JCTC Journal of Chemical Theory and Computation

SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute−Water Clusters

Casey P. Kelly, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry and Supercomputing Institute, 207 Pleasant St. SE, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received June 28, 2005

Abstract: A new charge model, called Charge Model 4 (CM4), and a new continuum solvent model, called Solvation Model 6 (SM6), are presented. Using a database of aqueous solvation free energies for 273 neutrals, 112 ions, and 31 ion−water clusters, parameter sets for the mPW0 hybrid density functional of Adamo and Barone (Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675) were optimized for use with the following four basis sets: MIDI!6D, 6-31G(d), 6-31+G(d), and 6-31+G(d,p). SM6 separates the observable aqueous solvation free energy into two different components: one arising from long-range bulk electrostatic effects and a second from short-range interactions between the solute and solvent molecules in the first solvation shell. This partition of the observable solvation free energy allows SM6 to effectively model a wide range of solutes. For the 273 neutral solutes in the test set, SM6 achieves an average error of ∼0.50 kcal/mol in the aqueous solvation free energies. For solutes, especially ions, that have highly concentrated regions of charge density, adding an explicit water molecule to the calculation significantly improves the performance of SM6 for predicting solvation free energies. The performance of SM6 was tested against several other continuum models, including SM5.43R and several different implementations of the Polarizable Continuum Model (PCM). For both neutral and ionic solutes, SM6 outperforms all of the models against which it was tested. Also, SM6 is the only model (except for one with an average error 3.4 times larger) that improves when an explicit solvent molecule is added to solutes with concentrated charge densities. Thus, in SM6, unlike the other continuum models tested here, adding one or more explicit solvent molecules to the calculation is an effective strategy for improving the prediction of the aqueous solvation free energies of solutes with strong local solute−solvent interactions. This is important, because local solute−solvent interactions are not specifically accounted for by bulk electrostatics, but modeling these interactions correctly is important for predicting the aqueous solvation free energies of certain solutes. Finally, SM6 retains its accuracy when used in conjunction with the B3LYP and B3PW91 functionals, and in fact the solvation parameters obtained with a given basis set may be used with any good density functional or fraction of Hartree−Fock exchange.

1. Introduction

Continuum solvent models are an attractive alternative to explicit solvent approaches, because they require less computational effort, making them applicable to larger solutes, extensive conformational analysis, and large libraries of compounds. Continuum solvent models have advanced to a point where aqueous solvation free energies of typical neutral organic solutes can usually be predicted accurately to better than 1 kcal/mol. However, the development of methods for accurately predicting aqueous solvation free energies of ionic solutes has been much less successful. In part, this is due to

* Corresponding author e-mail: cramer@chem.umn.edu (C.J.C.) and truhlar@chem.umn.edu (D.G.T).

10.1021/ct050164b CCC: $30.25
Published on Web 10/22/2005

the limited availability of experimental solvation free energies for ionic solutes. Unlike neutral solutes, for which aqueous solvation free energies can be obtained directly from partition coefficients of solutes between the gas phase and dilute aqueous solution, aqueous solvation free energies of ionic solutes must be determined from other experimental observables. Because of this, there is a certain degree of uncertainty associated with making direct comparisons between calculated and experimental aqueous solvation free energies for ions, making the development of a model that is able to treat neutral and ionic solutes at the same level of accuracy a very challenging task. In addition, because of strong electrostatic effects arising from localized solute−solvent interactions, the magnitudes of solvation free energies are much greater for ions than for neutrals, requiring smaller percentage errors for the same absolute accuracy. Because the differential solvation free energy between a given acid/base pair can be used in various thermodynamic cycles to determine pK_a, developing a model that can be used to predict these free energies accurately is a high priority.

A number of different strategies have been used to account for short-range interactions within the framework of continuum solvation theory.[1,2] For example, the SM*x* series of models developed by our co-workers[3−17] and us augments the electrostatic portion of the calculated solvation free energy with an empirical term that accounts for, among other things, deviations of short-range interactions, primarily those in the first solvation shell, from the bulk electrostatic limit. Although this approach has been very successful in predicting solvation free energies of neutral solutes, it remains unclear whether this type of correction to the solvation free energy can be applied to ionic solutes with the same success.

A key issue in predicting the large electrostatic effects involved in solvation of ions is determining the shape of the cavity that is used to define the boundary between the electronic distribution of the solute and the continuum solvent. In all of our recent SM*x* models[7−16] (including the one presented in this article), a single set of radii that are dependent only on the atomic number of the given atom are used to build up the molecular cavity. More elaborate methods for assigning atomic radii depend on the local chemical environment of the atom.[18−23] One such prescription, called the united atom for Hartree−Fock (UAHF) method,[23] assigns radii to atoms based on their hybridization states, what other atoms are bonded to them, and their formal charge. These radii are often used in conjunction with the popular polarizable continuum models (PCMs)[24−34] to predict aqueous solvation free energies. Other methods have been proposed in which the atomic radii depend on partial atomic charge, and several groups have had some success using charge-dependent atomic radii in continuum solvation calculations,[35−41] although the work has been limited to a small number of solutes. Other methods define the solute cavity as the contour on which the solute electronic density is equal to some constant value.[42] For example, Chipman has recently shown[43] that using a value of 0.001 e/a_0^3 for the isodensity contour along with a continuum model[44] results in accurate aqueous solvation free energies for a series of protonated amines. However, Chipman also showed in his work[43] that for a series of oxygen-containing anions, no single common contour value could be used to accurately predict their absolute aqueous solvation free energies. Chipman attributed this finding to the inability of a continuum model alone to account for strong anion−water interactions in the first solvation shell and suggested that better results might be obtained by augmenting continuum solvent calculations with other complementary methods that are especially designed to account for specific short-range interactions.

Adding explicit solvent molecules has been a popular strategy for trying to incorporate the effects of specific solute−solvent interactions into continuum solvent calculations. Often, this involves treating enough solvent molecules classically or quantum mechanically to account for at least the entire first solvation shell around the given solute.[45] Depending on the solute, this may require a large number of explicit solvent molecules, which can lead to a significant increase in the amount of computational effort expended. In addition to this problem, there are several other potential problems associated with treating solvent molecules explicitly. First, for many solutes, there is no easy way to determine the number or orientation of explicit water molecules in the first solvation shell. For example, X-ray diffraction experiments[46] and various theoretical calculations[47−51] lead to average coordination numbers ranging from 6 to 9.3 for the Ca^{2+} ion, suggesting that several different solvation structures exist. Second, even for solutes for which the first solvation shell is well defined, to properly treat even a few solvent molecules explicitly will most likely involve the need to sample over a large number of conformations that are local minima. Finally, introducing explicit solvent molecules will not yield more accurate solvation free energies if the level of theory used to treat the system is not high enough. Because properly treating nonbonded interactions usually requires treatment of electron correlation and the use of fairly large basis sets,[52] any realistic attempt at using solute−water clusters to calculate aqueous solvation free energies requires an accurate treatment of the entire solute−solvent system, which is practical only for small numbers of solvent molecules. Despite these problems, hybrid approaches combining quantal and classical treatments of solvent molecules have had some success in predicting aqueous solvation free energies of ions. For example, Pliego and Riveros showed[53] that for a test set of 17 ions, including several explicit water molecules in the continuum calculation significantly improved the performance of the model. To determine the number of explicit solvent molecules required in the calculation, these workers developed an approach in which the aqueous solvation free energy of the bare solute is minimized with respect to the number of coordinating waters.[54] Besides predicting solvation free energies of ions, this approach has also been used to predict solvatochromic shifts,[55,56] where explicit solute−solvent effects between the electronically excited solute and surrounding solvent molecules can have large effects on both the magnitude and direction of the shift.

In the present paper, we will present a new continuum solvent model called Solvation Model 6 (SM6). This model is similar to our most recently developed previous continuum model, called SM5.43R,[15,16] but improves on it in a number of significant ways. In both of these models, SM6 and

SM5.43R, the aqueous solvation free energy is calculated as a sum of free energies arising from long-range bulk electrostatic effects, which are calculated by a self-consistent reaction field (SCRF) calculation,[12,57,58] and those from nonbulk electrostatic effects, which are calculated using the solvent accessible surface area (SASA)[59,60] of the solute and a set of atomic surface tensions that depend on a set of empirical parameters and the geometry of the solute. SM6 differs from SM5.43R in two important ways. First, SM6 uses an improved charge model, called Charge Model 4 (CM4), for assigning partial atomic charges. CM4 is a new charge model developed as part of the present effort, and it is presented later in the text and in the Supporting Information. Second, SM6 is parametrized with a training set of aqueous solvation free energies that has been improved in three ways: (1) the neutral portion of the training set has been extended to include molecules containing certain functionalities that were not present in the SM5.43R training set, (2) we use a larger and improved set of data for ionic solutes, and (3) aqueous solvation data for various ion−water clusters and the water dimer have been added, and the entire philosophy of the parametrization of Coulomb radii is changed to reflect the use of cluster data in place of bare-ion data for cases where continuum solvent models are expected to be inadequate for bare ions.

SM6 calculations may use any reasonable gas-phase or liquid-phase geometry of the solute to calculate its aqueous solvation free energy. In addition, geometry optimizations in the liquid phase using analytical free-energy gradients can be efficiently carried out.[13] We previously denoted the case for which aqueous solvation free energies were calculated using gas-phase geometries with the suffix "R" and those in which they were calculated using liquid-phase geometries by dropping the "R" suffix (which stands for "rigid"); here we will drop the "R" suffix in all cases and use the standard Pople notation. For example, a single-point SM6 calculation at the MPW25/6-31+G(d,p) level using a gas-phase geometry optimized at the MPW25/MIDI! level of theory would be written as SM6/MPW25/6-31+G(d,p)//MPW25/MIDI!, whereas if the consistent liquid-phase geometry were used, this calculation would be written as SM6/MPW25/6-31+G(d,p). A solvation free energy calculated by SM6/MPW25/6-31+G(d,p) at a gas-phase geometry computed by the same electronic structure level (i.e., MPW25/6-31+G(d,p)) can be denoted SM6/MPW25/6-31+G(d,p)//MPW25/6-31+G(d,p) or for short SM6/MPW25/6-31+G(d,p)//*g*, where //*g* denotes a gas-phase geometry at the same level.

Four new parametrizations of SM6 for the MPW*X* hybrid density functional will be presented, where each parametrization uses a particular basis set. These four basis sets are MIDI!6D[61,62] and Pople's[63] popular 6-31G(d), 6-31+G(d), and 6-31+G(d,p) basis sets. The MPW*X* functional uses Barone and Adamo's[64] modified version of Perdew and Wang's exchange functional,[65] Perdew and Wang's PW91 correlation functional, and a percentage *X* of Hartree−Fock exchange.[16] The parameters presented here can be used with any value of *X*, which is a stability feature pointed out in a previous paper.[16] This is particularly useful, because depending on the problem, it may be advantageous to optimize *X* in the gas phase or in solution. For example, $X = 42.8$ has been optimized for kinetics (resulting in the MPW1K functional[66]), $X = 40.6$ for $Y^- + RY$ nucleophilic substitution reactions (Y = F, Cl; the MPW1N functional[67]), $X = 6$ for conformations of sugars,[68] and $X = 25$ has been suggested for predicting heats of formation[64] (this is the mPW1PW91 functional of Barone and Adamo,[64] which they also call mPW0 and which we also refer to as MPW25, or MPW*X*, with $X = 25$). We chose to base the present parametrizations on the MPW*X* hybrid density functional for two reasons. First, as mentioned above, methods for predicting various gas-phase properties that employ different values of *X* with this functional have already been developed, and it is useful to have a set of solvation parameters that can also be used with any *X*. Second, the MPW*X* functional has been shown to be more accurate than the popular B3LYP[69] and HF[70] methods for predicting energies of reaction and barrier heights.[66,71] Furthermore, two of the parametrizations presented here are for basis sets containing diffuse functions. This is important because diffuse functions are often required for accurate calculations of conformational energies and barrier heights.[72] Thus, the parametrizations based on the 6-31+G(d) and 6-31G+(d,p) basis sets are of special interest, because they can be applied in cases where one wants to use the same level of theory for calculating relative energies in the gas phase and in the aqueous phase.

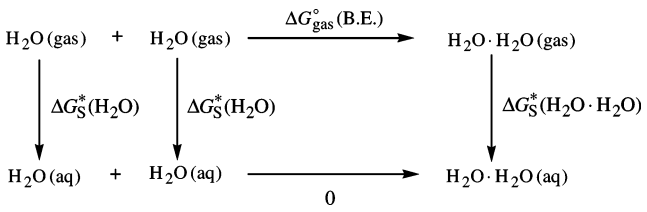
In addition to parametrizing a new charge model and a new aqueous solvent model, the present article has a third goal, namely, to ascertain what effect adding explicit solvent molecules has on the accuracy of continuum solvent models for predicting aqueous solvation free energies. For this, we added a *single* explicit water molecule to some of the solutes in our training set, and we used the resulting solute−solvent cluster to calculate the aqueous solvation free energy. Since the effort in this approach is modest because we are limiting the number of explicit water molecules to one and because the solute−solvent system will be modeled as a single, rigid conformation, the approach is very practical and does not have most of the problems associated with adding several explicit solvent molecules that were outlined above. Furthermore, the comparison between aqueous solvation free energies calculated using bare solutes to those calculated using solute−solvent clusters provides insight into whether continuum solvent models are appropriate for calculating aqueous solvation free energies of solute−water clusters as well as whether the performance of these models can be improved in cases where specific localized solute−solvent interactions are expected to play a large role in determining an aqueous solvation free energy.

Section 2 presents the experimental data used to train and test the new model, which is itself presented in section 3. Section 4 is concerned with parametrization, and section 5 gives the results. Sections 6 and 7 present discussion and conclusions, respectively.

2. Experimental Data

2.1. Standard States. All data and calculations are for 298 K. All experimental and calculated gas-phase free energies are tabulated using an ideal gas at 1 atm as the reference state. Free energies that employ this standard state definition will be denoted by the superscript "°". All experimental and

Scheme 1



calculated solvation free energies are tabulated for an ideal gas at a gas-phase concentration of 1 mol/L dissolving as an ideal solution at a liquid-phase concentration of 1 mol/L. Free energies that employ this standard state definition will be denoted by the superscript "*". The relationship between these two standard states is

$$\Delta G^*_{gas} = \Delta G^\circ_{gas} + \Delta G^{\circ\rightarrow *} \quad (1)$$

and

$$\Delta G^*_S = \Delta G^\circ_S - \Delta G^{\circ\rightarrow *} \quad (2)$$

where[73]

$$\Delta G^{\circ\rightarrow *} = RT\ln(24.46) \quad (3)$$

At 298 K $\Delta G^{\circ\rightarrow *}$ equals 1.89 kcal/mol.

2.2. Neutral Solutes. For neutral solutes, we start with all of the experimental aqueous solvation free energy data from the previously described SM5.43R training set,[15] which includes 257 aqueous solvation free energies for solutes containing at most H, C, N, O, F, P, S, Cl, and/or Br. To these data, we added additional aqueous solvation data for various reasons. Aqueous solvation free energies were added for methylhydrazine and 1,1-dimethylhydrazine in order to test the performance of SM6 without the surface tension previously[7−17,74,75] applied to solutes containing hydrogen atoms in the vicinity of two nitrogen atoms (e.g., hydrazines and hydrazones with a hydrogen attached to one of the individual nitrogen atoms). Data for hydrogen peroxide, methyl peroxide, and ethyl peroxide were added because the SM5.43R training set does not contain any data for peroxides. The SM5.43R training set contains an aqueous solvation free energy for aniline but no other aniline analogues, so we added data for *ortho*-, *meta*-, and *para*-methylaniline as well as *N*-methyl-, *N*-ethyl-, and *N*,*N*-dimethylaniline. We also added 1,2-ethanediamine and 3-aminoaniline because the SM5.43R training set does not contain any data for solutes with more than one amino group in the same solute. Finally, we added urea and benzamide because the SM5.43R training set contains only one solute with urea functionality (1-dimethyl-3-phenylurea) and three solutes with amide functionality (acetamide, Z-*N*-methylacetamide, and *E*-*N*-methylacetamide).

We also examined the accuracy of several experimental solvation free energies that were used in earlier versions of our training sets. In our previous training sets that include hydrazine,[7−17,74,75] we used a value of −9.30 kcal/mol for the aqueous solvation free energy. Here, this value has been replaced by a value of −6.26 kcal/mol, which was obtained using experimental values for the vapor pressure and aqueous solubility.[76,77] For methyl benzoate, the value of −2.22 kcal/mol that was used in our previous training sets[15−17] has been replaced by a value of −3.91 kcal/mol that was also obtained using experimental values for the vapor pressure and aqueous solubility.[76,77]

We also added the aqueous solvation free energy for the water dimer. This free energy can be determined using the free energy cycle shown in Scheme 1, according to

$$\Delta G^*_S(H_2O{\cdot}H_2O) = 2\Delta G^*_S(H_2O) - \Delta G^\circ_{gas}(B.E.) + \Delta G^{\circ\rightarrow *} \quad (4)$$

where $\Delta G^*_S(H_2O)$ is the aqueous solvation free energy of water, and $\Delta G^\circ_{gas}(B.E.)$ is the gas-phase binding free energy, which equals $G^\circ_{gas}(H_2O{\cdot}H_2O) - 2G^\circ_{gas}(H_2O)$. Substituting experimental values of −6.31 kcal/mol for $\Delta G^*_S(H_2O)$ and 3.34 kcal/mol[78] for $\Delta G^\circ_{gas}(B.E.)$ into eq 4 gives −14.06 kcal/mol for the aqueous solvation free energy of the water dimer (at 298 K).

Adding the new data described to the SM5.43R training set and correcting the two values mentioned in the previous paragraph results in a new data set with a total of 273 aqueous solvation free energies for neutral solutes containing at most H, C, N, O, F, P, S, Cl, and/or Br. This will be called the SM6 neutral-solute aqueous free-energy-of-solvation data set. These solvation free energies are listed in Table S1 of the Supporting Information.

2.3. Ionic Solutes. For our previous models that included data for ionic soutes,[3−17,74,75] aqueous solvation free energies were taken from Florián and Warshel[79] and Pearson[80] and then updated based on changes in the accepted absolute aqueous solvation energy of the proton. Based on a careful analysis of the ionic data in the SM5.43R training set (which contains aqueous solvation free energies for 47 ionic solutes), we decided to develop two entirely new data sets of experimental solvation free energies for ionic solutes. The first new data set, which is listed in Tables 1 and 2, contains aqueous solvation free energies for 112 ionic solutes (60 anions and 52 cations). This will be called the SM6 unclustered-ion data set, and it is described in the next two paragraphs and further discussed in the two paragraphs after that.

For the aqueous solvation free energy of the proton, $\Delta G^*_S(H^+)$, we used Zhan and Dixon's value of −264 kcal/mol.[81] For the remaining cations, we used the thermodynamic cycle shown in Scheme 2 along with eq 5

$$\Delta G^*_{aq} = 2.303RT\mathrm{p}K_a \quad (5)$$

where pK_a is the negative common logarithm of the aqueous acid dissociation constant of AH and ΔG^*_{gas} is the same as ΔG^*_{aq} except for the gas phase. Using Scheme 2 then yields the standard-state aqueous solvation free energy of AH^+ as

$$\Delta G^*_S(AH^+) = \Delta G^\circ_{gas}(A) + \Delta G^{\circ\rightarrow *} + \Delta G^*_S(A) + \Delta G^*_S(H^+) - 2.303RT\mathrm{p}K_a(AH^+) \quad (6)$$

where $\Delta G^*_S(A)$ is the aqueous solvation free energy of the neutral species A, $\Delta G^*_S(H^+)$ is the aqueous solvation free energy of the proton, and $\Delta G^\circ_{gas}(A)$ is the gas-phase basicity of A, equal to $G^\circ_{gas}(A) + G^\circ_{gas}(H^+) - G^\circ_{gas}(AH^+)$. The

Table 1. Aqueous Solvation Free Energies (kcal/mol) of Bare Anions[a]

A^-	AH	ΔG°_{gas} (AH)[b]	pK_a (AH)[c]	ΔG^*_S(AH)[d]	$\Delta G^*_S(A^-)$
HC_2^-	acetylene	370.0 ± 1.8	21.7[e]	0.0	−74 ± 3
HCO_2^-	formic acid	338.3 ± 1.5	3.8	−7.0[f]	−78 ± 2
$CH_3CO_2^-$	acetic acid	341.4 ± 2.0	4.8	−6.7	−80 ± 3
$CH_3CH_2CO_2^-$	propanoic acid	340.4 ± 2.0	4.9	−6.5	−78 ± 3
$CH_3(CH_2)_4CO_2^-$	hexanoic acid	339.0 ± 2.1	4.9	−6.2	−76 ± 3
$H_2C{=}CHCO_2^-$	acrylic acid	337.2 ± 2.8	4.3	−6.6[f]	−76 ± 3
$CH_3COCO_2^-$	pyruvic acid	326.5 ± 2.8	2.5	−9.4[f]	−70 ± 3
$C_6H_5CO_2^-$	benzoic acid	333.0 ± 2.0	4.2	−7.9[f]	−73 ± 3
CH_3O^-	methanol	375.0 ± 0.6	15.5	−5.1	−97 ± 2
$C_2H_5O^-$	ethanol	371.3 ± 1.1	15.9	−5.0	−93 ± 2
$CH_3CH_2CH_2O^-$	1-propanol	369.4 ± 1.4	16.1	−4.8	−90 ± 2
$(CH_3)_2CHO^-$	2-propanol	368.8 ± 1.1	17.1	−4.8	−88 ± 2
$CH_3CH_2CHOCH_3^-$	2-butanol	367.5 ± 2.0	17.6	−4.7[f]	−86 ± 3
$C(CH_3)_3O^-$	*t*-butanol	367.9 ± 1.1	19.2	−4.5	−84 ± 2
$H_2C{=}CHCH_2O^-$	allyl alcohol	366.6 ± 2.8	15.5	−5.1	−88 ± 3
$C_6H_5CH_2O^-$	benzyl alcohol	363.4 ± 2.0	15.4	−6.6[f]	−87 ± 3
$CH_3OCH_2CH_2O^-$	2-methoxyethanol	366.8 ± 2.0	14.8	−6.8	−91 ± 3
$C_6H_5O^-$	phenol	342.9 ± 1.3	10.0	−6.6	−74 ± 2
o-$CH_3C_6H_4O^-$	2-methylphenol	342.4 ± 2.0	10.3	−5.9	−72 ± 3
m-$CH_3C_6H_4O^-$	3-methylphenol	343.3 ± 2.0	10.1	−5.5	−73 ± 3
p-$CH_3C_6H_4O^-$	4-methylphenol	343.8 ± 2.0	10.3	−6.1	−74 ± 3
$CH_2OHCH_2O^-$	1,2-ethanediol	360.9 ± 2.0	15.4	−9.3	−87 ± 3
m-$HOC_6H_4O^-$	3-hydroxyphenol	339.1 ± 2.0	9.3[g]	−11.4[f]	−76 ± 3
p-$HOC_6H_4O^-$	4-hydroxyphenol	343.1 ± 2.0	9.9[g]	−11.9[f]	−80 ± 3
CH_3OO^-	methyl hydroperoxide	367.6 ± 0.7	11.5	−5.3[h]	−95 ± 2
$CH_3CH_2OO^-$	ethyl hydroperoxide	363.9 ± 2.0	11.8	−5.3[h]	−91 ± 3
$CH_2(O)CH^-$	acetaldehyde	359.4 ± 2.0	16.5	−3.5[f]	−78 ± 3
$CH_3C(O)CH_2^-$	acetone	362.2 ± 2.0	19.0	−3.9	−78 ± 3
$CH_3CH_2C(O)CHCH_3^-$	3-pentanone	361.4 ± 2.0	19.9	−3.3[f]	−76 ± 3
CH_2CN^-	acetonitrile	366.0 ± 2.0	25.0	−3.9	−74 ± 3
$NCNH^-$	cyanamide	344.0 ± 2.0	10.3[g]	−6.2[f]	−74 ± 3
$C_6H_5NH^-$	aniline	359.1 ± 2.0	27.7	−5.5	−65 ± 3
$(C_6H_5)_2N^-$	diphenylamine	343.8 ± 2.0	22.4	−5.3[f]	−56 ± 3
CN^-	hydrogen cyanide	343.7 ± 0.3	9.2[i]	−3.1[f]	−72 ± 2
o-$NO_2C_6H_4O^-$	2-nitrophenol	329.5 ± 2.0	7.2	−4.5[f]	−62 ± 3
m-$NO_2C_6H_4O^-$	3-nitrophenol	327.6 ± 2.0	8.4	−9.6[f]	−64 ± 3
p-$NO_2C_6H_4O^-$	4-nitrophenol	320.9 ± 2.0	7.1	−10.6[f]	−60 ± 3
$CH_2NO_2^-$	nitromethane	350.4 ± 2.0	10.2	−4.0[f]	−78 ± 3
p-$NO_2C_6H_5NH^-$	4-nitroaniline	336.2 ± 2.0	18.2	−9.9[f]	−59 ± 3
CH_3CONH^-	acetamide	355.0 ± 2.0	15.1	−9.7	−82 ± 3
CH_3S^-	methanethiol	350.6 ± 2.0	10.3	−1.2	−76 ± 3
$CH_3CH_2S^-$	ethanethiol	348.9 ± 2.0	10.6	−1.3	−74 ± 3
$C_3H_7S^-$	1-propanethiol	347.9 ± 2.0	10.7	−1.1	−72 ± 3
$C_6H_5S^-$	thiophenol	333.8 ± 2.0	6.6	−2.6	−65 ± 3
$CH_3S(O)CH_2^-$	dimethyl sulfoxide	366.8 ± 2.0	33.0	−9.8[f]	−70 ± 3
CCl_3^-	chloroform	349.7 ± 2.0	24.0	−1.1	−56 ± 3
$CF_3CO_2^-$	trifluoroacetic acid	316.7 ± 2.0	0.5	−7.3[f]	−61 ± 3
$CH_2ClCO_2^-$	chloroacetic acid	328.9 ± 2.0	2.9	−8.7[f]	−72 ± 3
$CHCl_2CO_2^-$	dichloroacetic acid	321.5 ± 2.0	1.4	−6.6[f]	−64 ± 3
$CF_3CH_2O^-$	2,2,2-trifluoroethanol	354.1 ± 2.0	12.4	−4.3	−80 ± 3
$CH(CF_3)_2O^-$	1,1,1,3,3,3-hexafluoropropan-2-ol	338.4 ± 2.0	9.3	−3.8	−67 ± 3
$ClC_6H_4O^-$	2-chlorophenol	337.1 ± 2.0	8.5	−4.5[f]	−68 ± 3
$ClC_6H_4O^-$	4-chlorophenol	336.5 ± 2.0	9.4	−6.2[f]	−68 ± 3
HO^-	water	383.7 ± 0.2	15.7	−6.3	−107 ± 2
HO_2^-	hydrogen peroxide	368.6 ± 0.6	11.7	−8.6[h]	−99 ± 2
O_2^-	hydroperoxyl radical	346.7 ± 0.8	4.7	−7.0[j]	−85 ± 2
HS^-	hydrogen sulfide	344.9 ± 1.2	7.0	−0.7	−74 ± 2
F^-	hydrofluoric acid				−102 ± 2[k]
Cl^-	hydrochloric acid				−73 ± 2[k]
Br^-	hydrobromic acid				−66 ± 2[k]

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] Gas-phase basicities taken from ref 83. [c] From ref 87, unless otherwise noted. [d] From the current data set unless otherwise noted. [e] Reference 89. [f] Reference 76. [g] Reference 91. [h] Reference 86. [i] Reference 92. [j] Reference 85. [k] Reference 82.

Table 2. Aqueous Solvation Free Energies of Bare Cations[a]

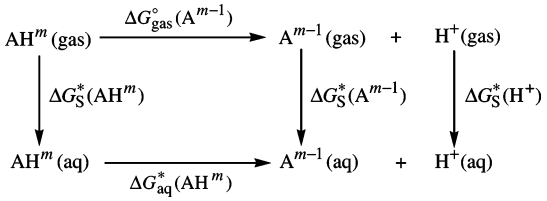
AH^+	AH	ΔG°_{gas} (AH^+)[b]	pK_a (AH^+) [c]	ΔG^*_S(AH)[d]	ΔG^*_S(AH^+)
$CH_3OH_2^+$	methanol	173.2 ± 2.0	−2.1	−5.1	−91 ± 3
$CH_3CH_2OH_2^+$	ethanol	178.0 ± 2.0	−1.9	−5.0	−86 ± 3
$(CH_3)_2OH^+$	dimethyl ether	182.7 ± 2.0	−2.5	−1.8	−78 ± 3
$(C_2H_5)_2OH^+$	diethyl ether	191.0 ± 2.0	−2.4	−1.8	−70 ± 3
$CH_3C(OH)CH_3^+$	acetone	186.9 ± 2.0	−2.9	−3.9	−75 ± 3
$CH_3COHC_6H_5^+$	acetophenone	198.2 ± 2.0	−4.3	−4.6	−63 ± 3
$CH_3NH_3^+$	methylamine	206.6 ± 2.0	10.6	−4.6	−74 ± 3
$CH_3(CH_2)_2NH_3^+$	*n*-propylamine	211.3 ± 2.0	10.6	−4.4	−70 ± 3
$(CH_3)_2CHNH_3^+$	isopropylamine	212.5 ± 2.0	10.6	−3.7[e]	−68 ± 3
$C(CH_3)_3NH_3^+$	*t*-butylamine	215.1 ± 2.0	10.7	−3.9[e]	−65 ± 3
c-$C_6H_{11}NH_3^+$	cyclohexanamine	215.0 ± 2.0	10.7	−5.1[e]	−67 ± 3
$H_2C{=}CHCH_2NH_3^+$	allylamine	209.2 ± 2.0	9.5	−4.3[e]	−70 ± 3
$(CH_3)_2NH_2^+$	dimethylamine	214.3 ± 2.0	10.7	−4.3	−67 ± 3
$(C_2H_5)_2NH_2^+$	diethylamine	219.7 ± 2.0	11.0	−4.1	−62 ± 3
(n-$C_3H_7)_2NH_2^+$	di-*n*-propylamine	222.1 ± 2.0	11.0	−3.7	−59 ± 3
$(H_2C{=}CHCH_2)_2NH_2^+$	diallylamine	219.0 ± 2.0	9.3	−4.0[e]	−60 ± 3
$(CH_3)_3NH^+$	trimethylamine	219.4 ± 2.0	9.8	−3.2	−59 ± 3
$(C_2H_5)_3NH^+$	triethylamine	227.0 ± 2.0	10.8	−3.0[e]	−53 ± 3
(n-$C_3H_7)_3NH^+$	tri-*n*-propylamine	229.5 ± 2.0	10.3	−2.5[e]	−49 ± 3
$C_6H_5NH_3^+$	aniline	203.3 ± 2.0	4.6	−5.5	−70 ± 3
o-$CH_3C_6H_4NH_3^+$	2-methylaniline	205.3 ± 2.0	4.5	−5.6[e]	−68 ± 3
m-$CH_3C_6H_4NH_3^+$	3-methylaniline	206.5 ± 2.0	4.7	−5.7[e]	−68 ± 3
p-$CH_3C_6H_4NH_3^+$	4-methylaniline	206.7 ± 2.0	5.1	−5.6[e]	−68 ± 3
m-$NH_2C_6H_4NH_3^+$	3-aminoaniline	214.9 ± 2.0	5.0	−9.9[e]	−64 ± 3
$C_6H_5NH_2CH_3^+$	*N*-methylaniline	212.7 ± 2.0	4.9[f]	−4.7[e]	−61 ± 3
$C_6H_5NH_2CH_2CH_3^+$	*N*-ethylaniline	213.4 ± 2.0	5.1[f]	−4.6[e]	−60 ± 3
$C_6H_5NH(CH_3)_2^+$	*N*,*N*-dimethylaniline	217.3 ± 2.0	5.1	−3.6[e]	−55 ± 3
p-$CH_3C_6H_4NH(CH_3)_2^+$	4-methyl-*N*,*N*-dimethylaniline	219.4 ± 2.0	5.6	−3.7[e]	−54 ± 3
$C_6H_5NH(CH_2CH_3)_2^+$	*N*,*N*-diethylaniline	221.8 ± 2.0	6.6	−2.9[e]	−52 ± 3
$C_{10}H_7NH_3^+$	1-aminonaphthalene	209.2 ± 2.0	3.9	−7.3[e]	−66 ± 3
$C_2H_4NH_2^+$	aziridine	208.5 ± 2.0	8.0	−4.5[e]	−69 ± 3
$C_3H_6NH_2^+$	azetidine	217.2 ± 2.0	11.3	−5.6	−66 ± 3
$C_4H_8NH_2^+$	pyrrolidine	218.8 ± 2.0	11.3	−5.5	−64 ± 3
$C_5H_{10}NH_2^+$	piperidine	220.0 ± 2.0	11.1	−5.1	−62 ± 3
$C_6H_{12}NH_2^+$	azacycloheptane	220.7 ± 2.0	11.1	−4.9[e]	−61 ± 3
$C_4H_5NH^+$	pyrrole	201.7 ± 2.0	−3.8	−4.3[e]	−60 ± 3
$C_5H_5NH^+$	pyridine	214.7 ± 2.0	5.2	−4.7	−59 ± 3
$C_9H_7NH^+$	quinoline	220.2 ± 2.0	4.8	−5.7[e]	−54 ± 3
$C_4H_8NHNH_2^+$	piperazine	218.6 ± 2.0	9.7	−7.4	−64 ± 3
CH_3CNH^+	acetonitrile	179.0 ± 2.0	−10.0[g]	−3.9	−73 ± 3
p-$CH_3OC_6H_4NH_3^+$	4-methoxyaniline	207.6 ± 2.0	5.3	−7.6[e]	−69 ± 3
p-$NO_2C_6H_4NH_3^+$	4-nitroaniline	199.4 ± 2.0	1.0	−9.9[e]	−74 ± 3
$C_4H_8ONH_2^+$	morpholine	213.0 ± 2.0	8.4	−7.2	−68 ± 3
$CH_3COHNH_2^+$	acetamide	199.0 ± 2.0	−0.6	−9.7	−72 ± 3
$C_6H_5COHNH_2^+$	benzamide	205.8 ± 2.0	−1.4	−10.9[e]	−65 ± 3
$(CH_3)_2SH^+$	dimethyl sulfide	191.5 ± 2.0	−7.0	−1.5	−63 ± 3
$(CH_3)_2SOH^+$	dimethyl sulfoxide	204.0 ± 2.0	−1.5	−9.8[e]	−66 ± 3
m-$ClC_6H_4NH_3^+$	3-chloroaniline	199.9 ± 2.0	3.5	−5.8[e]	−73 ± 3
p-$ClC_6H_4NH_3^+$	4-chloroaniline	201.2 ± 2.0	4.0	−5.9[e]	−72 ± 3
NH_4^+	ammonia	195.7 ± 2.0	9.3[h]	−4.3	−83 ± 3
$HNNH_2^+$	hydrazine	196.6 ± 2.0	8.1	−6.3[e]	−83 ± 3
H_3O^+	water	157.7 ± 0.7	−1.7	−6.3	−108 ± 2

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] Gas-phase basicities from ref 84. [c] From ref 87, unless otherwise noted. [d] From the current data set, unless otherwise noted. [e] Reference 76. [f] Reference 91. [g] Reference 88. [h] Reference 92.

experimental aqueous solvation free energies of F^-, Cl^-, and Br^- were taken from Tissandier et al.[82] and adjusted for a change in the standard state and the value used here for $\Delta G^*_S(H^+)$. For the remaining anions, we used the free energy cycle shown in Scheme 2 and eq 5 to determine the aqueous solvation free energy according to

$$\Delta G^*_S(A^-) = -\Delta G^\circ_{gas}(A^-) - \Delta G^{\circ\rightarrow *} + \Delta G^*_S(AH) - \Delta G^*_S(H^+) + 2.303RT\mathrm{p}K_a(AH) \quad (7)$$

Scheme 2



where $\Delta G^{\circ}_{gas}(A^-)$ is the gas-phase basicity of A^-, equal to $G^{\circ}_{gas}(A^-) + G^{\circ}_{gas}(H^+) - G^{\circ}_{gas}(AH)$.

Experimental gas-phase basicities of anions and acidities of neutral species were taken from the National Institute of Standards and Technology (NIST) database;[83] experimental gas-phase basicities of neutral species were taken from the most recent compilation of Lias et al.[84] For neutral species, experimental aqueous solvation free energies were taken from the data set described in section 2.2 and from several additional sources.[76,85,86] A large part of the experimental aqueous pK_a data used here was taken from the compilation of Stewart;[87] pK_a data not available in this compilation were taken from several additional sources.[88−92]

We note that gas-phase free energies and aqueous solvation energies for neutral solutes can be calculated fairly accurately, and several other compilations of aqueous solvation free energies of ions[79,80,93] use theoretical data for these quantities. Although we did not use theoretical values in deriving any of the free energies shown in Tables 1 and 2, future extensions of the present database could include aqueous solvation energies determined from calculated data.

The aqueous solvation free energies reported here can be compared to the recent compilation of Pliego et al.,[93] who used the same thermodynamic cycle, along with Tissandier et al.'s[82] value of −266 kcal/mol for $\Delta G^*_S(H^+)$, in determining the aqueous solvation free energies for 56 ions. (Note that the above value of −266 kcal/mol for $\Delta G^*_S(H^+)$ reflects the standard-state correction required in order to adjust Tissandier et al.'s reported value of −264 kcal/mol,[82] which is for a gas-phase standard state of 1 atm combined with an aqueous phase standard state of 1 mol/L, to a standard state that uses a concentration of 1 mol/L in both the gas and aqueous phases. Also note that Tissandier's reported value of −264 kcal/mol has sometimes been misinterpreted as corresponding to a standard state of 1 mol/L in both the gas and aqueous phases.[15,16,94−96]) Making an adjustment to account for the difference between the value used for $\Delta G^*_S(H^+)$ here and the value used in Pliego and Riveros' work brings the two sets of data into very good agreement with one another.

2.4. Water−Solute Clusters. We also compiled another data set, to be called the selectively clustered-ion data set, in which 31 bare ions in the unclustered-ion data set were replaced by the ion−water clusters that are listed (along with the water dimer) in Table 3. Using the free energy cycle shown in Scheme 3, the aqueous solvation free energies of these 31 solute−water clusters were determined according to

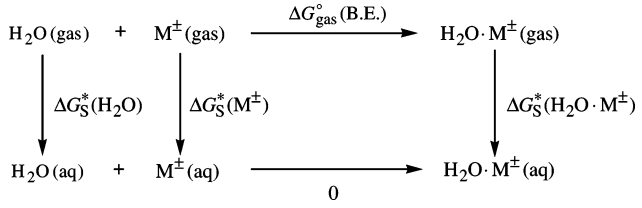
$$\Delta G^*_S(H_2O{\cdot}M^{\pm}) = \Delta G^*_S(H_2O) + \Delta G^*_S(M^{\pm}) - \Delta G^{\circ}_{gas}(\text{B.E.}) + \Delta G^{\circ\rightarrow *} \quad (8)$$

Table 3. Aqueous Solvation Free Energies for Solute−Water Clusters[a]

A·H$_2$O[b]	ΔG°_{gas}(B.E.)[c]	ΔG^*_S(A)[d]	ΔG^*_S(A·H$_2$O)
$H_2O(H_2O)$	3.34 ± 0.50[e]	−6.31 ± 0.20	−14.06 ± 0.57
$CH_3OH_2(H_2O)^+$	−18.5 [f]	−91.1 ± 2.8	−77 ± 3
$CH_3CH_2OH_2(H_2O)^+$	−16.8[f]	−86.5 ± 2.8	−74 ± 3
$(CH_3)_2OH(H_2O)^+$	−15.4[f]	−77.8 ± 2.8	−67 ± 3
$(C_2H_5)_2OH(H_2O)^+$	−11.4 ± 2.0[g]	−69.6 ± 2.8	−62 ± 3
$CH_3C(OH)CH_3(H_2O)^+$	−12.8[f]	−75.1 ± 2.8	−67 ± 3
$CH_3COHC_6H_5(H_2O)^+$	−10.8[f]	−62.6 ± 2.8	−56 ± 3
$NH_4(H_2O)^+$	−12.6[f]	−83.3 ± 2.8	−75 ± 3
$H_3O(H_2O)^+$	−27.0 ± 2.0[g]	−108.4 ± 2.0	−86 ± 3
$C_2H(H_2O)^-$	−10.6 ± 1.0[f]	−78.4 ± 2.6	−72 ± 3
$CN(H_2O)^-$	−8.3 ± 0.7[f]	−72.2 ± 1.9	−68 ± 2
$CH_3O(H_2O)^-$	−17.00 ± 0.30[f]	−96.9 ± 2.0	−84 ± 2
$C_2H_5O(H_2O)^-$	−14.2 ± 2.0[g]	−92.6 ± 2.2	−83 ± 3
$CH_3CH_2CH_2O(H_2O)^-$	−14.6 ± 2.0[g]	−90.2 ± 2.4	−80 ± 3
$(CH_3)_2CHO(H_2O)^-$	−12.3 ± 2.0[g]	−88.2 ± 2.2	−80 ± 3
$CH_3CH_2CHOCH_3(H_2O)^-$	−9.9 ± 2.0[g]	−86.1 ± 2.8	−81 ± 3
$C(CH_3)_3O(H_2O)^-$	−12.2 ± 2.0[g]	−84.2 ± 2.2	−76 ± 3
$H_2C{=}CHCH_2O(H_2O)^-$	−13.5 ± 2.0[g]	−88.5 ± 3.4	−79 ± 4
$C_6H_5CH_2O(H_2O)^-$	−11.6 ± 2.0[g]	−87.0 ± 2.8	−80 ± 3
$CH_3OCH_2CH_2O(H_2O)^-$	−13.5 ± 2.0[g]	−91.4 ± 2.8	−82 ± 3
$CH_2OHCH_2O(H_2O)^-$	−14.0 ± 2.0[g]	−87.2 ± 2.8	−78 ± 3
$CF_3CH_2O(H_2O)^-$	−11.6 ± 2.0[g]	−79.5 ± 2.8	−72 ± 3
$CH(CF_3)_2O(H_2O)^-$	−6.0 ± 2.0[g]	−67.4 ± 2.8	−66 ± 3
$CH_3OO(H_2O)^-$	−14.6 ± 2.0[g]	−95.2 ± 2.0	−85 ± 3
$CH_3CH_2OO(H_2O)^-$	−14.1 ± 2.0[g]	−91.1 ± 2.8	−91 ± 3
$HO(H_2O)^-$	−19.8 ± 1.4[f]	−106.6 ± 1.9	−91 ± 2
$HO_2(H_2O)^-$	−17.0 ± 2.0[g]	−99.2 ± 2.0	−87 ± 3
$O_2(H_2O)^-$	−12.1 ± 2.0[f]	−85.2 ± 2.1	−78 ± 3
$HS(H_2O)^-$	−8.6 ± 2.0[f]	−74.0 ± 2.3	−70 ± 3
$F(H_2O)^-$	−12.5 ± 1.6[f]	−102.5 ± 1.9	−94 ± 3
$Cl(H_2O)^-$	−9.0 ± 4.0[f]	−72.7 ± 1.9	−68 ± 4
$Br(H_2O)^-$	−7.1 ± 2.0[f]	−66.3 ± 1.9	−64 ± 3

[a] Aqueous solvation free energies are for a temperature of 298 K. [b] B97-1/MG3S optimized geometries. [c] Water−solute binding free energies. [d] Aqueous solvation free energy of the bare solute. [e] Experimental value, taken from ref 78. [f] Experimental value, taken from ref 100. [g] Theoretical value, calculated at the B97-1/MG3S level of theory.

Scheme 3



In the above equation, aqueous solvation free energies of the unclustered ions $\Delta G^*_S(M^{\pm})$ were taken from Tables 1 or 2. When available, experimental values for the gas-phase binding energies were used in the above equation. When experimental values were not available, we calculated them at the B97-1[97]/MG3S[98] level of theory, which has recently been shown[52] to perform well for nonbonded interactions in the gas phase.

2.5. Uncertainty of Experimental Data. We have previously estimated an average uncertainty of 0.2 kcal/mol for aqueous solvation free energies of neutral solutes in our data sets.[15] The uncertainties in the aqueous solvation free energies for ionic solutes are significantly greater due in part to their

large magnitudes but also due to uncertainties associated with each of the experimental quantities (except pK_a, as discussed below) used to determine them. The uncertainties in aqueous solvation free energies for ionic solutes are reported here as the root-sum-of-squares combination of each of the uncertainties associated with the individual experimental measurements used to determine them.

All of the gas-phase basicities of anions were taken from the NIST tables.[83] In several cases, more than one experimental measurement was available for a single molecule, in which case the average value was used. For molecules where more than one experimental measurement was available, we calculated the standard deviation from the mean and compared this value to the smallest value for the uncertainty associated with any of the individual measurements. The uncertainty reported here is the larger of these two values. For the gas-phase basicities of neutrals, the only molecule used in this work for which an absolute experimental value has been reported is water; this measurement has an experimental uncertainty of 0.7 kcal/mol. For the remaining molecules considered here, the experimental gas-phase basicities are relative values that have been obtained from bracketing experiments. Following Hunter and Lias[84] we have assigned an uncertainty of 2.0 kcal/mol to data obtained by bracketing.

An additional contribution to the overall uncertainty arises from the value used for the aqueous solvation free energy of the proton $\Delta G^*_S(H^+)$. We use Zhan and Dixon's value of -264 kcal/mol[81] for $\Delta G^*_S(H^+)$, to which we assign an uncertainty of 2 kcal/mol.

Experimental pK_a values that fall in the range 0−14 can be measured quite accurately, and thus their uncertainty is not included in calculating the overall uncertainties in the aqueous solvation free energies of ions. Experimental pK_a values that fall outside the range 0−14 are somewhat more uncertain, especially those pK_a values that are 2 pK_a units or more outside this range.[99] Despite this, we feel that the relative uncertainties associated with the experimental pK_a values considered in this work that do fall outside this range are small in comparison to the uncertainties associated with the other experimental data, and thus we will not take them into account.

Finally, for the solute−water clusters, an additional source of uncertainty from the value used for the gas-phase binding free energy ΔG°_{gas}(B.E.) must be considered. For the anion−water clusters, all experimental values for ΔG°_{gas}(B.E.) were taken from the NIST tables,[100] in which the uncertainty of each measurement is reported. The uncertainty associated with the experimental ΔG°_{gas}(B.E.) values of the cations, which were also taken the NIST tables, is negligible. For theoretical ΔG°_{gas}(B.E.) values calculated at the B97-1/MG3S level of theory, we have estimated an uncertainty of 2.0 kcal/mol.

Based on the error analysis presented above, we estimate that the uncertainty in the aqueous solvation free energy of a typical ionic solute is approximately 3 kcal/mol, which is lower than our previously estimated[15] uncertainty of 4−5 kcal/mol.

3. Definition of SM6

In the SM*x* models, the solvation free energy ΔG^*_S is partitioned according to

$$\Delta G^*_S = \Delta E_{elec} + \Delta E_{relax} + \Delta G^*_{conc} + G_P + G_{CDS} \quad (9)$$

where ΔE_{elec} is the change in the solute's internal electronic energy in moving from the gas phase to the liquid phase at the same geometry, ΔE_{relax} is the change in the solute's internal energy due to changes in the geometry accompanying the solvation process, and ΔG^*_{conc} accounts for the concentration change between the gas-phase and the liquid-phase standard states. Following the notation used in our previous models, we will refer to the sums $\Delta E_{elec} + G_P$ and $\Delta E_{relax} + \Delta E_{elec} + G_P$ as ΔG_{EP} and ΔG_{ENP}, respectively. Since we use the same concentrations (1 mol/L) in both phases, ΔG^*_{conc} is zero.[73,101] Also, all calculations reported here are based on gas-phase geometries (although the present model can be used to optimize geometries in the liquid phase[13]), so ΔE_{relax} is assumed to be zero.

The ΔG_{EP} contribution to the total solvation free energy is calculated from a self-consistent molecular orbital calculation,[12,57,58] where the generalized Born approximation[102−106] is used to calculate the polarization contribution to the total free energy according to

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k \gamma_{kk'} q_{k'} \quad (10)$$

In the above equation, the summation goes over all atoms k in the solute, ϵ is the dielectric constant of the solvent, q_k is the partial atomic charge of atom k, and $\gamma_{kk'}$ is a Coulomb integral involving atoms k and k'. For water, we use $\epsilon = 78.3$.[107]

The partial atomic charges are obtained from Charge Model 4 (CM4). This new charge model is similar in most ways (except one that is described below) to our most recent previous charge model, CM3.[108−112] In particular, CM4 empirically maps atomic charges obtained from a Löwdin population analysis (LPA)[113−116] or a redistributed Löwdin population analysis (RLPA).[117] The Supporting Information provides a detailed description of CM4, although we note here that an important difference between CM3 and CM4 is the performance of these two models for hydrocarbons and molecules containing aliphatic functional groups. For CM3, the parameter that is used to map Löwdin or redistributed Löwdin C−H bond dipoles was optimized[108,109,112,118] by requiring the average CM3 charge on H in benzene and ethylene to be 0.11, a value that had been justified in a previous paper.[118] More recently, careful analysis of partial atomic charges calculated using CM3 revealed that in some cases, CM3 yielded C−H bonds that were too polar, and this, in turn, had a negative impact on the performance of our solvation models for some solutes as well as the performance of other methods that use CM3 partial atomic charges. Because of this deficiency, we developed a different procedure for optimizing the C−H parameter. In particular, this parameter was optimized by minimizing the error between calculated partial atomic charges and those partial atomic charges used in Jorgenson et al.'s OPLS force field[119]

for a series of 19 hydrocarbons. Once this parameter was optimized it was fixed, and the remaining parameters were optimized in a fashion completely analogous to that used for CM3,[108,109,112,118] resulting in a new charge model called CM4. Details of the CM4 parametrization are given in the Supporting Information. It is important to note that CM4 retains all of the qualities of CM3, except that the new model gives more reasonable partial atomic charges for hydrocarbons and molecules containing aliphatic functional groups, which is important for accurately modeling hydrophobic effects.

The Coulomb integrals $\gamma_{kk'}$ are calculated according to

$$\gamma_{kk'} = [R_{kk'}^2 + \alpha_k\alpha_{k'} \exp(-R_{kk'}^2/d\alpha_k\alpha_{k'})]^{-1/2} \qquad (11)$$

where $R_{kk'}$ is the distance between atoms k and k' and α_k is the effective Born radius of atom k, which is described below. When $k = k'$, eqs 10 and 11 lead to Born's equation[120] for the polarization free energy of a monatomic ion, whereas at large $R_{kk'}$ Coulomb's law for the interaction energy of two point charges in a dielectric continuum is recovered. A d value of 4 was originally proposed by Still et al.[105] because for intermediate values of $R_{kk'}$, it gives polarization free energies that are close to those predicted using the classical equation for a dipolar sphere embedded in a dielectric continuum. Because modeling most solutes as a dipolar sphere is itself a drastic approximation, and because Jayaram et al.[121] demonstrated for a test set of dicarboxylic acids that using different values of d in eq 11 led to improvements in calculated pK_a shifts, we will treat d as a parameter that can be adjusted.

The following equation[122] is used to calculate the Born radius

$$\alpha_k = \left(\frac{1}{R'} + \int_{\rho_{Z_k}}^{R'} \frac{A_k(\mathbf{R},r,\{\rho_{Z'}\})}{4\pi r^4}\mathrm{d}r\right)^{-1} \qquad (12)$$

where R' is the radius of the sphere centered on atom k that completely engulfs all other spheres centered on the other atoms of the solute, and $A_k(R,r,\{\rho_Z\})$ is the exposed area[122] of a sphere of radius r that is centered on atom k. This area depends on the geometry of the solute, **R**, and the radii of the spheres centered on all the other atoms in the solute. The radii of these spheres are given by a set of intrinsic Coulomb radii ρ_{Z_k} that depend on the atomic number Z_k of the atom k. Other methods that have recently been proposed for determining the Born radius include one by Onufriev et al.[123] that takes into account interior regions of the solute inaccessible by solvent molecules (which can have an effect on the calculated solvation free energies of macromolecules[124,125]) and one by Zhang et al.[126] that assigns Born radii based on atom types. We recently tested[16] Onufriev et al.'s method for calculating Born radii during the development of SM5.43R and found that it significantly worsened the performance of that model for predicting the aqueous solvation free energies of ionic solutes. Therefore, we will not test this method here.

The G_{CDS} term is calculated according to

$$G_{CDS} = \sum_k \sigma_k A_k(\mathbf{R},\{R_{Z_k} + r_s\}) \qquad (13)$$

where σ_k is the atomic surface tension of atom k, A_k is the solvent-accessible surface area (SASA)[59,60]of atom k, which depends on the geometry, **R**, atomic radius R_{Z_k}, and solvent radius r_s, which is added to each of the atomic radii. Adding a nonzero value for solvent radius to the atomic radii defines the SASA of a given solute. Note that the atomic radii in eq 13 are not constrained to be equal to the intrinsic Coulomb radii ρ_{Z_k} of eq 12. Although the same values for the atomic radii might have been used in eqs 12 and 13, our experience has shown that the overall performance of our models is relatively insensitive to the values used for the atomic radii in eq 13 (but not eq 12). Therefore, in eq 13 we use Bondi's van der Waals radii.[127] For the solvent radius, we use a value of 0.4 Å, which has been justified in earlier work. [15] The atomic surface tensions σ_k are given by

$$\sigma_k = \sum_k \tilde{\sigma}_{Z_k} + \sum_{k,k'} \tilde{\sigma}_{Z_kZ_{k'}} T_k(\{Z_{k'},R_{kk'}\}) \qquad (14)$$

where $\tilde{\sigma}_Z$ is an atomic-number-specific parameter, $\tilde{\sigma}_{ZZ'}$ is a parameter that depends on the atomic numbers of atoms k and k', and $T_k(\{Z_{k'},R_{k,k'}\})$ is a geometry-dependent switching function called a cutoff tanh or COT. The general form for this function is

$$T(R_{kk'}|\bar{R}_{ZZ'},\Delta R) = \begin{cases} \exp\left[-\left(\dfrac{\Delta R}{\Delta R - R_{kk'} + \bar{R}_{ZZ'}}\right)\right] & R_{kk'} \le \bar{R}_{ZZ'} + \Delta R \\ 0 & \text{otherwise} \end{cases} \qquad (15)$$

where $\bar{R}_{ZZ'}$ is the midpoint of the switch, and ΔR determines the range over which the function switches. These values, which are listed in Table 4, have been assigned based on an analysis of distances between certain atoms for molecules in our database. The COT function has the property that it vanishes identically for all $R_{kk'}$ greater than $\bar{R} + \Delta R$ but is continuous and has an infinite number of continuous derivatives for all $R_{kk'}$. These properties are important because they allow the model to optimize geometries efficiently in the liquid phase.[13] The functional forms that we use for the

Table 4. $\bar{R}_{ZZ'}$ and ΔR Values (Å)

Z	Z'	$\bar{R}_{ZZ'}$
H	C, N, or O	1.55
H	S	2.14
C	C, N	1.84
C	C	1.27[a]
C	O	1.33
C	F or P	2.20
C	Cl	2.10
C	Br	2.30
N	N	1.85
N	O	1.50
O	O	1.80
O	P	2.10
S	S	2.75
S	P	2.50
W		0.30
$W_{CC(2)}$		0.07
W_{CO}		0.10
W_{NC}		0.065

[a] $\bar{R}_{CC}^{(2)}$.

atomic surface tensions are given below, where we write Z for Z_k to simplify the notation:

$$\sigma_Z|_{Z=1} = \tilde{\sigma}_Z + \sum_{\substack{k' \\ Z_{k'}=6,7,8,16}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (16)$$

$$\sigma_Z|_{Z=6} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(2)}_{ZZ'}[T(R_{kk'}|\bar{R}^{(2)}_{ZZ'},W_{CC(2)})] + \sum_{\substack{k' \neq k \\ Z_{k'}=7}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)]^2 \quad (17)$$

$$\sigma_Z|_{Z=7} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}\{[T(R_{kk'}|\bar{R}_{ZZ'},W)] \sum_{\substack{k'' \neq k \\ k'' \neq k'}} [T(R_{kk'}|\bar{R}_{ZZ'},W)]^2\}^{1.3} + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(2)}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \sum_{\substack{k'' \\ Z_{k''}=8}} [T(R_{k'k''}|\bar{R}_{Z'Z''},W)] + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}^{(3)}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W_{NC})] \quad (18)$$

$$\sigma_Z|_{Z=8} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=6}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W_{OC})] + \sum_{\substack{k' \neq k \\ Z_{k'}=7,8,15}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (19)$$

$$\sigma_Z|_{Z=9} = \tilde{\sigma}_Z \quad (20)$$

$$\sigma_Z|_{Z=15} = \tilde{\sigma}_Z \quad (21)$$

$$\sigma_Z|_{Z=16} = \tilde{\sigma}_Z + \sum_{\substack{k' \neq k \\ Z_{k'}=15,16}} \tilde{\sigma}_{ZZ'}[T(R_{kk'}|\bar{R}_{ZZ'},W)] \quad (22)$$

$$\sigma_Z|_{Z=17} = \tilde{\sigma}_Z \quad (23)$$

$$\sigma_Z|_{Z=35} = \tilde{\sigma}_Z \quad (24)$$

The functional forms shown above are a convenient way to treat different types of chemical environments that a particular atom in a solute might encounter. Unlike models that require the user to assign types to atoms (e.g., molecular mechanics force fields), these functional forms do not require the user to make assignments. This feature means that the user is never in doubt about which parameter to use. Furthermore, because the above equations are smooth functions of the solute geometry, the present model can be applied to systems containing nonbonded or partially bonded pairs of atoms, such as transition states and solute−solvent clusters.

It is important to point out that although separating the free energy of solvation into the above components allows us to effectively model a wide variety of solutes, only the total free energy is a state function, so there is a certain degree of ambiguity associated with separating the aqueous solvation free energy, which is an experimental observable, into several contributions that cannot be measured independently. Thus, there is some flexibility in how to interpret these various contributions.

We interpret the contribution to the free energy arising from G_P as accounting for electrostatic interactions between the charge distribution of the solute (which is modeled as a collection of point charges distributed over atomic spheres) and the bulk electric field of the solvent when it is assumed to begin at a boundary defined by the effective Born radii. For ionic solutes, G_P makes the largest contribution to the overall solvation free energy. Because G_P is calculated under the assumption that the solvent responds linearly to the electronic distribution of the solute (hence[58] the factor $^1/_2$ appearing in eq 10), nonlinear solvent effects, such as changes in the dielectric constant of the solvent in the vicinity of the solute, and strong solute−solvent hydrogen bonds, which prevent the solvent from fully responding to the solute charge distribution, are not explicitly accounted for by this term. One strategy that we have used to partly account for these nonlinear effects has been to empirically adjust the values that we use for the intrinsic Coulomb radii. For neutrals, previous experience has shown that the overall performance of our models is relatively insensitive to the values used for these atomic radii, whereas for ions the performance of our models is quite sensitive to the choice of these radii. By making adjustments to the atomic radii, our previous models have achieved an accuracy of ~5 kcal/mol in aqueous solvation free energies for the majority of the ions used to train these models. However, the present data set is much larger and more diverse than our previous data sets for ions. Thus, one of the goals of this work will be to see if the deficiencies described above can adequately be accounted for by making empirical adjustments to the intrinsic Coulomb radii.

A second strategy to account for deficiencies of a bulk electrostatic model is to include G_{CDS}. In the past, we have interpreted the G_{CDS} term as formally accounting for cavitation (i.e. the free energy cost associated with creating a cavity in the solvent to accommodate the solute), dispersion interactions between the solute and solvent, and specific solute effects on the solvent structure (e.g. the loss of orientational freedom of water molecules around a nonpolar solute). However, because the G_{CDS} term is empirical in nature, it can be more accurately interpreted as accounting for *any* contribution to the total solvation free energy of a given solute that is not explicitly accounted for by the bulk electrostatic (G_P) term. Such effects include, but are not limited to, the nonlinear solvent effects described above, deviations of the true solute−solvent interface defined by the atomic radii, short-range exchange and repulsion forces between the solute and solvent, neglect of charge transfer between the solute and solvent, and any systematic errors that may arise from the GB approximation, the ability of partial atomic charges to represent the true solute charge distribution, or the level of theory used to calculate the electronic wave function of the solute. In addition, several other effects are implicitly accounted for by G_{CDS} that could, in principle, be explicitly calculated, such as the change in the solute's translational, vibrational, or rotational free energy in moving from the gas phase to solution. By using atomic surface tensions, our previous models have been quite

successful at predicting aqueous solvation free energies for a wide variety of neutral solutes (an average error of ~0.5 kcal/mol), including many hydrogen bonding solutes. It is important to point out that because (in the present work and in most, but not all, of our previous work) we optimize the parameters contained in the G_{CDS} term using only neutral solutes, applying this scheme also to ions involves the assumption that first solvation shell interactions are similar for a given pair of solutes containing the same functionality but different formal charge; this may be a major contributor to the residual error. We could of course eliminate this problem by using different surface tension parameters for ions, but, as stated above, unlike some of the other methods in the literature, we avoid using molecular mechanics types.

4. Parameters To Be Optimized

Three different types of parameters will be optimized as part of this work: (1) the atomic radii in eq 12, (2) the value for d in eq 11, and (3) the atomic surface tension parameters $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ in eqs 14−24. The above parameters will be optimized using gas-phase geometries. In all cases, the solutes in our database were represented by a single, lowest-energy conformation. For some of the solutes, in particular the solute−water clusters, this involved performing a conformational analysis to identify the global minimum on the potential energy surface. For the acetamide cation, we used the geometry corresponding to the oxygen-protonated species, which is 9.7 kcal/mol lower in free energy in the gas phase than the nitrogen-protonated species (MPW25/MIDI! level of theory). For the acetamide anion, we used the deprotonated imidate form, which is lower in free energy in the gas phase than the enolate by 15.7 kcal/mol (MPW25/MIDI! level of theory).

For all of the unclustered solutes used in this work, we used geometries optimized at the MPW25/MIDI! level of theory. The MIDI! basis set[61,62,128] is an especially economical basis set for calculations on large organic systems, but it was designed to give particularly accurate geometries. All of the solute−water clusters were optimized at the B97-1/MG3S level of theory.

5. Results

5.1. Partial Atomic Charges. Although there is formally no "correct" method for assigning partial atomic charges because partial atomic charge it is not a quantum mechanical observable,[129] several qualities make CM4 partial atomic charges more suitable for use in the present model than charges obtained from other methods. First, dipole moments derived from CM4 point-charges are generally more accurate than point-charge-derived dipole moments obtained using other charge partitioning schemes. This is demonstrated by the data in Table 5, which lists the mean unsigned errors between experimental dipole moments for 397 neutral molecules and those dipole moments calculated using CM4 point charges and those obtained from a redistributed Löwdin population analysis (RLPA).[117] Also listed are the average errors for 107 unclustered ionic molecules (experimental dipole moments are not available for ionic solutes, so we used density dipole moments calculated at the MPW25/MG3S//MPW25/MIDI! level of theory for comparison). The data in Table 5 show that for most of the molecules tested here, CM4-point-charge-derived dipole moments are more accurate than those calculated using RLPA partial atomic charges (for nondiffuse basis sets, RLPA partial atomic charges are equivalent to those obtained from a Löwdin population analysis[113−116]). One notable exception is for the anions tested here, where at the MPW25/MIDI!6D level of theory, RLPA-point-charge-derived dipole moments are more accurate than the CM4-point-charge-derived dipole moments (for other levels of theory, the CM4-point-charge derived dipole moments are more accurate).

Table 5. Mean Unsigned Errors (Debyes) in Dipole Moments Calculated with Partial Atomic Charges Obtained from CM4 and RLPA at the MPW25 Level of Theory[a]

	CM4			RLPA[b]		
basis set	neutrals	cations	anions	neutrals	cations	anions
MIDI!6D	0.19	0.20	0.61	0.37	0.24	0.35
6-31G(d)	0.23	0.24	0.36	0.62	0.21	0.46
6-31+G(d)	0.27	0.32	0.44	0.76	0.43	0.55
6-31+G(d,p)	0.26	0.33	0.40	0.81	0.67	0.67

[a] For the neutral solutes, point-charge-derived dipole moments were compared to dipole moments taken from the CM3 training set, which is described in ref 112 (397 total dipole moments). For the dipolar ions in Tables 1 and 2, point-charged derived dipole moments were compared to density dipole moments calculated at the MPW25/MG3S level of theory (107 total dipole moments). [b] For nondiffuse basis sets, RLPA partial atomic charges are equivalent to Löwdin partial atomic charges.

A second reason we prefer using CM4 partial atomic charges is because they are less sensitive to changes in the basis set than partial atomic charges obtained from other models. This becomes especially true when diffuse functions are added. Table 5 shows that as polarization and diffuse functions are added to the basis set, the quality of RLPA charges progressively worsens, whereas a much smaller dependency on basis set is observed for the CM4 charges.

Finally, it has recently been shown[112] that CM3 (a model quite similar to CM4) delivers more accurate charges for interior or buried atoms than partial atomic charges calculated from a fit to the electrostatic potential calculated around the molecule of interest (e.g. the ChElPG charge model[130]). In this same paper, it was also shown that CM3 charges are much less sensitive to small conformational changes and to the level of treatment of electron correlation than are ChElPG charges.

5.2. Aqueous Solvation Free Energies Calculated Using Previously Defined Atomic Radii. First, we tested several previously defined sets of atomic radii for predicting aqueous solvation free energies. For this, we developed three intermediate models using three different sets of atomic radii in eq 12, in particular the following: Bondi's atomic radii (which we also use to calculate the solvent-accessible-surface area in eq 13) and the radii used by our SM5.42R[10−14] and SM5.43R[15,16] models. These sets of atomic radii are listed in Table 6. For each intermediate method, we calculated ΔG_{EP} values at the MPW25/6-31+G(d,p) level of theory for all of the solutes in our data set, with d fixed at 4. With calculated values for ΔG_{EP} in hand, we then optimized a set of surface tension coefficients $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ for each intermediate model by minimizing the errors between the experimental

Table 6. Atomic Radii Used by Various Models

atom	Bondi	SM5.42R	SM5.43R	SM6
H	1.20	0.91	0.79	1.02
C	1.70	1.78	1.81	1.57
N	1.55	1.92	1.66	1.61
O	1.52	1.60	1.63	1.52[c]
F	1.47	1.50	1.58	1.47[c]
P	1.80	2.27	2.01	1.80[c]
S	1.80	1.98	2.22	2.12
Cl	1.75	2.13	2.28	2.02
Br	1.85	2.31	2.38	2.60

[a] Not optimized, held fixed at Bondi's value.

Table 7. Mean Unsigned Errors in Aqueous Solvation Free Energies (kcal/mol) Obtained Using Various Sets of Atomic Radii[a]

		ions	
atomic radii[b]	neutrals	unclustered[c]	selectively clustered[d]
Bondi	0.56	6.87	5.55
SM5.42R	0.52	5.64	4.73
SM5.43R	0.52	6.06	5.32

[a] For each set of radii, a different set of atomic surface tensions was optimized. All *d* values were fixed at 4 for the calculations in this table, and (in the whole article) we always use Bondi's radii in eq 13. The calculations in this table were carried out using MPW25/6-31+G(d,p). [b] Intrinsic Coulomb radii. [c] This data set contains all 112 ions listed in Tables 1 and 2. [d] This data set contains 81 of the ions listed in Tables 1 and 2 (those that do not appear in clustered form in Table 3) plus the 31 clustered ions listed in Table 3.

and calculated solvation free energies for all 273 of the neutral solutes. This step was accomplished using a NAG Fortran 90 routine, in particular the linear least squares solver routine.[131] The performance of the three intermediate models is summarized in Table 7.

The data in Table 7 show that all three models lead to similar errors in the solvation free energies of neutral solutes. Although not shown explicitly in Table 7, all three of the intermediate models predict an aqueous solvation free energy for water that is ~4 kcal/mol too negative. We encountered a similar error during the development of earlier models and removed it by including a surface tension that identified oxygen atoms in the vicinity of two hydrogen atoms. We did not include this surface tension here, because doing so would have a negative impact on the performance of the model for hydronium, protonated alcohols, and solute−water clusters. For the water *dimer*, all three of the intermediate models predict its solvation free energy correctly to within 1.2 kcal/mol without using the special surface tension, which is a significant improvement compared to the performance of these models for the bare water solute.

For ions, the errors shown in Table 7 are broken down into two different subsets: the unclustered-ion data set and the selectively clustered-ion data set. The unclustered-ion data set contains all of the experimental aqueous solvation free energies listed in Tables 1 and 2 but none of the solute−water clusters in Table 3 (112 ionic solutes). The selectively clustered-ion data set contains all of the experimental aqueous solvation free energies in Table 3 plus those aqueous solvation free energies that are in Tables 1 or 2 but not Table 3 (i.e. solutes that are included in the selectively clustered-ion data set as solute−water clusters are not included as their analogous bare ions). The criteria we used to decide to which of the solutes in our training set to add an explicit water molecule (i.e. which bare solutes would be deleted from the selectively clustered-ion data set and replaced by their analogous water−solute cluster) is as follows. First, we added an explicit water molecule to any ionic solute containing three or fewer atoms. Second, we added an explicit water molecule to any ionic solute with one or more oxygen atoms bearing a more negative partial atomic charge than bare water solute (as judged by comparison of CM4 gas-phase charges computed at the MPW25/6-31+G(d,p) level of theory). Finally, we added an explicit water molecule to ammonium and to all of the oxonium ions. We singled out solutes that satisfied one or more of these three criteria because we felt these solutes were likely to form strong solute−solvent hydrogen bonds with water and therefore would serve as a useful indicator as to whether including a single explicit water molecule in the calculation is an effective strategy for accounting for strong solute−solvent hydrogen bonding effects.

The data in Table 7 are consistent with the fact that the solvation free energies of ions are more sensitive to the choice of atomic radii than the neutrals. This is not surprising, since the ΔG_{EP} term is the major contributor to the total solvation free energy for ions and because the surface tensions are optimized for a given set of values of ΔG_{EP}. The data in Table 7 also show that all three intermediate models give significantly lower errors for the selectively clustered-ion data set than for the unclustered-ion data set. This indicates that for the above intermediate models including a *single* explicit water molecule in the calculation is at least partly effective in accounting for strong specific solute−solvent hydrogen bonding interactions.

5.3. Aqueous Solvation Free Energies Calculated Using Optimized Atomic Radii and *d* Parameter. Of the three intermediate models presented above, the one based on SM5.42R radii performs the best for ionic solutes, giving a mean unsigned error of 5.64 kcal/mol for the ions in the unclustered-ion data set and 4.73 kcal/mol for the ions in the selectively clustered-ion data set. Next, we examined whether optimizing a new set of radii would lead to more accurate solvation free energies. For this, we again used MPW25/6-31+G(d,p). Throughout the parameter optimization process, for each set of intrinsic Coulomb radii, we optimized a different set of surface tension coefficients for each set of radii by minimizing the overall error between the calculated and experimental aqueous solvation free energies for all of the neutral solutes. This two-step procedure (where first a set of intrinsic Coulomb radii are chosen, and then surface tension coefficients are optimized for that set of atomic radii) was implemented into a genetic algorithm,[132] and the average error between the calculated and experimental aqueous solvation free energies for all of the ionic solutes in the selectively clustered-ion data set (the reason we chose to use only the data from the selectively clustered-ion data set is discussed in section 6.2) was minimized. To ensure that a physical parametrization was achieved, we constrained the optimization in the following ways. First, any set of intrinsic Coulomb radii that yielded positive ΔG_{EP}

values for any of the solutes in our data sets, or for *n*-hexadecane, was disqualified. Second, any set of intrinsic Coulomb radii that yielded a ΔG_{EP} value more negative than −0.40 kcal/mol for any of the *n*-alkanes was disqualified. This second constraint was added because several initial optimizations led to unusually small values for intrinsic Coulomb radii of hydrogen and carbon atoms. A value of −0.40 kcal/mol was chosen as the cutoff because 0.40 kcal/mol is twice the value that we have previously estimated[15] for the average uncertainty in the free energy of solvation of a typical neutral solute.

Throughout the optimization, we found that small changes in the value used for *d* led to some improvement in the model. Therefore, we decided to optimize the *d* parameter simultaneously along with the intrinsic Coulomb radii. For this, we constrained the value of *d* between 3.5 and 4.5. We chose not to vary *d* more widely because it appears in the exponential term of eq 11.

With the above constraints, we found an optimum of 3.7 for *d*, which is close to the value of 4 originally suggested by Still et al.[105] The intrinsic Coulomb radii resulting from the above optimization are listed in the last column of Table 6. For oxygen and fluorine, we found that using Bondi's atomic radii[127] instead of the optimized radii had little effect on the overall performance of the model, so we used Bondi's values instead. For phosphorus, we held the value fixed at Bondi's value throughout the optimization because the current data set does not contain any phosphorus-containing ions.

5.4. Atomic Surface Tension Coefficients. With the intrinsic Coulomb radii and *d* parameter fixed at the values obtained above, again using all 273 of the neutral solutes in our data set, we optimized four different sets of atomic surface tension coefficients for the following levels of theory: MPW25/MIDI!6D, MPW25/6-31G(d), MPW25/6-31+G(d), and MPW25/6-31+G(d,p). The ΔG_{EP} values calculated at the MPW25/6-31+G(d) and MPW25/6-31+G(d,p) levels of theory for the solute *O*-ethyl *O*′-4-bromo-2-chlorophenyl *S*-propyl phosphorothioate are large outliers, so we omitted this solute from the determination of the atomic surface tension coefficients for these two levels of theory (although we did use their calculated aqueous solvation free energies to determine the errors shown in Table 9). The optimized surface tension coefficients for each of the above levels of theory are listed in Table 8.

Shown in the first column of Table 9 are the various classes of solutes in our data set. The next four columns list the average errors in aqueous solvation free energies for each solute class by level of theory (subsequent columns of this table are discussed in the last paragraph of this section). For neutral solutes, the overall performance of SM6 is relatively insensitive to changes in basis set. Closer inspection of the individual solute classes in Table 9 does, however, reveal that for the neutrals there are a few systematic differences between the aqueous solvation free energies calculated with and without diffuse functions. For example, using diffuse functions leads to calculated solvation free energies for the carboxylic acids that are on average 0.2−0.3 kcal/mol less accurate than those calculated without diffuse functions. Conversely, using diffuse functions reduces the average error by a factor of around 3 for the two halogenated sulfur compounds in the data set.

Table 8. SM6 Surface Tension Coefficients (cal/Å^2), Optimized for MPW*X* with Various Basis Sets

Z,*Z*′	MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)
H	60.3	60.7	55.2	55.2
C	96.8	92.5	114.6	108.0
N	49.5	14.9	7.1	9.8
O	−120.4	−126.9	−142.8	−152.5
H,C	−79.5	−80.7	−75.2	−73.7
C,C	−77.9	−71.8	−82.4	−80.2
C,C(2)	−21.5	−19.8	−29.5	−30.3
H,O	−116.7	−80.3	−59.4	−55.1
O,C	208.4	209.4	208.6	227.3
O,O	95.3	109.6	127.9	139.5
H,N	−135.9	−110.1	−89.9	−91.8
C,N	−11.6	20.8	17.4	7.5
N,C	−46.1	−50.7	−57.1	−58.0
N,C(2)	−226.1	−158.8	−257.5	−267.4
N,C(3)	15.0	75.9	44.7	54.6
O,N	228.2	236.8	254.0	267.4
F	38.4	32.7	26.3	26.3
Cl	−2.0	−1.6	0.4	0.7
Br	−19.2	−19.2	−22.2	−21.8
S	−59.9	−52.7	−75.6	−74.9
H,S	18.4	6.6	71.0	72.3
S,S	21.6	12.3	51.0	55.3
P	−35.0	12.1	21.7	14.9
O,P	153.9	135.9	196.3	202.7
S,P	114.3	104.9	104.8	122.5

Including diffuse basis functions has a much more significant impact on the calculated aqueous solvation free energies for the anionic solutes. Depending on whether diffuse basis functions are used, the average errors in the selectively clustered-ion data set range from 3.56 kcal/mol to 5.56 kcal/mol. For anions, the MPW25/MIDI!6D level of theory performs the poorest of all the levels of theory, which is not surprising since it was also shown to give the least accurate point-charge-derived dipole moments for anions. For the cations, including diffuse functions has almost no effect on the calculated aqueous solvation free energies—the average errors for the cationic portion of the selectively clustered-ion data set range from only 2.72 kcal/mol to 2.82 kcal/mol, and they do not follow the same trend as for the anions. Although, based on the above results, it is tempting to suggest that diffuse basis functions are a necessary requirement for calculating aqueous solvation free energies of anions, the role diffuse functions play in the liquid phase is not as well understood as their role in the gas phase. Therefore, we will avoid making any general statements regarding the relationship between using diffuse functions and the accuracy of various calculated liquid-phase properties.

Earlier work revealed[16] that for a given basis set and hybrid density functional using the same set of atomic surface tensions for any fraction *X* of Hartree−Fock exchange had little effect on the overall performance of our models. We also found this to be true here, so the surface tension coefficients in Table 8 can be used for any fraction of

Table 9. Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using SM6, by Solute Class

		MPW25					
solute class	*N*	MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	B3LYP/ 6-31+G(d,p)[a]	B3PW91/ 6-31+G(d,p)[a]
inorganic compounds	8	0.96	0.99	1.11	1.08	1.09	1.09
n-alkanes	8	0.53	0.58	0.62	0.59	0.53	0.57
branched alkanes	5	0.40	0.44	0.39	0.38	0.34	0.38
cycloalkanes	5	0.51	0.48	0.52	0.50	0.43	0.49
alkenes	9	0.30	0.29	0.28	0.26	0.24	0.26
alkynes	5	0.22	0.18	0.28	0.27	0.35	0.29
arenes	8	0.26	0.19	0.30	0.29	0.30	0.26
all hydrocarbons	40	0.37	0.36	0.40	0.38	0.36	0.37
alcohols	12	0.62	0.53	0.46	0.43	0.48	0.44
phenols	4	0.69	0.63	0.67	0.65	0.95	0.70
ethers	12	0.49	0.47	0.51	0.49	0.49	0.48
aldehydes	6	0.42	0.28	0.34	0.39	0.34	0.37
ketones	12	0.31	0.30	0.55	0.57	0.54	0.55
carboxylic acids	5	0.57	0.65	0.90	0.83	0.93	0.84
esters	13	0.33	0.40	0.52	0.46	0.41	0.44
bifunctional H,C,O compounds	5	0.70	0.53	0.47	0.46	0.50	0.46
peroxides[b]	3	0.21	0.20	0.27	0.26	0.22	0.26
all H,C,O compounds[c]	115	0.42	0.40	0.46	0.44	0.45	0.44
aliphatic amines	15	0.67	0.63	0.60	0.61	0.59	0.60
anilines	7	0.48	0.55	0.72	0.79	1.12	0.84
aromatic nitrogen heterocycles (1N in ring)	10	0.21	0.22	0.32	0.35	0.53	0.37
aromatic nitrogen heterocycles (2Ns in ring)	2	0.71	0.56	0.53	0.49	0.51	0.50
hydrazines[d]	3	0.97	0.86	0.91	0.91	0.84	0.90
nitriles	4	0.41	0.34	0.63	0.62	0.69	0.61
bifunctional H,C,N compounds	3	0.33	0.32	0.39	0.39	0.51	0.39
all H,C,N compounds[c]	85	0.45	0.44	0.50	0.50	0.54	0.50
amides	4	1.07	0.89	1.02	0.97	1.10	0.99
ureas	2	0.59	0.31	0.61	0.45	0.25	0.40
nitrohydrocarbons	7	0.58	0.28	0.24	0.30	0.35	0.29
bifunctional H,C,N,O compounds	4	0.67	0.58	0.40	0.46	0.34	0.46
H,C,N,O compounds[c]	177	0.50	0.46	0.53	0.52	0.55	0.52
fluorinated hydrocarbons	6	0.51	0.46	0.41	0.42	0.43	0.42
chlorinated hydrocarbons	27	0.43	0.34	0.37	0.40	0.40	0.39
brominated hydrocarbons	14	0.24	0.20	0.34	0.33	0.22	0.31
multihalogen hydrocarbons	12	0.31	0.30	0.43	0.42	0.42	0.43
halogenated bifunctional compounds	9	1.00	0.95	1.24	1.21	1.23	1.20
thiols	4	0.29	0.32	0.30	0.28	0.29	0.27
sulfides[e]	5	0.65	0.78	0.48	0.46	0.40	0.44
disulfides	2	0.21	0.16	0.37	0.34	0.38	0.34
sulfur heterocycles	1	0.40	0.11	0.86	0.84	0.60	0.80
halogenated sulfur compounds	2	1.70	1.90	0.47	0.29	0.36	0.25
all non-phosphorus sulfur compounds	14	0.61	0.67	0.44	0.39	0.38	0.37
all phosphorus compounds [f]	14	0.64	0.72	1.30	1.18	1.10	1.22
neutrals	273	0.50	0.47	0.55	0.54	0.55	0.54
H,C,N,O ions[c,g]	91	4.26	3.83	3.39	3.30	3.46	3.34
F,Cl,Br,S ions[g,h]	21	4.17	3.63	2.82	2.83	3.01	2.85
anions[g]	60	5.56	4.66	3.68	3.56	3.76	3.62
cations[g]	52	2.72	2.80	2.82	2.81	2.93	2.82
ions[g]	112	4.24	3.80	3.28	3.21	3.37	3.25

[a] Surface tension coefficients optimized for the MPW25/6-31+G(d,p) level of theory were used for these calculations. [b] The inorganic solute hydrogen peroxide is included in this solute class as well as the inorganic compound solute class. [c] Solutes containing at most the listed elements. [d] The inorganic solute hydrazine is included in this solute class as well as the inorganic compound solute class. [e] The inorganic solute hydrogen sulfide is included in this solute class as well as the inorganic compound solute class. [f] The inorganic compound phosphine is included in this solute class as well as the inorganic compound solute class. [g] The ion data in this table are all taken from the selectively clustered-ion data set. [h] Solutes containing at least one of the listed elements plus, in most cases, elements from the previous row.

Hartree−Fock exchange. Furthermore, we propose that they can be used with any good density functionals. To test the accuracy of the surface tensions in Table 8 when used with density functionals other than MPW*X*, we calculated aqueous solvation free energies at the SM6/B3LYP[69]/6-31+G(d,p)//MPW25/MIDI! and SM6/B3PW91[133]/6-31+G(d,p) //MPW25/MIDI! levels of theory (both of these functionals have $X = 20\%$) using the atomic surface tension coefficients in Table

8. The resulting errors are shown in the final two columns of Table 9. In almost all cases, the aqueous solvation free energies calculated at these levels of theory are very close to those calculated at the SM6/MPW25/6-31+G(d,p)//MPW25/MIDI! level of theory. This invariance is due to the ability of these density functionals to deliver accurate electronic wave functions, and the above results are encouraging because they show that the above parameters can be applied to a wide variety of different density functionals, assuming that the given density functional is able to provide a reasonably accurate electronic wave function for the solute of interest.

6. Discussion

6.1. Optimizing Solvation Parameters Based on Gas-Phase Geometries. The first point worth some discussion is the fact that the method is parametrized using gas-phase geometries. We optimize the parameters based on gas-phase geometries here as well as in recent previous work[7,10−12,14−17,74,75] for two reasons. First, for many solutes, less expensive (e.g., semiempirical or molecular mechanics methods) can yield accurate gas-phase geometries. Second, our experience is that optimizing solvation parameters based on gas-phase geometries does not cause a problem because for all or almost all of the molecules in the parametrization set (with the possible exception of those containing an explicit solvent molecule), the difference in solvation free energy between using a gas-phase geometry and using an aqueous geometry is smaller than the mean error of the model. Having obtained the parameters with such a training set, they can be used more broadly. Thus, when the geometry does change significantly in solution, the molecule should be optimized in the aqueous phase, and such optimization would be expected to give more accurate results in such a case. The ASA algorithm[122] that we use for the solvation calculations has excellent analytic gradients that allow for efficient and stable geometry optimization in solution.

6.2. Optimizing Atomic Radii Against the Unclustered-Ion Data Set. The results presented above suggest that for some of the ions considered here as well as water adding an explicit water molecule to the calculation is one way to increase the accuracy of the model for these solutes. However, since the solute−water clusters in Table 3 were included in the training set used to obtain the parameters contained in the present model, an interesting question is whether better results might be obtained if only the unclustered ions were considered. One strategy that could be used to improve the performance of the model for the unclustered-ion data set would simply be to repeat the above optimization using the unclustered instead of the selectively clustered-ion data set. Indeed, it was shown in section 5.3 that the solvation free energies of the ions are quite sensitive to the choice of radii. Based on a careful examination of the solvation free energies for all of the ions in our data set though, we determined that some of the bare ions required a much different set of parameters than the majority of the other solutes in our data set. Therefore, fitting the unclustered-ion data set with a single set of atomic radii gives uneven results, and that strategy was abandoned.

Another strategy that could be used to try to fit all of the data in the unclustered-ion data set would be to make the atomic radius of a given atom not only a function of its atomic number but also a function of its partial atomic charge. In fact, our earliest models (SM1-SM3[3−6]) used charge-dependent intrinsic Coulomb radii. We experimented with several different functional forms for charge-dependent radii, including (as just one example) making intrinsic Coulomb radii a quadratic function of partial atomic charge, and we found that while making the intrinsic Coulomb radii a function of partial atomic charge did improve the performance of the model for many of the ions it did so at the cost of having a deleterious effect on some of the other ions. We eventually abandoned the idea of using charge-dependent radii because of this finding and for two additional reasons. First, a model that uses charge-dependent radii would be more likely to be highly basis-set dependent than a model that uses atomic-number-dependent radii.[35,36] Second, charge-dependent atomic radii might lead to highly questionable results in cases where this dependence has not been carefully examined (e.g. our data set does not contain any zwitterions or large biomolecules) or in cases where the atomic radius might not be a smooth function of its partial atomic charge (e.g. transition states that involve the displacement of a charged or partially charged leaving group).

Several additional strategies that were not considered here, but that have been used by others, include using atom-typed radii[18−23] or using different sets of atomic radii for neutrals and ions.[134] An example of this former type of approach is the united atom for Hartree−Fock (UAHF)[23] method of Barone and co-workers, in which the atom's radius depends on its hybridization state, connectivity, and formal charge (so, in a sense, this method falls under both of the above categories). Here, we did not consider using atom-typed radii because it is not completely clear whether they can be used for modeling chemical reactions or predicting activation free energies (which require modeling a transition state) in the liquid phase.

We feel that adding one or more explicit water molecules to the calculation is the most reasonable approach to modeling some ionic solutes with a continuum solvent model. Ideally we would give a definite prescription for when the user should add a specific water molecule in using this model. However, it is not possible to do this in a way that covers the great diversity of possible cases that occur in applications, especially if one includes reaction paths and enzymes. One prescription would be to add an explicit water whenever one wants to improve the accuracy, since adding an explicit water should almost always improve the accuracy when the effect is large, but it is relatively safe because it cannot make the accuracy much worse when the effect is small. In considering this question, we should note that although we obtain better results when we add an explicit water in cases where the given solute satisfies one or more of the three criteria explained in section 5.2, we also obtain reasonably good results in most cases even without the explicit water (see Table 10).

Although not tested as part of this work, the above strategy of adding one or more explicit solvent molecules should also improve the accuracy for predicting solvation free energies of some solutes in nonaqueous solvents, in particular those where specific solute−solvent hydrogen bonds are expected

Table 10. Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using Different Continuum Solvent Models

	MPW25/6-31+G(d,p)		HF/6-31G(d)					MPW25/6-31+G(d,p)
	SM6	SM5.43R	DPCM/98	DPCM/03	CPCM/98	CPCM/03	IEF-PCM/03	IEF-PCM/03
neutrals[a]	0.54	0.62	1.02	1.40	1.06	1.11	1.10	1.21
unclustered ions[b]	4.19	6.12	4.40	14.95	5.02	6.32	6.37	7.78
clustered ions[b]	3.21	6.16	5.86	14.32	6.33	7.58	7.63	8.98
all ions[c]	4.38	6.92	5.83	15.60	6.40	7.47	7.53	8.88
all data[d]	1.86	2.79	2.69	6.28	2.91	3.30	3.31	3.85

[a] 273 molecules. [b] 112 ions. [c] 143 ions. [d] 416 data.

to occur. Of course, testing this strategy for solvents other than water would require a careful comparison between experimental and calculated solvation free energies of solutes in nonaqueous solvents, which is beyond the scope of the present work.

6.3. Performance of Other Continuum Models. Using our database of aqueous solvation free energies, along with gas-phase geometries optimized at the MPW25/MIDI! level of theory, we tested the performance of the SM5.43R and PCM continuum models for predicting aqueous solvation free energies. For all PCM calculations, we used the UAHF method for assigning atomic radii,[23] which is the recommended method for predicting aqueous solvation free energies with PCM according to the *Gaussian 03* manual.[135] Because the parameters contained in the UAHF method were originally optimized for use with the HF/6-31G(d) level of theory,[23] we used this level of theory to calculate aqueous solvation free energies for all of the PCM methods tested here. (Thus the PCM methods are tested in a way that should allow them to perform at their best.) However, we also wanted to see what effect changing the level of theory had on the accuracy of PCM, so we tested one of the PCM methods described below at the PCM/MPW25/6-31+G(d,p)//MPW25/MIDI! level of theory also.

There are several different varieties of PCM, and most of these are implemented differently in *Gaussian 98* [136] and *Gaussian 03*.[135] Here, we tested the dielectric version[26−28] of PCM (DPCM) as implemented in both *Gaussian 98* and *Gaussian 03*.[29] These two models will be referred to as DPCM/98 and DPCM/03, respectively. We also tested CPCM/98[25,137,138] and CPCM/03[24,137,138] as well as the default PCM method in *Gaussian 03*, IEF-PCM/03.[29−32] The IEF-PCM/03 model is particularly interesting because Chipman[139] found that it includes charge penetration effects "extremely well for all solutes". The results of these calculations are summarized in Table 10.

The data in Table 10 show that SM6 outperforms all of the other models tested above for both neutral and ionic solutes. For PCM, the most accurate solvation free energies are obtained using the older DPCM/98 implementation. The data in Table 10 also show that changing the level of theory has a negative effect on the performance of IEF-PCM/03, which is not surprising, since the UAHF method for assigning atomic radii was optimized[23] using DPCM/98 at the HF/6-31G(d) level of theory. Of course, the performance of SM6 for anions is dependent on the basis set used (see the data in Table 9), although its performance improves as the basis set size in increased (for neutrals and cations, there is very little basis-set dependence in aqueous solvation free energies calculated with SM6).

Comparing the overall errors for the unclustered-ion data set to those in the selectively clustered-ion data set shows that only the performance of SM6 improves significantly when a single explicit solvent molecule is added to the calculation (the overall error decreases from 4.19 kcal/mol for the ions in the unclustered-ion data set to 3.21 kcal/mol for the ions in the selectively clustered-ion data set). Again, this suggests that including a small number of explicit water molecules in SM6 calculations may be an effective strategy for predicting the aqueous solvation free energies of some ions in cases where strong solute−solvent hydrogen bonds are expected to play an important role in the aqueous phase. Furthermore, we feel that this is a much more reasonable strategy than trying to use drastically scaled values for the atomic radii or atom-typed or charge-dependent radii. The excellent performance of the SM6 model as compared to all the models in the popular *Gaussian* packages is especially remarkable when one remembers that the atomic radii in SM6 are functions of only atomic number (and the radii for O and F are not even optimized), whereas the recommended radii used in *Gaussian* depend on connectivity, hybridization state, and formal charge.

7. Concluding Remarks

We have presented a new database of experimental aqueous solvation free energies that contains 273 neutral and 143 ionic solutes, including 31 ion−water clusters. Using these data, we developed a new continuum solvent model called SM6. This model can be used to calculate aqueous solvation free energies and, although not demonstrated here, liquid-phase geometries in aqueous solution. SM6 uses partial atomic charges obtained from a new charge model, Charge Model 4 (CM4), which has been shown to give accurate partial charges for both neutral and ionic solutes. In addition, we have shown that the partial atomic charges obtained from CM4 are much less dependent upon changes in the basis set than partial atomic charges obtained from a Löwdin or Redistributed Löwdin population analysis of the wave function.

For some of the ions in our data set, we showed that the addition of a single explicit water molecule to the calculation (i.e., modeling the solute as a solute−water cluster) improved the performance of SM6 for predicting aqueous solvation free energies, indicating that large numbers of solvent molecules are not necessarily required for improving the treatment of some strong solute−solvent hydrogen bonds in

the first solvation shell. This is encouraging, because treating large numbers of solvent molecules explicitly often presents many problems. Furthermore, we feel that this strategy is more reasonable than using unphysical values for the atomic radii or using atom-typed or charge-dependent radii.

We also used our new database of aqueous solvation free energies to test the performance of several other continuum solvent models, namely SM5.43R and several different implementations of PCM. For both neutral and ionic solutes, SM6 outperforms all of the methods against which it was tested for predicting aqueous solvation free energies. Furthermore, we found that SM6 is the only model of those tested here (except for one model with a mean error 3.4 times larger) that improves upon the addition of a single explicit water molecule to the calculation. Thus, unlike the other models tested here, adding a single explicit water molecule to SM6 calculations in cases where strong solute−solvent hydrogen bonds are expected to occur in the aqueous phase appears to be both practical and effective for improving the accuracy of the present model for these types of solutes.

Finally, it was shown that SM6 retains its accuracy when used in conjunction with the B3LYP and B3PW91 functionals. Based on this analysis, we proposed that the charge and solvation parameters obtained with a given basis set (charge and solvation parameters have been optimized for the MIDI!6D, 6-31G(d), 6-31+G(d), and 6-31+G(d,p) basis sets) may be used with any good density functional or fraction of Hartree−Fock exchange.

Availability of SM6. All of the SM6 parametrizations presented in this article are available in the SMXGAUSS[140] program. This program can read a *Gaussian* output file corresponding to a gas-phase calculation of a given solute and carry out a single-point calculation with SM6. In addition, the above program allows liquid-phase geometry optimizations and Hessian calculations to be carried out with SM6. Although SMXGAUSS requires only a *Gaussian* output file to perform SM6 calculations, users that have a *Gaussian 03* executable can use SMXGAUSS in conjunction with the powerful geometry optimizers available in *Gaussian*. For non-*Gaussian* users, the CM4 and SM6 parametrizations are also available in the GAMESSPLUS[141] and HONDOPLUS[142,143] software programs. All three of these programs are available free of charge and can be downloaded from our Web site, http://comp.chem.umn.edu/software.

Acknowledgment. We thank Adam Moser for his help during preparation of the ionic portion of our data set. This work was supported by the NIH training grant for Neurophysical-computational Sciences, by the U.S. Army Research Office under Multidisciplinary Research Program of the University Research Initiative (MURI) through grant number DAAD19-02-1-0176, by the Minnesota Partnership for Biotechnology and Medical Genomics, and by the National Science Foundation (CHR-230446).

Supporting Information Available: Further details regarding CM4, all 273 experimental aqueous solvation free energies for the neutral solutes, and a breakdown of errors by data set for all of the SM6 methods tested in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.

(2) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.

(3) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(4) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *6*, 629−666.

(5) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213−217.

(6) Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089−1097.

(7) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257−3271.

(8) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(9) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

(10) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293−298.

(11) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc*. **1999**, *103*, 9−63.

(12) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117−9133.

(13) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503−5513.

(14) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178−2182.

(15) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532−6542.

(16) Thompson, D. J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 107−131.

(17) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem*. **2000**, *21*, 340−366.

(18) Banavali, N. K.; Roux, B. *J. Phys. Chem. B* **2002**, *106*, 11026−11035.

(19) Curutchet, C.; Bidon-Chanal, A.; Soteras, I.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2005**, *109*, 3565−3574.

(20) Marten, B.; Kim, K.; Cortis, C.; Friesner, R.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11778.

(21) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239−5248.

(22) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. *J. Chem. Theory. Comput.* **2005**, *1*, 484−493.

(23) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(24) Cossi, M.; Rega, N.; Scalamani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.

(25) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(26) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.

(27) Miertus, S.; Scrocco, E.; Tomasi, J. *J. Chem. Phys.* **1981**, *55*, 117−129.

(28) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239−245.

(29) Cossi, M.; Scalamani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54.

(30) Cancés, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(31) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253−260.

(32) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151−5158.

(33) Mennucci, B.; Cancés, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506−10517.

(34) Tomasi, J.; Mennucci, B.; Cancés, E. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211−226.

(35) Aguilar, M. A.; Martin, M. A.; Tolosa, S.; Olivares del Valle, F. J. *J. Mol. Struct. (THEOCHEM)* **1988**, *166*, 313−318.

(36) Aguilar, M. A.; Olivares del Valle, F. J. *Chem. Phys.* **1989**, *129*, 439−450.

(37) Gonçalves, P. F. B.; Livotto, P. R. *Chem. Phys. Lett.* **1999**, *304*, 438−444.

(38) Olivares del Valle, F. J.; Aguilar, M. A.; Contador, J. C. *Chem. Phys.* **1993**, *170*, 161−165.

(39) Rick, S. W.; Berne, B. J. *J. Am. Chem. Soc.* **1994**, *116*, 3949−3954.

(40) Takahashi, O.; Sawahata, H.; Ogawa, Y.; Kikuchi, O. *J. Mol. Struct. (THEOCHEM)* **1997**, *393*, 141−150.

(41) Camaioni, D. M.; Dupuis, M.; Bentley, J. *J. Phys. Chem. A* **2003**, *107*, 5778−5788.

(42) Rinaldi, D.; Rivail, J.-L.; Rguini, N. *J. Comput. Chem.* **1992**, *13*, 675−680.

(43) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413−7422.

(44) Chipman, D. M. *J. Chem. Phys.* **2002**, *116*, 10129−10138.

(45) Rode, B. M.; Schwenk, C. F.; Tongraar, A. *J. Mol. Liq.* **2004**, *110*, 105−112.

(46) Licheri, G.; Piccaluga, G.; Pinna, G. *J. Chem. Phys.* **1976**, *64*, 2437−2441.

(47) Kalko, S. G.; Sesé, G.; Padró, J. A. *J. Chem. Phys.* **1996**, *104*, 9578−9585.

(48) Obst, S.; Bradaczek, H. *J. Phys. Chem.* **1996**, *100*, 15677−15687.

(49) Probst, M. M.; Radnai, T.; Heinzinger, K.; Bopp, P.; Rode, B. M. *J. Phys. Chem.* **1985**, *89*, 753−759.

(50) Tongraar, A.; Liedl, K. R.; Rode, B. M. *J. Phys. Chem. A* **1997**, *101*, 6299−6309.

(51) Kerdcharoen, T.; Morokuma, K. *J. Chem. Phys.* **2003**, *118*, 8856−8862.

(52) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2005**, *1*, 415−432.

(53) Pliego Jr., J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434−7439.

(54) Pliego Jr., J. R.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241−7247.

(55) Fox, T.; Rösch, N. *Chem. Phys. Lett.* **1992**, *191*, 33−37.

(56) Karelson, M.; Zerner, M. C. *J. Am. Chem. Soc.* **1990**, *112*, 9405−9406.

(57) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, 1980; pp 25ff.

(58) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; pp 1−80.

(59) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754−2759.

(60) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, 379−400.

(61) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281−301.

(62) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1998**, *99*, 192−196.

(63) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(64) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664−675.

(65) Perdew, J. P. *Electronic Structure of Solids '91*; Zieesche, P., Eshrig, H., Eds.; Akademie: Berlin, 1991.

(66) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.

(67) Kormos, B. L.; Cramer, C. J. *J. Phys. Org. Chem.* **2002**, *15*, 712−720.

(68) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384−1388.

(69) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(70) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69−89.

(71) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *111*, 335−344.

(72) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384.

(73) Cramer, C. J.; Truhlar, D. G. In *Free Energy Calculations in Rational Drug Design*; Reddy, M. R., Erion, M. D., Eds.; Kluwer/Plenum: New York, 2001; pp 63−95.

(74) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147−7157.

(75) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305−4313.

(76) *Physical/Chemical Property Database (PHYSPROP);* SRC Environmental Science Center: Syracuse, NY, 1994.

(77) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661−1670.

(78) Curtiss, L. A.; Frurip, D. L.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703−2711.

(79) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583−5595.

(80) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109−6114.

(81) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 11534−11540.

(82) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. J.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.

(83) Lias, S. G.; Bartness, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. Ion Energetics Data. 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, March 2003.

(84) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413−656.

(85) Régimbal, J. M.; Mozurkewich, M. *J. Phys. Chem. A* **1997**, *101*, 8822−8829.

(86) O'Sullivan, D. W.; Lee, M.; Noone, B. C.; Heikes, B. G. *J. Phys. Chem.* **1996**, *100*, 3241−3247.

(87) Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic: New York, 1985.

(88) Arnett, E. M. *Prog. Phys. Org. Chem.* **1963**, *1*, 223−403.

(89) Kresge, A. J.; Pruszynski, P.; Stang, P. T.; Williamson, B. J. *J. Org. Chem.* **1991**, *56*, 4808−4811.

(90) Perrin, D. D. *Ionisation Constants of Inorganic Acids and Bases in Aqueous Solution*; Pergamon: New York, 1982.

(91) Albert, A.; Serjeant, E. P. *The Determination of Ionization Constants: A Laboratory Manual*; Chapman and Hall: New York, 1984.

(92) *Ionisation Constants of Organic Acids in Aqueous Solution*; Serjeant, E. P., Dempsey, B., Eds.; Pergamon: New York, 1979.

(93) Pliego Jr., J. R.; Riveros, J. M. *Phys. Chem. Chem. Phys.* **2002**, *4*, 1622−1627.

(94) Lewis, A.; Bumpus, J. A.; Truhlar, D. G.; Cramer, C. J. *J. Chem. Educ.* **2004**, *81*, 596−604.

(95) Winget, P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2004**, *112*, 217−227.

(96) Palascak, M. W.; Shields, G. C. *J. Phys. Chem. A* **2004**, *108*, 3692−3694.

(97) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264−6271.

(98) Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407−410.

(99) Lowry, T. H.; Schueller-Richardson, K. *Mechanism and Theory in Organic Chemistry*, 3rd ed.; Harper Collins: New York, 1987.

(100) Meot-Ner, M. M.; Lias, S. G. Binding Energies Between Ions and Molecules, and the Thermochemistry of Cluster Ions. 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, March 2003.

(101) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987.

(102) Hoijtink, G. J.; Boer, E. D.; Meij, P. H. v. D.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas.* **1956**, *75*, 487−503.

(103) Peradejori, F. *Cah. Phys.* **1963**, *17*, 393−447.

(104) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164−170.

(105) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(106) Cramer, C. J.; Truhlar, D. G. In *Rev. Comp. Chem.*; Boyd, D. B., Lipkowitz, K. B., Eds.; VCH Publishers: New York, 1995; Vol. 6; pp 1−72.

(107) Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *Minnesota solvent* descriptor database [http://comp.chem.umn.edu/solvation/mnsddb.pdf], 1999.

(108) Thompson, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291−1304.

(109) Winget, P.; Thompson, D. J.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 10707−10717.

(110) Brom, J. M.; Schmitz, B. J.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 6483−6488.

(111) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545−2549.

(112) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133−151.

(113) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365−375.

(114) Golebiewski, A.; Rzescowska, E. *Acta Phys. Pol.* **1974**, *45*, 563−568.

(115) Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221−229.

(116) Kar, T.; Sannigrahi, A. B.; Mukherjee, D. C. *J. Mol. Struct. (THEOCHEM)* **1987**, *153*, 93−101.

(117) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117−134.

(118) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.

(119) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, T. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(120) Born, M. *Z. Physik.* **1920**, *1*, 45−48.

(121) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465−1471.

(122) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422−440.

(123) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.

(124) Qui, D.; Shenkin, P.; Hollinger, F.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.

(125) Srinvasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426−434.

(126) Zhang, W.; Hou, T.; Xu, X. *J. Chem. Inf. Comput. Sci.* **2005**, *45*, 88−93.

(127) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.

(128) Thompson, J. D.; Winget, P.; Truhlar, D. G. *PhysChemComm* **2001**, *4*, 72−77.

(129) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.

(130) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.

(131) *NAG Fortran 90 Library*, 4th ed.; The Numerical Algorithms Group, Inc.: Oxford, 2000.

(132) Carroll, D. L. *GA Driver, Version 1.7a*; CU Aerospace, University of Illinois: Urbana, IL, 2001.

(133) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(134) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588−5593.

(135) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.;

Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalamani, G.; Rega, N.; Petersson, G. A.; Nagatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; 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.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.01*; Pittsburgh, PA, 2003.

(136) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; 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. W.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; 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.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.11*; Pittsburgh, PA, 1998.

(137) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312−9320.

(138) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(139) Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80−89.

(140) Kelly, C. P.; Thompson, D. J.; Lynch, B. J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Volobuev, Y.; Dupuis, M.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. SMXGAUSS-version 3.0; University of Minnesota: Minneapolis, MN 55455, 2005.

(141) Kelly, C. P.; Pu, J.; Thompson, D. J.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS-version 4.7; University of Minnesota, Minneapolis, 2005, based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(142) Dupuis, M.; Marquez, A.; Davidson, E. R. *HONDO 99.6*; based on HONDO 95.3, Dupuis, M.; Marquez, A.; Davidson, E. R., Quantum Chemistry Program Exchange (QCPE); Indiana University: Bloomington, IN, 1999.

(143) Kelly, C. P.; Nakamura, H.; Xidos, J. D.; Thompson, D. J.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. HONDOPLUS-version 4.7; University of Minnesota, Minneapolis, 2005, based on HONDO v. 99.6.

CT050164B