with experimental values of ΔG_s* for protonated hydrazine[37,38] and H^+ in combination with the pK_a of $NH_3NH_3^{2+}$, to estimate ΔG_s*($NH_3NH_3^{2+}$) = −305.0 kcal/mol. To calculate the ΔG_s* value using the continuum model, the united atom radius definition eq 1f was used (gas and solution CHELPG charge of the group is +1). We found that reoptimizing the structure of $NH_3NH_3^{2+}$ in solution changed the free energy of solvation from −301.8 to −304.5 kcal/mol. As a result, NH_2NH_2 and $NH_2NH_3^+$ were optimized in solution for the pK_a calculations.[39−42] The calculated pK_a value for $NH_3NH_3^{2+}$ is within 2 pK_a units of experiment. For $NH_2NH_3^+$, NH_4^+, $CH_3OH_2^+$, and $CH_3NH_3^+$, the agreement ranges between 0.3 and 1.8 pK_a units. Takano and Houk[22] have calculated pK_a values in water for NH_4^+, $CH_3OH_2^+$, and $CH_3NH_3^+$ that show better agreement with experiment using the CPCM model with United Atom Kohn−Sham (UAKS) cavities to determine the free energy of solvation. Since UAKS cavities do not give free energies of solvation as accurately as our CD-COSMO cavities do, we suggest the apparent agreement with the experimental pK_a values of the CPCM UAKS model is due to a fortuitous cancelation of errors.

These above results demonstrate a promise of robustness for the approach taken in this work for defining cavity radii for a variety of species. A systematic study of acidities in carboxylic acids will be reported.

Zwitterionic Equilibrium. Glycine plays an important role in chemistry because of its tautomeric forms with dramatically different chemical character. The two forms of glycine, the neutral form (NT) and the zwitterionic form (ZT) have been extensively characterized experimentally and theoretically in gas phase and in aqueous phase. It is known that the ZT form is more stable and predominant in aqueous phase. Experimental energy estimates for transferring glycine from the gas phase to aqueous phase and the free energy difference between the two forms in solution are available and often serve to benchmark new theoretical models. The free energy difference between the NT and ZT forms can in fact be viewed as an appropriate and sensitive test of our model, given that we give a special significance to the partial charge of the solute atoms and that those are dramatically different between the NT and the ZT forms.

Table 5 shows results calculated using our CD-COSMO model and compares them with values derived from known physical and thermochemical data for glycine.

For the process NT(g) ⇌ ZT(aq), $\Delta G^* = -20.5 \pm 1.5$ kcal/mol,[43−47] for NT(aq) ⇌ ZT(aq), $\Delta G^* = -7.3$ kcal/mol,[48] and therefore, ΔG_s* (NT) = −13.2 ± 2.0 kcal/mol. To calculate the free energies for these processes, we performed geometry optimizations at B3LYP/6-311+G** level for NT in vacuum and for NT and ZT in solution using the CD-COSMO protocol described above. G3B3 calculations were also performed on these geometries and used to correct the DFT electronic energies to G3B3 energies. For the process NT(g) ⇌ ZT(aq), we calculate $\Delta G^* = -21.7$ and −20.3 kcal/mol using DFT and G3B3 theory, respectively. Both values are within the uncertainty of the measured value. For the process, NT(aq) ⇌ ZT(aq), the DFT calculation gives 11.7 kcal/mol, which differs from experiment by 4.4 kcal/mol. The calculation at G3B3 level gives 10.0 kcal/mol and differs by 3 kcal/mol from experiment. The uncertainty in the measured value for NT(aq) ⇌ ZT(aq) is small. Therefore, the error is due to a combination of errors in electronic and solvation energies. Considering the average error for G3B3 is ∼1 kcal/mol and for our solvation model, 0.5−1 kcal/mol,[49] the agreement is within expectations.

TABLE 5: Standard Free Energy Changes (ΔG^*) for Glycine Neutral (NT) and Zwitterion (ZT) Equilibria in kcal/mol

reaction	B3LYP/ 6-311+G** CD-COSMO[a]	G3B3 CD-COSMO[b]	exp
NT(g) ⇌ ZT(aq)	−21.7	−20.3	−20.5[c]
ZT(aq) ⇌ NT(aq)	11.7	10.0	7.3[d]
NT(g) ⇌ NT(aq)	−10.0	−10.3	−13.2

[a] Geometries of solutes reoptimized and frequencies and thermal corrections calculated using CD-COSMO. [b] B3LYP/6-311+G** CD-COSMO values corrected to the G3 electronic energies calculated with B3LYP/6-311+G** geometries. [c] Reference 43. [d] Reference 48.

As a comparison, Truong and Stevanovich[50] used the DFT(B3LYP)/6-31G** level of theory also in conjunction with the COSMO model of solvation but with nonvarying atomic radii that they had fitted to solvation energies for a training set of molecules. Thus, in contrast with the present work, the radii used by these authors are fixed and do not change when going from the NT form to the ZT form. They found the difference in solvation energies between the NT and ZT forms to be sensitive to the structures used, but when using the gas phase geometry for the NT form and the solvated structure for the ZT form, they obtained the computational result that the ZT form is more stable by up to 4.0 kcal/mol than the NT form, depending on the level of theory, compared to the experimental value of 7.3 kcal/mol.

Reaction in Solutions. With its emphasis on solute atomic charges, the present scheme is ideally suited to the calculation of aqueous solvation energies of transition states. It is well understood that the differential strength of solvation between reactants/products and transition states gives rise to increase or decrease in activation barriers. As a rule of thumb, the localized

TABLE 6: Activation Barriers (kcal/mol) for Selected Self-Exchange S_N2 Reactions in Gas and Aqueous Phase

$X^- + RX \rightarrow [X–R–X]^{-‡}$		$\Delta G^‡$ (gas)		$\Delta G^‡$ (aqueous)				
				CD-COSMO				
X	R	G3B3[a]	B3LYP	G3B3[a]	B3LYP	MCFEP[b]	CPCM[c]	expt[d]
F	CH_3	4.6	2.0	34.2	31.7			31.8
Cl	CH_3	9.2	4.7	27.2	22.7	26.0	35.3	26.5
Br	CH_3		3.3		22.0			23.7
HO	CH_3	20.1	17.4	42.4	39.7			41.8
Cl	CH_3CH_2	10.6	7.0	28.4	24.9	23.9	38.4	
Cl	$(CH_3)_3CCH_2$	17.1	14.3	36.4	33.6	30.4	47.6	

[a] G3B3 is G3 variant using B3LYP model in Gaussian98. [b] MCFEP denotes the Monte Carlo Free Energy Perturbation method.[52,53] [c] Reference 53, CPCM is the COSMO model in Gaussian98. [d] Reference 54.

TABLE 7: ΔG_s* in (kcal/mol) Calculated with CD-COSMO and Charges and Radii Used in the Calculation

		charges					radii				
species	ΔG_s*	halogen	CH_3	H	O	CH_2	halogen	CH_3	H	O	CH_2
F^-	−104.4	−1.000					1.443				
Cl^-	−74.6	−1.000					1.998				
Br^-	−68.4	−1.000					2.153				
OH^-	−103.9			0.417	−1.417				1.649	1.356	
FCH_3	−0.2	−0.329	0.329				2.000	2.113			
$ClCH_3$	−0.6	−0.245	0.245				2.180	2.126			
$BrCH_3$	−0.8	−0.248	0.248				2.268	2.124			
$HOCH_3$	−5.1			0.475	−0.804				1.349	1.492	
EtCl	−0.4	−0.294	0.044			0.250	2.168	2.173			2.126
			−0.157					2.214			
$(CH_3)_3CCH_2$ Cl	1.5	−0.254	−0.186			−0.003	2.178	2.221			2.180
$[FCH_3F]^{-‡}$	−73.0	−0.785	0.570				1.622	2.045			
$[ClCH_3Cl]^{-‡}$	−55.2	−0.751	0.502				2.065	2.058			
$[BrCH_3Br]^{-‡}$	−48.6	−0.748	0.496				2.192	2.263			
$[HOCH_3HO]^{-‡}$	−84.7		0.531	0.346	−1.112			2.053	1.544	1.355	
$[ClEtCl]^{-‡}$	−55.1	−0.799	0.019			0.578	2.047	2.176			2.042
			−0.139					2.210			
$[ClNpCl]^{-‡}$	−51.8	−0.800	−0.170			0.313	2.045	2.218			2.097

or delocalized character of the charge distribution is an easy indicator of weaker or stronger solvation of the transition state compared to reactants and products. The S_N2 exchange reactions $X^- + CH_3X$, involving halogen and other nucleophiles have attracted considerable theoretical attention since the early work of Shaik[51] and Jorgensen and co-workers[52] as prototypes for benchmarking theoretical models of solvation.[51−53] We carried out a parametrization of the Cl radius using a two-point fit between Cl^- and CH_3Cl, as those molecules span the range of Cl atoms carrying anywhere from a very strong charge to a small charge. The resulting simple functional form was given above. The same approach was applied to the parametrization for bromine atoms using Br^- and CH_3Br, and for fluorine atom using F^- and CH_3F; all compounds for which an experimental solvation free energy is available. With the availability of radius definitions for other functional groups, such as $-CH_3$, $-CH_2-$ (same united atom description for both), we calculated the activation barriers for a number of S_N2 reactions. They are given in Table 6. The details about the free energy of solvation for the species in the S_N2 reactions are provided in Table 7.

In gas phase, we calculated the activation barriers using the G3B3[34] scheme for all the reactions except for the Br self-exchange reaction, because this scheme is not available for bromine-containing species. The differential energy of solvation was then calculated employing the methodology described in this work, and the activation barriers in solution were estimated by combining the gas phase activation barrier calculated at the G3B3 level of theory with the differential energy of solvation using the approach of this work. For the reactions $CH_3X + X^-$, where X = F, Cl, Br, and HO, experimental barrier heights have been reported by Albery and Kreevoy[54] and summarized by Shaik.[24] The $RCl + Cl^-$ reactions, where $R = C_2H_5$ and C_5H_{11}, have no experimental barriers available, but have been studied using Monte Carlo free energy perturbation (MCFEP) by Jorgensen and co-workers.[53] Our calculated free energy barriers approach chemical accuracy compared to experiment. Note that values presented in Table 6 are not corrected for basis set superposition errors as those are small (between 0.2 kcal/mol for $BrCH_3Br^-$ and 1.9 kcal/mol for $HOCH_3HO^-$ at B3LYP/6-311+G** level of theory).

The plain application[53] of the COSMO/UAHF model with the B3LYP/6-31G* level of theory basis set yielded an activation energy of 35.3 kcal/mol for $ClCH_3$, which is in fact much worse than the earlier theoretical value of 26 kcal/mol by Jorgensen and co-workers[52] obtained with MC FEP calculations. Cossi et al.[55] obtained better results (E_a = 26.5 kcal/mol) using DFT (mPW1PW/6-311+G2df,2p) theory and the COSMO method with a united atom topology model of the cavity in which radii for Cl was allowed to vary with the formal charge. Truong and Stevanovich[56] reported an activation energy of 16 kcal/mol using the B3LYP/6-31+G* level of theory in connection with COSMO and their own set of fixed radii as already mentioned earlier in the glycine test application. We note the large variations in activation energies associated with varied levels of theory reported by these authors. The same observation can be made for the self-exchange reactions involving ethyl chloride (C_2H_5Cl) and neopentyl chloride ($C_5H_{11}Cl$) where the CPCM/UAHF method gives barriers that are 8.5 and 10.1 kcal/mol higher than those calculated with CD-COSMO. For these systems, our calcu-

lated barriers are 4.5 and 6.0 kcal/mol higher than the MC FEP barriers. The reasons for these discrepancies are not clear.

Summary and Conclusions

Dielectric continuum solvation models are widely used because they are a computationally efficient way to simulate equilibrium properties of solutes and have reached a high level of accuracy, in particular for neutral solutes. It is fair to say that until now they have been unable to consistently and accurately predict the effects of solvation on ions, especially anions. We presented here the extended development of a protocol put forth earlier and from now referred to as CD-COSMO, for defining the cavities of aqueous solutes with resulting advances that are most striking for anions. The radii of the spheres that make up the cavity are determined by simple empirically based expressions involving the effective atomic charges of the solute atoms (derived from molecular electrostatic potential) and base radii and interatomic distances in some appropriate cases. These terms were optimized for different types of atoms or functional groups in a training set of neutral and charged solutes. Our model uses the well established density functional level of theory with the B3LYP functional and the 6-311+G** basis set. Inherent to this approach, the cavity definitions reflect the strength of specific solute-water interactions. These findings offer encouragement that we can keep extending this scheme to other functional groups and obtain better accuracy in using continuum solvation models to predict equilibrium properties of aqueous ionic solutes. The approach was illustrated for a number of test cases, including the calculation of first and second acidities for conjugate acids of hydrazine, and the calculation of the relative energy of the neutral and zwitterionic forms of glycine in solution. The approach is also extended to calculating solvation energies of transition states for full characterization of reaction pathways in aqueous phase.[8] The reaction barriers in aqueous solution calculated with our model for prototypical S_N2 exchange reactions were found to be in excellent agreement with experimental values. Systematic studies of other challenging issues such as acidities in carboxylic acids and other studies are in progress and will be reported in future publications.

Acknowledgment. This research was conducted at the Pacific Northwest National Laboratory (PNNL). C.A.S. and Q.G. acknowledge support by the U.S. Department of Energy (DOE), Office of Science's Science Undergraduate Laboratory Internship (SULI) program. D.M.C. acknowledges support from the Office of Environmental and Biological Sciences' Environmental Research Sciences Program. B.G. and M.D. were supported by the DOE's Office of Basic Energy Sciences, Condensed Phase and Interfacial Molecular Science program. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy.

Appendix A

The CD-COSMO solvation model described in this paper is based on a definition of the molecular cavity as interlocked spheres whose radii are defined by simple functions of the effective atomic charges of the atoms in the solute. The effective atomic charges are charges that fit the molecular electrostatic potential (CHELPG charges). Water, a high dielectric solvent, interacts more strongly with charged compounds, and this strong interaction is reflected in smaller cavity radii. By using CHELPG charges as parameters for defining cavity radii, the strength of the interactions between the solvent and charged solutes and neutral solutes is well represented, and this representation leads to a more accurate prediction of solvation energies for ions than in earlier schemes.

Developing the functions defining the sphere radii required a fitting process that correlated the best relationship between CHELPG charges and cavity radii for solutes in the training set. Parameters in the cavity definitions were determined using a fitting program that minimized the residuals between experimentally derived ΔG_s^* and calculated $\Delta G_s^*{}_{calc}$. Making the empirical relations (eq 1a−1k) involved varying the parameters in the radii definitions until the sum of the residuals between experimental and calculated ΔG_s^* reached a minimum. To facilitate the procedure, functional dependences of CHELPG charge and ΔG_s^* on cavity radii were determined prior to the residual minimization. The functional dependences were determined such that consistent sets of radii and solution charges could be found through an iterative process. To this end, Gaussian98 calculations were performed for ranges of radii required for each molecule in the training set. Radii ranges were chosen such that the calculated ΔG_s^* bracketed the experimental ΔG_s^*. For ammonia, for example, calculations were performed for a nitrogen radius range between 1.20 and 2.20 Å and a hydrogen radius range between 0.90 and 1.45 Å. Each set of radii values gave different CHELPG charges and ΔG_s^* and acted as a data point, allowing a polynomial fit of the CHELPG charges and ΔG_s^* to be determined as functions of the nitrogen and hydrogen radii. The process was repeated for all the molecules in the training set. Adding a molecule to the training set requires scanning of the solvation energy ΔG_s^* and atomic CHELP charges over ranges of atomic radii values, to extract polynomial fits of these quantities as functions of the radii.

To this date the training set giving rise to the parametrization given earlier included (i) *Anions*, HO^-, O^-, O_2^-, HCO_2^-, O_3^-, NO_2^-, ClO_2^-, NO_3^-, Cl^-, Br^-, F^-; (ii) *Cations*, H_3O^+, NH_4^+, $CH_3NH_3^+$, $NH_2NH_3^+$; and (iii) *Neutrals*, H_2, H_2O_2, HO_2, H_2O, CH_3OH, CH_3NH_2, NH_3, NH_2NH_2, HO, SO_2, ClO_2, O_3, CO_2, NO_2, O_2, CH_3CH_3, CH_4, CH_3Cl, CH_3Br, CH_3F. It should be noted that the definition for oxoanions and neutrals with central X atom (eqs 1b and 1c.) were inferred using somewhat different optimization procedure, as described in our previous publication,[1] and were here adopted without a change.

As indicated earlier, functional groups which lack lone pairs for accepting hydrogen bonds such as $-CH_3$ and $-NH_3^+$ are assigned a single radius, as in the united atom topology. Otherwise, hydrogen atoms are assigned spheres with radii that depend on their effective electrostatic potential-fitting charges. This allows the radius to vary with the electronic charge character of the group and to more accurately represent the interactions between solute and solvent. The CHELPG charges show significant differences for anions and cations compared to their neutral analogs.

References and Notes

(1) Camaioni, D. M.; Dupuis, M.; Bentley, J. *J. Phys. Chem. A* **2003**, *107*, 5778.

(2) Pettitt, B. M. *Theo. Chem. Acc.* **1999**, *103*, 171.

(3) (a) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093. (b) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(4) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(5) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-Lopez, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284.

(6) Cramer, C. J. *Essentials of Computational Chemistry, Theories and Models*; John Wiley & Sons, Ltd.: West Sussex, England, 2002; pp 346−374.

(7) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

(8) Ginovska, B.; Camaioni, D. M.; Dupuis, M. *J. Chem. Phys.* **2008**, *129*, 014506.
(9) (a) Cossi, M.; Mennucci, B.; Pitarch, J.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 833. (b) Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1995**, *16*, 1449.
(10) Chipman, D. M. *Theo. Chem. Acc.* **2002**, *107*, 80.
(11) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744.
(12) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.
(13) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
(14) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493.
(15) Frisch, M. J. ; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery Jr., J. A.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K. ; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian98*, revision A.11.4; Gaussian, Inc.: Pittsburgh, PA, 1998.
(16) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(17) (a) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294. (b) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. (c) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639. (d) Curtiss, L. A.; McGrath, M. P.; Blandeau, J. -P.; Davis, N. E.; Binning, R. C., Jr.; Radom, L. *J. Chem. Phys.* **1995**, *103*, 6104.
(18) (a) Langlet, J.; Claverie, P.; Caillet, J.; Pullman, A. *J. Phys. Chem.* **1988**, *92*, 1617.
(19) Floris, F. M.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784.
(20) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.
(21) Throughout this work we use ΔG^*_s to refer to standard free energy for transfer from gas to solution where the solute and gas concentrations are specified in number density concentration units, that is, mol/L; see : Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(22) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70.
(23) Curutchet, C.; Bidon-Chanal, A.; Soteras, I.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2005**, *109*, 3565.
(24) Experimental value estimated using gas phase basicity, $\Delta G°_b$ = 369.5 kcal/mol: Ramond, T. M.; Blanksby, S. J.; Kato, S.; Bierbaum, V. M.; Davico, G. E.; Schwartz, R. L.; Lineberger, W. C.; Elliso, *J. Phys. Chem. A* **2002**, *106*, 9641.
(25) (a) Fattahi and Kass correlated proton affinity of anions with electron affinity of their corresponding radicals, breaking the compounds down into classes. The HOO^-, in the heterosubstituted oxyanions class, has proton affinity that is 0.8 kcal/mol larger than the regressed value for this class. A smaller proton affinity would make $\Delta G_s°$ (HOO^-) less favorable. (b) Fattahi, A.; Kass, S. R. *J. Org. Chem.* **2004**, *69*, 9176.
(26) Wan, S.; Stote, R. H. *J. Chem. Phys.* **2005**, *121*, 9539.
(27) Tissandier, M. D.; Cowen, K. A.; Fen, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787.
(28) The values of ΔG_s^* for ions are defined based on the most accurate current value of $\Delta G_s^*(H^+)$ as reported by Tissandier et al.[27] See also ref 14.
(29) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634.
(30) Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373*, 411.
(31) Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327.
(32) Camaioni, D. M.; Scherdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795.
(33) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066.
(34) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.
(35) Cramer, C. J. *Essentials of Computational Chemistry, Theories and Models*; John Wiley & Sons, Ltd.: West Sussex, England, 2002; pp 412.
(36) The average error in the G3B3 method is ±1 kcal/mol; see ref 34.
(37) On the basis of $\Delta G_s^* = -9.3$ kcal/mol[38] for hydrazine.
(38) Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schumm, R. H.; Halow, I.; Bailey, S. M.; Churney, K. L.; Nuttal, R. L. *J. Phys. Chem. Ref. Data* **1982**, *11*, Suppl. No. 2.
(39) Hunter, E. P.; Lias, S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413–6563.
(40) (a) Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry*; Wiley Interscience: New York, 1988; p 316. (b) Cardulla, F. *J. Chem. Educ.* **1983**, *60*, 505–508.
(41) Albert, A.; Serjeant, E. P. *The Determination of Ionization Constants*; Chapman and Hall: New York, 1984.
(42) Perdoncin, G.; Scorrano, G. *J. Am. Chem. Soc.* **1977**, *99*, 6983.
(43) From the following thermochemical and physical data for glycine: $\Delta H°_{sub} = 33.0 \pm 1.1$ kcal/mol;[44] $S°$ (c) = 24.74 cal/mol•K;[45] $S°$ (g) = 74.8 cal/mol•K (B3LYP/6-311+G** frequency calculation); aqueous solubility of glycine is 250 g/L (~3.3 M) at 25 °C;[46] activity coefficient at 25 °C is 0.729.[47]
(44) Ngauv, S. N.; Sabbah, R.; Laffitte, M. *Thermochim. Acta* **1977**, *20*, 371.
(45) Hutchens, J. O.; Cole, A. G.; Stout, J. W. *J. Am. Chem. Soc.* **1960**, *82*, 4813.
(46) Beilstein, F. K. *Handbuck der Organischen Chemie*, 4th Ed.; Springer-Verlag: New York, *Vol. 4*, p. 333.
(47) Smith, E. R. B.; Smith, P. K. *J. Biol. Chem.* **1937**, *117*, 209.
(48) (a) Edsall, J. T.; Blanchard, M. H. *J. Am. Chem. Soc.* **1933**, *55*, 2337. (b) Wada, G.; Tamura, E.; Okina, M.; Nakamura, M. *Bull. Chem. Soc. Jpn.* **1982**, *55*, 3064.
(49) Although average is 0.5 kcal/mol for species in this work, the average error for oxo anions and neutral parent molecules was 1 kcal/mol.[1].
(50) Truong, T. N.; Stefanovich, E. V. *J. Chem. Phys.* **1995**, *103*, 3709.
(51) Shaik, S. S. *J. Am. Chem. Soc.* **1984**, *106*, 227.
(52) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049.
(53) Vayner, G.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I. *J. Am. Chem. Soc.* **2004**, *126*, 9054.
(54) Albery, W. J.; Krevoy, M. M. *Adv. Phys. Org. Chem.* **1978**, *16*, 87.
(55) Cossi, M.; Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *297*, 1.
(56) Truong, T. N.; Stefanovich, E. V. *J. Phys. Chem.* **1995**, *99*, 14700.
(57) Sander, R. Henry's Law Constants. In *NIST Chemistry WebBook*; NIST Standard Reference Database Number 69; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, June 2005.
(58) (a) Ruscic, B.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Harding, L. B.; Asher, R. L.; Wagner, A. F. *J. Phys. Chem. A* **2001**, *105*, 1. (b) Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X.; Ng, C.-Y.; Liu, J.; Chen, W.; Schwenke, D. W. *J. Phys. Chem. A* **2002**, *106*, 2727. (c) Poskrebyshev, G. A.; Neta, P.; Huie, R. E. *J. Phys. Chem. A* **2002**, *106*, 11488.
(59) (a) Pliego, J. R.; Riveros, J. M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1622. (b) Pliego, J. R.; Riveros, J. M. *Chem. Phys. Lett.* **2000**, *332*, 597.
(60) O'Sullivan, D. W.; Lee, M.; Noone, B. C.; Heikes, B. G. *J. Phys. Chem.* **1996**, *100*, 3241.
(61) Flowers, B. A.; Szalay, P. G.; Stanton, J. F.; Kállay, M.; Gauss, J.; Császár, A. G. *J. Phys. Chem. A* **2004**, *108*, 3195.

JP804092V