

## Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges

Aleksandr V. Marenich, Ryan M. Olson, Casey P. Kelly, Christopher J. Cramer,\* and Donald G. Truhlar\*

Department of Chemistry and Supercomputing Institute, University of Minnesota,  
207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431

Received June 10, 2007

**Abstract:** A new universal continuum solvation model (where “universal” denotes applicable to all solvents), called SM8, is presented. It is an implicit solvation model, also called a continuum solvation model, and it improves on earlier SMx universal solvation models by including free energies of solvation of ions in nonaqueous media in the parametrization. SM8 is applicable to any charged or uncharged solute composed of H, C, N, O, F, Si, P, S, Cl, and/or Br in any solvent or liquid medium for which a few key descriptors are known, in particular dielectric constant, refractive index, bulk surface tension, and acidity and basicity parameters. It does not require the user to assign molecular-mechanics types to an atom or group; all parameters are unique and continuous functions of geometry. It may be used with any level of electronic structure theory as long as accurate partial charges can be computed for that level of theory; we recommend using it with self-consistently polarized Charge Model 4 or other self-consistently polarized class IV charges, in which case analytic gradients are available. The model separates the observable solvation free energy into two components: the long-range bulk electrostatic contribution arising from a self-consistent reaction field treatment using the generalized Born approximation for electrostatics is augmented by the non-bulk-electrostatic contribution arising from short-range interactions between the solute and solvent molecules in the first solvation shell. The cavities for the bulk electrostatics calculation are defined by superpositions of nuclear-centered spheres whose sizes are determined by intrinsic atomic Coulomb radii. The radii used for aqueous solution are the same as parametrized previously for the SM6 aqueous solvation model, and the radii for nonaqueous solution are parametrized by a training set of 220 bare ions and 21 clustered ions in acetonitrile, methanol, and dimethyl sulfoxide. The non-bulk-electrostatic terms are proportional to the solvent-accessible surface areas of the atoms of the solute and have been parametrized using solvation free energies for a training set of 2346 solvation free energies for 318 neutral solutes in 90 nonaqueous solvents and water and 143 transfer free energies for 93 neutral solutes between water and 15 organic solvents. The model is tested with three density functionals and with four basis sets: 6-31+G(d,p), 6-31+G(d), 6-31G(d), and MIDI!6D. The SM8 model achieves mean unsigned errors of 0.5–0.8 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 2.2–7.0 kcal/mol for ions. The model outperforms the earlier SM5.43R and SM7 universal solvation models as well as the default Polarizable Continuum Model (PCM) implemented in *Gaussian 98/03*, the Conductor-like PCM as implemented in *GAMESS*, *Jaguar*’s continuum model based on numerical solution of the Poisson equation, and the GCOSMO model implemented in *NWChem*.

### 1. Introduction

Realistic solvation models must include long-range electrostatic polarization effects, which decrease as  $R^{-4}$  ( $R$  is the

distance between the solute and a given solvent molecule), shorter-range polarization effects, and shorter-range non-electrostatic effects such as cavitation, dispersion, and solvent structural effects (CDS), the latter including both hydrogen bonding and exchange repulsion effects.<sup>1–6</sup> These effects can be treated either in terms of explicit (atomistic) solvent or

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

implicit solvent, where the latter is usually represented by a continuous (also called continuum) medium characterized by both macroscopic properties, such as dielectric constant and bulk surface tension, and microscopic properties, such as polarizability and effective solvent radius. The “electrostatic” effect may be described as the electric polarization of the solvent by the polar or nonuniform charge distribution of the solute, and it also includes the effect of the self-consistent distortion of the solute by the polarized solvent. Although some efforts have been made<sup>4,5</sup> to treat nonelectrostatic terms (at least, dispersion) self-consistently (in so-called direct reaction field methods), a much more common assumption in self-consistent solvation models is that the solute charge distribution polarizes due to the electrostatic effects but not due to the nonelectrostatic ones. Thus the solute properties depend on the way that these effects are separated. Unfortunately though, there is no unique way to separate electrostatic effects from solvent structural effects. The ambiguity in current models is well illustrated by a comparison, a few years ago, of three successful aqueous solvation models based on different assumptions and model parameters.<sup>7</sup> For typical solutes (nitroethane, acetone, acetonitrile, benzaldehyde, and tagged water), the average difference between models of predicted standard free energies of solvation is 0.7 kcal/mol, whereas for the same cases the average difference from experiment is 0.6 kcal/mol, and the average difference between models of the electrostatic component is 2.1 kcal/mol.<sup>7</sup> Clearly the nonelectrostatic terms have been parametrized in a way that compensates for the differences in electrostatics.

Solvation models are usually parametrized and/or validated in terms of their ability to predict free energies of solvation, and implicit solvation models approximate such free energies as a sum of electrostatic and nonelectrostatic effects without cross terms. However, because the cross terms are not negligible, there is no unambiguous way to sort out the electrostatic and nonelectrostatic components of free energy. In fact the only possible separation of the free energy changes into components that are state functions (and hence independent of path) is the separation into enthalpy and entropy contributions<sup>8,9</sup> with a further possible separation, usually of little interest, of enthalpy into internal energy and work of compression.

A given separation of the free energy of solvation into electrostatic and nonelectrostatic contributions may therefore be associated with a particular implicit path for thermodynamic integration, and some paths may have more predictive power for modeling than others do.<sup>9</sup> One particularly relevant issue in this regard is that the magnitudes of solvation free energies of ions are much larger than those of neutral solutes and are dominated by large electrostatic contributions. Therefore a parametrization that is carried out in such a way that free energies of solvation of ions are accurate must be doing a good job of modeling electrostatics. By using the same parameters for neutrals one might also achieve a physical estimation of the electrostatics for cases where electrostatic and nonelectrostatic terms are comparable. The nonelectrostatic contribution can then be defined as the difference between the experimentally accessible and path-

independent total free energy of solvation and the modeled electrostatic contribution.

The most important parameters for modeling the electrostatics are the atomic radii; we call the radii used in the bulk electrostatics calculation the Coulomb radii (to distinguish them from van der Waals radii or covalent radii and from the radii used in the nonelectrostatic calculation). In the SM $x$  series of solvation models ( $x = 1, 2, \dots, 8$ ),<sup>3,10–12</sup> the Coulomb radii are calculated by a dielectric descreening approximation<sup>2</sup> from a set of intrinsic atomic Coulomb radii, and it has been our usual practice to optimize these intrinsic atomic Coulomb radii in calculations on ions, then fix these parameters and optimize the nonelectrostatic terms on data for neutrals. There have, however, been two flies in the ointment.

The first problem is that most ionic solvation data have involved an uncertainty related to the partition of the free energy of solvation of a salt or Brønsted acid into separate contributions associated with the cation and the anion because only their sum is well defined in classical thermodynamics.<sup>13</sup> This is resolved by molecular statistical mechanics by determining one absolute ionic solvation free energy, traditionally that of the proton.<sup>14</sup> However, there have been controversies about the value of that key quantity. Recent work, though, has largely eliminated these uncertainties,<sup>15–17</sup> and this enabled us to make a large database of ionic free energies of aqueous solvation.<sup>17</sup> These data were used<sup>17</sup> to test the performance of 13 solvation models for aqueous solvation energies of ions (see ref 18 for details), and the best performance was found for the SM6 solvation parameters<sup>18</sup> with the mPW1PW density functional<sup>19</sup> (also called MPW25, mPW0, and mPW1PW91), the Charge Model 4 (CM4),<sup>18</sup> and the 6-31G(d)<sup>20</sup> basis set.

A second problem though is that implicit solvation models have not been well studied for nonaqueous ionic solvation. One reason is the complication of ion pairing in media with low dielectric constants. Even in dilute solutions in more strongly solvating nonaqueous media, where ionic pairing may be neglected in calculating solvation free energies, although it is not totally absent,<sup>21–24</sup> there is a paucity of data, and until recently there has been no accurate determination of an absolute single-ion solvation energy, which is required, just as in water, to anchor the separate cationic and anionic scales. Recently, the absolute solvation free energy of the proton has been determined in methanol, acetonitrile, and dimethyl sulfoxide (DMSO),<sup>25</sup> and these absolute values can be used to determine databases of ionic solvation data in all three solvents. This now allows us to extend to nonaqueous solutions the strategy of adjusting Coulomb radii to fit ionic data and using the electrostatic model thusly parametrized even for neutrals where non-bulk-electrostatic effects are comparable to electrostatic ones.

In the present article we parametrize a new solvation model for both aqueous and nonaqueous solvents by using the Coulomb radii of SM6 for water and by parametrizing new Coulomb radii for nonaqueous solvents with the new database. The new model is called SM8. Note that SM7<sup>26</sup> denotes an unpublished universal solvation model in which the SM6 Coulomb radii are used in both aqueous and

nonaqueous media (the modeling strategy of employing the same Coulomb radii in all media was also used in SM5.2,<sup>27</sup> SM5.4,<sup>28,29</sup> SM5.42,<sup>30–32</sup> SM5C,<sup>33</sup> and SM5.43<sup>34</sup>); however, SM7 yielded some large errors for nonaqueous free energies of solvation of a subset of the ions.

One potential approach that could be used to parametrize the SM8 solvation model would be to develop individual sets of parameters for each solvent. For example, an adequate amount of experimental data exists in solvents like 1-octanol and hexadecane so that developing reasonably accurate sets of solvation parameters in these solvents would be possible. Indeed, two earlier SMx models for hexadecane<sup>28</sup> and for chloroform<sup>35</sup> used this approach. However, a major disadvantage that is associated with following this approach is that adequate experimental data do not exist in most other organic solvents, so that developing separate sets of solvation parameters in these solvents is not practical.

To circumvent this problem, a series of universal SMx models that can be applied to any solvent has been developed.<sup>18,26,30–34,36–39</sup> In the universal SMx models the solvation parameters are functions of a small set of solvent descriptors that are transferable to *any* condensed-phase medium. In this way, a single set of solvation parameters can be developed against a training set that includes experimental data in all solvents, including those solvents for which very little data exist.

In the most recent previously published universal solvation model, SM5.43,<sup>34,39</sup> the solvation parameters were optimized against a training set of data that contained 2237 solvation free energies for 295 solutes in 91 different solvents, including water, 79 transfer free energies between water and 12 organic solvents for an additional 51 solutes, and 47 aqueous solvation free energies for 47 ionic solutes. No experimental data for ionic solutes in nonaqueous solvents were included in this training set. Furthermore the aqueous ion data set used for SM5.43 is smaller than that used for SM6, which was parametrized only for aqueous solution. In this article, an updated version of the SM5.43 neutral training set and a new parametrization strategy involving a smaller number of parameters will be used to develop a new universal solvation model called SM8. It is especially noteworthy that the parametrization will involve new single-ion solvation free energies in acetonitrile, DMSO, and methanol.

All SMx solvation models except SM5C are based on discrete partial atomic charges, whereas the SM5C solvation model<sup>33</sup> was based on the continuous electronic density  $\rho(\mathbf{r})$  where  $\mathbf{r}$  denotes a point in space. The partial atomic charges in models SM1–SM5.2 were obtained by Mulliken<sup>40</sup> population analysis (which yields class II<sup>41</sup> charges), and those in models SM5.4, SM5.42, and SM5.43 were obtained by class IV charge models CM1,<sup>41</sup> CM2,<sup>42</sup> and CM3,<sup>43</sup> respectively. The SM8 model, like SM6<sup>18</sup> and SM7,<sup>26</sup> will be parametrized for the CM4<sup>18</sup> class IV charge model.

We have already mentioned one key respect in which SM8 differs from all previous SMx models, namely it involves intrinsic Coulomb radii adjusted to improve the solvation energies of ions in nonaqueous media. A second key difference is the catholicity of the parametrization. In solvation models SM1–SM7, there was a separate set of

solvation parameters for each electronic structure level, for example, separate sets for AM1, HF/6-31G(d), mPW1PW/6-31G(d), and mPW1PW/6-31+G(d,p) where AM1 denotes the Austin Model 1 semiempirical molecular orbital theory, HF denotes *ab initio* Hartree Fock theory, and 6-31+G(d,p)<sup>20</sup> is a basis set. The main reason for carrying out separate parametrizations is that the partial charges depend to some extent on the electronic structure level, and the parameters must be consistent with the partial atomic charges. However, this is true to a much greater extent for Mulliken<sup>40</sup> or Löwdin<sup>44–47</sup> charges than for class IV charges. Therefore in SM8 we will develop only a single set of solvation parameters that is designed to be used with any level of electronic structure theory that supports either the CM4 charge model or other comparably accurate charges. For example, it gives similar accuracy with any class IV charges, and we will show that it can also be used, although somewhat less accurately, with charges from population analysis. Although partial atomic charges are not physical observables, they can still be considered accurate within a given model context if they vary physically with molecular geometry and environment and can be used to accurately reproduce observables such as dipole moments or if they can be derived consistently and realistically from accurate experimental data, for instance, from the dipole moment of a diatomic molecule. The parameters of CM1 and CM2 depend on the specifics of the electronic structure level, but the parameters of CM3 are more general. They can be used with either HF theory or density functional theory (DFT), and they depend only on the basis set and the fraction of HF exchange. CM3 is also parametrized<sup>48</sup> for the self-consistent charge density-functional tight-binding model.<sup>49</sup> The parameters of CM4 are even more general and depend only on basis set. Thus CM3 and CM4 are parametrized for all density functionals (including hybrid ones with any amount of HF exchange), which means that SM8 can use class IV charges with all density functionals. Currently, CM4 parameter sets are available for the MIDI!6D<sup>50,51</sup> basis set and for Pople's<sup>20</sup> 6-31G(d), 6-31+G(d), and 6-31+G(d,p) basis sets. Additional CM4 parameters are under development and will be published soon.<sup>52</sup>

## 2. Description of the SM8 Universal Model

In the SMx models, the standard-state free energy of solvation  $\Delta G_s^0$  is partitioned according to

$$\Delta G_s^0 = \Delta E_E + \Delta E_N + \Delta G_{\text{conc}}^0 + G_p + G_{\text{CDS}} \quad (1)$$

where  $\Delta E_E$  is the change in the solute's internal electronic (*E*) energy in moving from the gas phase to the liquid phase at the same geometry,  $\Delta E_N$  is the change in the solute's internal energy due to changes in the equilibrium nuclear (*N*) positions in the solute that accompany the solvation process,  $\Delta G_{\text{conc}}^0$  (which is also called<sup>25,39,53</sup> the free energy of liberation or  $\Delta G_{\text{lib}}^0$ ) accounts for the concentration change between the gas-phase and the liquid-phase standard states,  $G_p$  is the polarization free energy, and  $G_{\text{CDS}}$  is the component of the free energy that is nominally associated with cavitation, dispersion, and solvent structure. Following the notation used



in previous SMx models, the sums  $\Delta E_E + G_P$  and  $\Delta E_E + \Delta E_N + G_P$  will be referred to as  $\Delta G_{EP}$  and  $\Delta G_{ENP}$ , respectively. Since the same concentration (1 mol/L) is used in both the gaseous and solution phases,  $\Delta G_{conc}^o$  is 0.<sup>53,54</sup> (If we used a gas-phase standard state of 1 atm,  $\Delta G_{conc}^o$  would be +1.9 kcal/mol.) All calculations reported here are based on gas-phase geometries (although the present model can be used to optimize geometries in the liquid phase<sup>55</sup>), so  $\Delta E_N$  is assumed to be 0 in this article, although not in the model in general. Since all free energies in this article are standard free energies, we will omit the standard-state modifier in most of the text for brevity.

The  $\Delta G_{EP}$  contribution to the total solvation free energy is calculated from a self-consistent molecular orbital calculation,<sup>30,56,57</sup> where the generalized Born approximation<sup>2,3,58–61</sup> is used to calculate the polarization contribution to the total free energy according to

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

In the above equation, the summations go over atoms  $k$  in the solute,  $\epsilon$  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  $k'$ .

As in the most recent previously published SMx solvation model, SM6,<sup>18</sup> the solvation parameters presented in this work are based on polarization free energies computed by eq 2 using CM4 partial atomic charges self-consistently polarized in solution. In CM4, the partial atomic charges are functions of the partial atomic charges obtained from a Löwdin population analysis<sup>44–47</sup> or a redistributed Löwdin population analysis (RLPA),<sup>62</sup> the gas-phase or liquid-phase Mayer bond orders,<sup>63–65</sup> and a set of atomic-number-dependent empirical parameters. These parameters have been optimized in earlier work and were chosen so as to minimize the errors between accurate gas-phase dipole moments and the dipole moments computed using gas-phase CM4 partial atomic charges. CM4 differs from the previous CMx model, CM3,<sup>43,48,66–68</sup> in that for hydrocarbons, CM4 is designed to accurately reproduce the partial atomic charges obtained from the Optimized Potentials for Liquid Simulations (OPLS) force field.<sup>69</sup> (Many of the hydrocarbons, e.g., ethane, in the CM4 training set do not have permanent dipole moments.)

One of the reasons it is preferable to optimize the parameters contained in the SMx solvation models using polarization free energies computed with class IV partial charges (as in CM4 and earlier CMx models<sup>41–43,48,66–68,70</sup>) is because these types of charge models are usually able to remove many of the systematic errors, in particular basis set dependence, that are present in partial atomic charges obtained from Mulliken,<sup>40</sup> Löwdin,<sup>44–47</sup> and redistributed Löwdin<sup>62</sup> population analyses. This helps to properly shift the focus of the modeling effort toward the various components of the solvation process. In addition, CM4 charges yield more accurate electrostatic potentials than population analysis charges, and this makes the solvation models more physical. It is worth noting that partial atomic charges obtained from any method can be used in eq 2 to compute polarization free

energies (see, for example, ref 71). However, one should be aware that in many cases, using different charge models can lead to very different partial atomic charges for a given molecule (and hence polarization free energies).<sup>72</sup> Because of this, it is recommended that, whenever possible, one should use the SM8 solvation parameters with CM4 partial atomic charges or with other charge models that have been validated to give partial atomic charges similar to those of CM4.

The Coulomb integrals  $\gamma_{kk'}$  are calculated according to ref 2

$$\gamma_{kk'} = [R_{kk'}^2 + \alpha_k \alpha_{k'} \exp(-R_{kk'}^2/d\alpha_k \alpha_{k'})]^{-1/2} \quad (3)$$

where  $R_{kk'}$  is the distance between atoms  $k$  and  $k'$ , and  $\alpha_k$  is the effective Born radius of atom  $k$ , which is described below. In the above equation,  $d$  is an empirical constant that is usually set equal to 4 (this value was originally proposed by Still et al.,<sup>2</sup> 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 medium), although during the development of SM6 and some earlier solvation models, it was found that optimizing this parameter increased the accuracy. In SMx models prior to SM6, when  $d$  was not set equal to 4 for all  $k$  and  $k'$ ,<sup>27,29,30,32–35,38,39,73–75</sup> it was always set equal to 4 or 3.9 except for the case where one of the atoms  $k$  and  $k'$  is carbon and the other is hydrogen. In SM6,  $d$  was made independent of  $k$  and  $k'$ , and it was optimized to the value of 3.7. We also used that value in SM7, and we will also use it in SM8.

The Born radius is calculated by<sup>76</sup>

$$\alpha_k = \left( \frac{1}{R'} + \int_{\rho_{Z_k}}^{R'} \frac{A_k(\mathbf{R}, r, \{\rho_Z\})}{4\pi r^4} dr \right)^{-1} \quad (4)$$

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  $A_k(\mathbf{R}, r, \{\rho_Z\})$  is the exposed area<sup>76</sup> of a sphere of radius  $r$  that is centered on atom  $k$ . This area depends on the geometry of the solute,  $\mathbf{R}$ , 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  $\rho_{Z_k}$  that depend on the atomic number  $Z_k$  of the atom  $k$ .

The final term on the right-hand-side of eq 1,  $G_{CDS}$ , is the first-solvation-shell contribution to the solvation free energy. Examples of first-solvation-shell effects include, but are not limited to, cavitation (C), dispersion (D), and structural (S) effects of solvent molecules in the first solvation shell. In SM8,  $G_{CDS}$  is given by

$$G_{CDS} = \sum_k^{\text{atoms}} \sigma_k A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) + \sigma^{[M]} \sum_k^{\text{atoms}} A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) \quad (5)$$

where  $\sigma_k$  and  $\sigma^{[M]}$  are the atomic and the molecular surface tensions of atom  $k$ , respectively, and  $A_k$  is the solvent-

accessible surface area (SASA)<sup>77,78</sup> of atom  $k$ . The SASA depends on the geometry  $\mathbf{R}$ , the set  $\{R_{Z_k}\}$  of all atomic van der Waals radii, and the solvent radius  $r_s$ , which is added to each of the atomic van der Waals radii. Adding a nonzero value for solvent radius to the atomic radii defines the spheres that are used to compute the SASA of a given solute.<sup>76</sup> Notice that the van der Waals radii used in the SASA calculation are not the same as the intrinsic Coulomb radii used in eq 4; in fact we use Bondi's values<sup>79</sup> for the van der Waals radii.

The atomic surface tensions are given by

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{k'}^{\text{atoms}} \tilde{\sigma}_{Z_k Z_{k'}} T_k(\{Z_{k'}, R_{kk'}\}) \quad (6)$$

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  $T_k(\{Z_{k'}, R_{kk'}\})$  is a geometry-dependent switching function called a cutoff tanh, or COT; this function is described in a previous publication.<sup>18</sup> For H, C, N, O, F, P, S, Cl, and Br, SM8 uses the same functional forms  $T_k$  as does SM6. The atomic surface tensions for these atoms were also presented in the previous publication.<sup>18</sup> For SM8, an additional atomic surface tension for Si was added; this atomic surface tension is set equal to the atomic-number-specific parameter for Si (i.e., this atomic surface tension does not include any COT functions)

$$\sigma_{Z|Z=14} = \tilde{\sigma}_Z \quad (7)$$

As in previous SMx universal solvation models, in SM8 the atomic surface tensions are made to depend on the solvent by making the parameters  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$  functions of a set of solvent descriptors. This dependence is given by

$$\tilde{\sigma}_i = \tilde{\sigma}_i^{[n]} n + \tilde{\sigma}_i^{[\alpha]} \alpha + \tilde{\sigma}_i^{[\beta]} \beta \quad (8)$$

where  $\tilde{\sigma}_i$  is either  $\tilde{\sigma}_Z$  or  $\tilde{\sigma}_{ZZ'}$ ,  $n$  is the refractive index of the solvent at room temperature (which is conventionally taken as 293 K for this quantity),  $\alpha$  is Abraham's<sup>80–83</sup> hydrogen bond acidity parameter of the solvent (which Abraham denotes as  $\Sigma\alpha_2$ ),  $\beta$  is Abraham's hydrogen bond basicity parameter of the solvent (which Abraham denotes as  $\Sigma\beta_2$ ), and  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ , and  $\tilde{\sigma}_i^{[\beta]}$  are empirical parameters that depend on  $i$ . (Note that Abraham developed  $\Sigma\alpha_2$  and  $\Sigma\beta_2$  as solute descriptors, but we use them as solvent descriptors.) Besides making the atomic surface tensions depend on the solvent through the use of eq 8, SM8 also uses a molecular surface tension that is multiplied by the total SASA of the given solute (see eq 5; the total SASA of the solute is equal to the sum of the SASAs of each of the individual atoms in the solute). The molecular surface tension is also a function of solvent descriptors, and it is given by

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]} \left( \frac{\gamma}{\gamma_o} \right) + \tilde{\sigma}^{[\phi^2]} \phi^2 + \tilde{\sigma}^{[\psi^2]} \psi^2 + \tilde{\sigma}^{[\beta^2]} \beta^2 \quad (9)$$

where  $\gamma$  is the macroscopic surface tension of the solvent at air/solvent interface at 298.15 K expressed in  $\text{cal mol}^{-1} \text{\AA}^{-2}$  (note that  $1 \text{ dyn/cm} = 1.43932 \text{ cal mol}^{-1} \text{\AA}^{-2}$ ),  $\gamma_o = 1 \text{ cal mol}^{-1} \text{\AA}^{-2}$ ,  $\phi^2$  is the square of the fraction of solvent atoms that are aromatic carbon atoms (carbon aromaticity),  $\psi^2$  is

the square of the fraction of solvent atoms that are F, Cl, or Br (electronegative halogenicity),  $\beta^2$  is the square of Abraham's hydrogen bond basicity parameter of the solvent, and  $\tilde{\sigma}^{[\gamma]}$ ,  $\tilde{\sigma}^{[\phi^2]}$ ,  $\tilde{\sigma}^{[\psi^2]}$ , and  $\tilde{\sigma}^{[\beta^2]}$  are empirical parameters that are independent of the solute.

The chosen solvent descriptors are physically meaningful.<sup>27,75</sup> For example, the refractive index  $n$  is a measure of solvent's polarizability, which in turn is related to dispersion interactions of the solvent. The acidity and basicity parameters are related to the solvent's ability to donate and accept hydrogen bonds, respectively. The solvent's macroscopic surface tension represents the energy required for cavitation (creation of a surface) in the solvent. The aromaticity and electronegative halogenicity factors are used to account for systematic differences in intermolecular interactions in aromatic solvents and solvents containing electronegative halogen atoms.

In SM8 as well as in previous universal solvation models water is treated as a special solvent that is given its own set of  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$  values, so that eq 8 is not needed for water. Also, for water, the molecular surface tension  $\sigma^{[M]}$  is set equal to zero. Thus, when SM8 is used to compute solvation free energies in aqueous solvent, eq 5 reduces to

$$G_{\text{CDS,water}} = \sum_k^{\text{atoms}} \sigma_k A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) \quad (10)$$

where  $\tilde{\sigma}_Z$  or  $\tilde{\sigma}_{ZZ'}$  used in eq 6 to obtain  $\sigma_k$  are simply numbers that do not depend on solvent descriptors.

### 3. Parameters to be Optimized

During the development of SM6,<sup>18</sup> the parametrization effort was focused on two types of parameters: (1) the atomic radii used in eq 4 and (2) the parameters  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$  used in eq 10. For previous universal SMx models, it has been shown that using solvent-independent values for the intrinsic Coulomb radii, the van der Waals radii, and the solvent radius  $r_s$  leads to relatively accurate solvation free energies in both water and nonaqueous solvents; that is, it was assumed that the solvent dependence of the solvation free energy is entirely contained in  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$ . This assumption is too restrictive for the present work, and we will allow the intrinsic atomic Coulomb radii to depend on solvent in nonaqueous solvents while retaining the SM6 values in water. Then the parameters to be determined are a solvent-dependent set of intrinsic atomic Coulomb radii for nonaqueous solvents and the full set of atomic surface tensions, namely  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ , and  $\tilde{\sigma}_i^{[\beta]}$  that appear in eq 8 and  $\tilde{\sigma}^{[\gamma]}$ ,  $\tilde{\sigma}^{[\phi^2]}$ ,  $\tilde{\sigma}^{[\psi^2]}$ , and  $\tilde{\sigma}^{[\beta^2]}$  that appear in eq 9. The  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$  parameters for water also are to be optimized as part of this work. The symbols  $\tilde{\sigma}_i^{[\text{water}]}$  will be used to denote the  $\tilde{\sigma}_Z$  and  $\tilde{\sigma}_{ZZ'}$  parameters that are optimized specifically for water, where  $i$  is either  $Z$  or  $ZZ'$ . For Si, which was not included in the SM5.43R<sup>34,39</sup> or the SM6 parametrizations, Bondi's value<sup>79</sup> of  $2.10 \text{ \AA}$  will be used in eq 4 for water and also to compute the SASA for all solvents; this is the same value for the atomic radius that was used by a previous universal SMx model that included Si.<sup>75</sup>

#### 4. SM8 Universal Model Training Set

**Standard States.** In this article, all gas-phase free energies use a standard-state gas-phase pressure of 1 atm. All solvation free energies in the present article are tabulated for the gas-phase solute having a standard state of an ideal gas at a gas-phase concentration of 1 mol/L and for the liquid-phase solute being dissolved in an ideal solution at a liquid-phase concentration of 1 mol/L. Transfer free energies between water and organic solvents use a standard state for which the concentration is equal in both phases.

**Experimental Data for Neutrals.** The present training set begins with the portions of the SM5.43 and SM6 training sets that contain experimental aqueous solvation free energies of neutral solutes. This subset contains aqueous solvation free energies for 273 neutral solutes (including the water dimer) containing one or more of the elements H, C, N, O, F, P, S, Cl, or Br. To this subset of data was added the experimental aqueous solvation free energy of tetramethylsilane, which was taken from an earlier training set.<sup>75</sup>

The nonaqueous portion of the SM5.43 training set<sup>34,39</sup> contains 1980 solvation free energies for 263 solutes containing one or more of the elements H, C, N, O, F, P, S, Cl, or Br, in 90 organic solvents. This training set also contains 79 transfer free energies between water and 12 organic solvents, which are a subset of the 90 organic solvents. These transfer free energies were determined directly from experimental partition coefficients according to

$$\Delta G_{o/w}^o = -2.303RT \log P_{o/w} \quad (11)$$

where  $\Delta G_{o/w}^o$  is the standard-state free energy associated with transferring the solute from the aqueous phase w to the organic phase o, and  $P_{o/w}$  is the corresponding partition coefficient, which is given by

$$P_{o/w} = \frac{[\text{solute}]_o}{[\text{solute}]_w} \quad (12)$$

where  $[\text{solute}]_o$  is the equilibrium concentration of the solute in the organic phase, and  $[\text{solute}]_w$  is the equilibrium concentration of the solute in the aqueous phase. Transfer free energy data are included in this as well as several previous SMx training sets,<sup>33,34,39,75</sup> because for many solutes the experimental data that are required to determine the solvation free energy between the gas and liquid phases are not available. Thus, if one were restricted to considering only solvation free energies, many solutes containing important functionality would not be well represented (or not represented at all) in the training set. It is worth noting that many of the solvation free energies in the SM5.43 training set are derived from experimental partition coefficients and experimental aqueous solvation free energies, that is

$$\Delta G_{o/a}^o = \Delta G_{w/a}^o - 2.303RT \log P_{o/w} \quad (13)$$

where  $\Delta G_{o/a}^o$  is the solvation free energy in the organic solvent o (a denotes the gas phase, or air),  $\Delta G_{w/a}^o$  is the aqueous solvation free energy, and  $P_{o/w}$  is the partition coefficient.

Before incorporating the experimental data taken from the SM5.43 training set into the current training set, several updates and corrections were made to these data. For nitromethane, the SM5.43 training set contains solvation free energies in carbon tetrachloride and cyclohexane as well as transfer free energies between water and carbon tetrachloride and water and cyclohexane. The two transfer free energies are redundant and were removed. Using an experimental value for the aqueous solvation free energy of nitromethane,<sup>18</sup> the transfer free energy of nitromethane between octanol and water was converted to the solvation free energy of nitromethane in octanol using eq 13. Similarly, the transfer free energies of 3,5-dimethylpyridine between benzene and water, 4-ethylpyridine between octanol and water,  $\gamma$ -butyrolactone between octanol and water, pyrrole between chloroform and water, octanol and water, and cyclohexane and water, and quinoline between chloroform and water, octanol and water, and cyclohexane and water were converted to solvation free energies in the above organic solvents using eq 13 and experimental values<sup>18,84</sup> for the aqueous solvation free energies of these solutes. During the development of SM6, the experimental value for the aqueous solvation free energy of hydrazine was updated from  $-9.30$  kcal/mol to  $-6.26$  kcal/mol. In the SM5.43 training set, experimental values for the solvation free energies of hydrazine in benzene, octanol, diethyl ether, and chloroform were determined using experimental partition coefficients and an experimental value of  $-9.30$  kcal/mol for the aqueous solvation free energy of hydrazine in eq 13; these solvation free energies were adjusted to reflect the above update in the aqueous solvation free energy of hydrazine. Finally, in the SM5.43 training set, the experimental value for the transfer free energy of phenylurea between chloroform and water is incorrectly listed as  $-0.86$  kcal/mol. The current training set uses the correct value, which is  $+0.86$  kcal/mol.<sup>85</sup>

To the above subset of data, experimental partition coefficients<sup>85</sup> and experimental aqueous solvation free energies were used to add 80 solvation free energies in organic solvents for the following 14 solutes: hydrogen peroxide, urea, benzamide, methylhydrazine, 2-methylaniline, 3-methylaniline, 4-methylaniline, *N*-methylaniline, *N*-methyl-2-pyrrolidinone, 2-pyrrolidinone, formamide, *N,N*-dimethylformamide, *N*-methylformamide, and *N,N*-dimethylacetamide. Sixty-three relative solvation free energies between water and organic solvents for 31 solutes, the majority of which contain amide groups, were also added. Finally, experimental values for the solvation free energy of tetramethylsilane in hexadecane and in octanol and experimental values for the transfer free energies of 13 other solutes containing Si between water and octanol were added. These experimental data were taken from an earlier training set.<sup>75</sup>

Combining all of the data from above results in a total of 2346 solvation free energies for 318 neutral solutes in 91 solvents (including water) and 143 transfer free energies for 93 neutral solutes between water and 15 organic solvents. These data will be referred to as the SM8 universal model neutral training set. Note that this training set does not contain any ionic solutes. Experimental data for ionic solutes will be discussed below.



**Table 1.** 90 Nonaqueous Solvents in the SM8 Neutral Training Set<sup>a</sup>

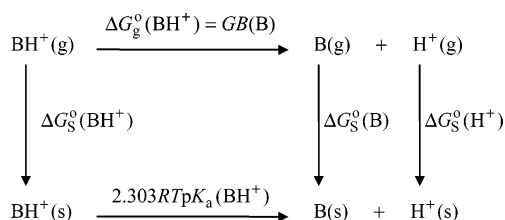
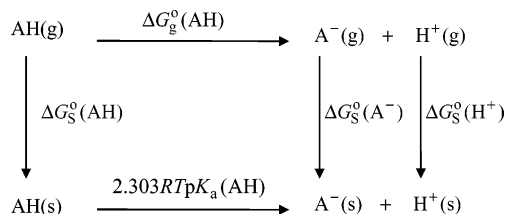
acetic acid	<i>1,2-dibromoethane</i>	methoxyethanol
acetonitrile*	<i>dibutyl ether</i>	methylene chloride*
acetophenone	<i>o</i> -dichlorobenzene	methylformamide
aniline*	<i>1,2-dichloroethane</i> *	4-methyl-2-pentanone
anisole	<i>diethyl ether</i> *	2-methylpyridine
<i>benzene</i> *	diisopropyl ether	<i>nitrobenzene</i>
benzonitrile	<i>N,N</i> -dimethylacetamide	nitroethane
benzyl alcohol	<i>N,N</i> -dimethylformamide	nitromethane*
bromobenzene	2,6-dimethylpyridine	<i>o</i> -nitrotoluene
bromoethane	dimethyl sulfoxide*	nonane
bromoform	dodecane	nonanol
bromooctane	ethanol*	octane
<i>1-butanol</i>	ethoxybenzene	<i>octanol</i>
<i>2-butanol</i>	<i>ethyl acetate</i>	pentadecane
butanone	ethylbenzene	pentane
butyl acetate	fluorobenzene	pentanol
<i>n</i> -butylbenzene	1-fluoro- <i>n</i> -octane	perfluorobenzene
<i>sec</i> -butylbenzene	<i>heptane</i> *	phenyl ether
<i>t</i> -butylbenzene	heptanol	propanol
carbon disulfide	hexadecane	pyridine
<i>carbon tetrachloride</i> *	hexadecyl iodide	tetrachloroethene
<i>chlorobenzene</i> *	<i>hexane</i>	tetrahydrofuran*
<i>chloroform</i> *	hexanol	tetrahydrothiophene dioxide
chlorohexane	iodobenzene	tetralin
<i>m</i> -cresol	isobutanol	toluene*
<i>cyclohexane</i> *	isooctane	tributylphosphate
cyclohexanone	isopropanol	triethylamine
decalin (mixture)	isopropylbenzene	1,2,4-trimethylbenzene
decane	<i>p</i> -isopropyltoluene	undecane
decanol	mesitylene	xylene (mixture)

<sup>a</sup> Methanol is not included in this training set because there are no data for neutral solutes in methanol. The asterisk denotes the nonaqueous solvents presently available with the default solvation model implemented in *Gaussian 03* in addition to methanol and water. The names of 15 solvents for which we used solvent–water transfer free energies are italicized.

Table 1 shows the 90 organic solvents used in the SM8 parametrization. A table containing experimental values for the 2346 solvation free energies and 143 transfer free energies contained in the SM8 universal model neutral training set is given in the Supporting Information. Also included in the Supporting Information are calculated values obtained using the SM8 model described below.

**Solvation Free Energies of Ions in Water.** The current ion training set for ions in aqueous solution was explained in the article where SM6 was parametrized.<sup>18</sup> In particular we use the data set called the selectively clustered set. In this set, there are 112 ions; 81 of these are unclustered and 31 are clustered with a single water molecule each (these ions are not included in unclustered form). The criterion for whether to cluster an ion is that it is clustered if it contains three or fewer atoms, or if the partial charge on any oxygen is more negative than the partial charge on oxygen in water, or if the ion is an oxonium or ammonium cation. The rationale for this criterion is explained in the SM6 paper.<sup>18</sup>

**Single-Ion Solvation Free Energies of Unclustered Ions in Acetonitrile, DMSO, and Methanol.** In previous work, the cluster pair approximation was used to estimate the

**Scheme 1.** Thermochemical Cycle Relating the Solvation Free Energy of BH<sup>+</sup> to the Gas-Phase Basicity (GB) of the Base B**Scheme 2.** Thermochemical Cycle Relating the Solvation Free Energy of A<sup>−</sup> to the Gas-Phase Acidity of the Acid AH

absolute solvation free energy of the proton in acetonitrile, DMSO, and methanol.<sup>25</sup> These values can be combined with experimental or calculated data to determine absolute solvation free energies of other ionic solutes in these solvents. For example, using thermochemical cycle 1 (illustrated in Scheme 1) the absolute solvation free energy of the cation BH<sup>+</sup> can be written as

$$\Delta G_{\text{S}}^{\circ}(\text{BH}^+) = \Delta G_{\text{g}}^{\circ}(\text{BH}^+) + \Delta G_{\text{S}}^{\circ}(\text{B}) - 2.303RT\text{p}K_{\text{a}}(\text{BH}^+) + \Delta G_{\text{S}}^{\circ}(\text{H}^+) + \Delta G^{\text{1atm} \rightarrow \text{1M}} \quad (14)$$

where  $\Delta G_{\text{g}}^{\circ}(\text{BH}^+)$  is the gas-phase acidity of BH<sup>+</sup>, which is equal to

$$\Delta G_{\text{g}}^{\circ}(\text{BH}^+) = G^{\circ}(\text{B}) + G^{\circ}(\text{H}^+) - G^{\circ}(\text{BH}^+) \quad (15)$$

$\Delta G_{\text{S}}^{\circ}(\text{B})$  is the solvation free energy of the neutral species B,  $\text{p}K_{\text{a}}$  is the negative common logarithm of the solution-phase acid dissociation constant of BH<sup>+</sup>, and  $\Delta G^{\text{1atm} \rightarrow \text{1M}}$  is the free energy change associated with moving from a gas-phase pressure of 1 atm to a liquid-phase concentration of 1 M. Similarly, thermochemical cycle 2 (illustrated in Scheme 2) can be used to write the absolute solvation free energy of the anion A<sup>−</sup>

$$\Delta G_{\text{S}}^{\circ}(\text{A}^-) = -\Delta G_{\text{g}}^{\circ}(\text{AH}) + \Delta G_{\text{S}}^{\circ}(\text{AH}) + 2.303RT\text{p}K_{\text{a}}(\text{AH}) - \Delta G_{\text{S}}^{\circ}(\text{H}^+) - \Delta G^{\text{1atm} \rightarrow \text{1M}} \quad (16)$$

In previous work, the above equations were combined with experimental  $\text{p}K_{\text{a}}$  values, gas-phase acidities, and aqueous solvation free energies of neutral species in order to determine single-ion solvation free energies in aqueous solution. However, for the solvents acetonitrile, DMSO, and methanol, finding solutes for which experimental  $\text{p}K_{\text{a}}$  values, gas-phase acidities, and solvation free energies (of the neutral species) exist is challenging. In particular, for those solutes where experimental  $\text{p}K_{\text{a}}$  values in one of the solvents above and gas-phase acidities are available, experimental solvation free energies in the given solvent are typically not available.

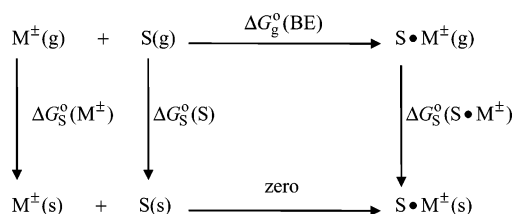
**Table 2.** Reference Free Energies of Solvation of Selected Ions (kcal/mol)<sup>a</sup>

neutral molecule (AH or B)	charge	$\Delta G_g^0$	$\Delta G_s^0$ (neutral)	$pK_a$	$\Delta G_s^0$ (ion)
Acetonitrile					
ammonia	+1	195.7	-4.29	16.5 <sup>90</sup>	-89.3
pyridine	+1	214.7	-6.34	12.3 <sup>89,90</sup>	-66.7
acetic acid	-1	341.4	-6.04	22.3 <sup>98,106,107</sup>	-58.8
phenol	-1	342.9	-7.20	27.0 <sup>97,113</sup>	-55.1
DMSO					
ammonia	+1	195.7	-3.95	10.5 <sup>114</sup>	-93.9
pyridine	+1	214.7	-5.71	3.5 <sup>114</sup>	-67.2
acetic acid	-1	341.4	-5.95	12.3 <sup>114</sup>	-59.2
phenol	-1	342.9	-7.22	18.0 <sup>114</sup>	-54.2
Methanol					
ammonia	+1	195.7	-5.05	10.8 <sup>125</sup>	-85.6
pyridine	+1	214.7	-6.57	5.4 <sup>125,126</sup>	-60.8
acetic acid <sup>b</sup>	-1	341.4	-6.25	9.7 <sup>110,125,128</sup>	-72.9
phenol	-1	342.9	-7.60	14.4 <sup>125,131,139</sup>	-69.3

<sup>a</sup> The free energies of solvation for all ions and the auxiliary data are listed in the Supporting Information. For cations  $BH^+$ ,  $\Delta G_g^0$  is the gas-phase acidity of  $BH^+$  equal to the gas-phase basicity<sup>87</sup> of the neutral base B (see eq 14 and Scheme 1). For anions  $A^-$ ,  $\Delta G_g^0$  is the gas-phase acidity<sup>88</sup> of the neutral acid AH (see eq 16 and Scheme 2). The free energies of solvation for neutral species  $\Delta G_s^0$ (neutral) are calculated at the SM7/mPW1PW/6-31G(d) level of theory.<sup>26</sup> The values of  $pK_a(BH^+)$  and  $pK_a(AH)$  are reference data. In case of multiple references, the  $pK_a$  values are averaged over the references. The absolute free energies of ions  $\Delta G_s^0$ (ion) are based on the following values<sup>25</sup> for the absolute solvation free energy of the proton in the three solvents: -260.2 (acetonitrile), -273.3 (DMSO), and -263.5 (methanol) kcal/mol. <sup>b</sup> See also refs 130, 131, and 133 for  $pK_a$ .

The SM8 universal solvent model as well as earlier SM $x$  universal solvent models can predict solvation free energies of neutral solutes in nonaqueous media to an accuracy of  $\sim 0.6$  kcal/mol. Thus, calculated instead of experimental values can be used for the solvation free energies of neutral solutes in the above equations to obtain relatively reasonable estimates of the solvation free energies of single ions. To do this, a data set<sup>86</sup> of experimental gas-phase acidities<sup>87,88</sup> and experimental  $pK_a$  values<sup>89–139</sup> in acetonitrile, DMSO, and methanol was created. For each of the species, solvation free energies were calculated at the SM7/mPW1PW/6-31G(d) level of theory.<sup>26</sup> Using these experimental and calculated data, single-ion solvation free energies were determined in acetonitrile, DMSO, and methanol. The resulting free energies for all the ions and the auxiliary data used to determine them are listed in the Supporting Information. Table 2 shows examples of these data for a few typical ions in the three solvents.

**Clustered Ions in Nonaqueous Solvents.** For all four cations in DMSO and for any ion in acetonitrile and DMSO that contains one or more halogen atoms, we calculated solvation free energies of clustered ions (with a single solvent molecule in the cluster) by using precisely the same procedure as used previously<sup>18,25</sup> to obtain solvation free energies of clustered ions in water. This procedure is illustrated in cycle 3 (see Scheme 3) that relates the solvation free energy of a clustered ion to the gas-phase binding free

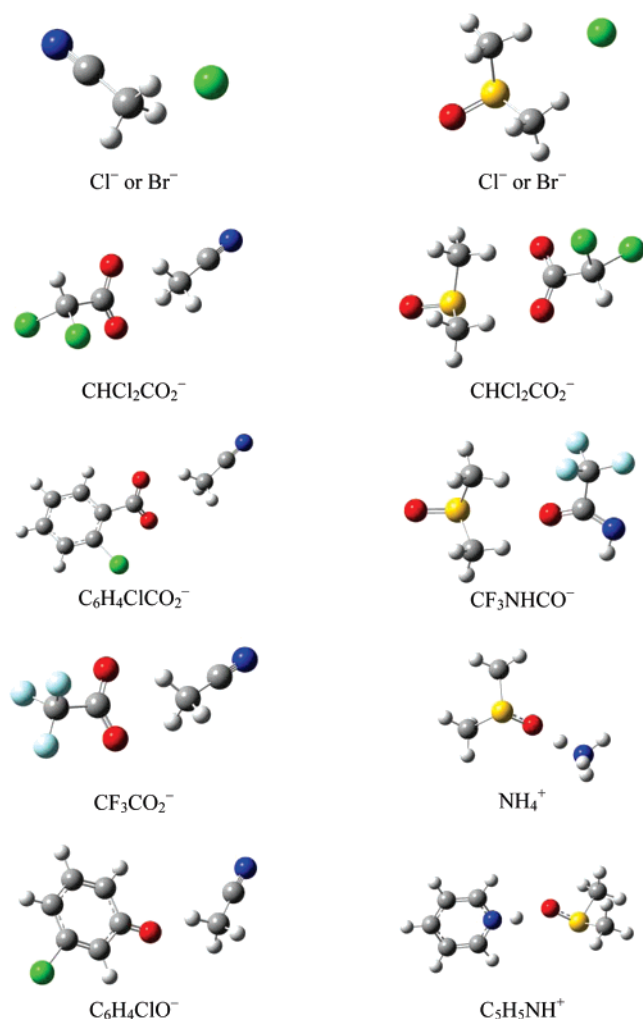
**Scheme 3.** Thermochemical Cycle Relating the Solvation Free Energy of an Ionic Solute  $M^\pm$  to the Solvation Free Energy of the Solvent–Solute Cluster  $S \cdot M^\pm$ **Table 3.** Solvation Free Energies of Solvent–Solute Clusters Used in Optimization of the SM8 Coulomb Radii<sup>a</sup>

neutral molecule (AH or B)	ion	$\Delta G_g^0$ (BE) <sup>b</sup>	$\Delta G_s^0$ (bare ion) <sup>c</sup>	$\Delta G_s^0$ (cluster) <sup>c</sup>
Acetonitrile Clusters				
hydrochloric acid	$Cl^-$	-9.5 <sup>d</sup>	-62.4	-55.9
hydrobromic acid	$Br^-$	-8.7 <sup>d</sup>	-59.3	-53.6
trifluoroacetic acid	$CF_3CO_2^-$	-6.3	-45.6	-42.2
3-trifluoromethylphenol	$CF_3C_6H_4O^-$	-5.0	-46.9	-44.8
chloroacetic acid	$CH_2ClCO_2^-$	-7.2	-54.6	-50.4
2-chlorobenzoic acid	$C_6H_4ClCO_2^-$	-5.7	-53.5	-50.8
3-chlorophenol	$C_6H_4ClO^-$	-5.8	-50.6	-47.7
dichloroacetic acid	$CHCl_2CO_2^-$	-5.7	-51.2	-48.5
3,4,5-trichlorophenol	$C_6H_2Cl_3O^-$	-4.0	-43.8	-42.7
DMSO Clusters				
ammonia	$NH_4^+$	-29.1 <sup>e</sup>	-93.9	-70.6
aniline	$C_6H_5NH_3^+$	-18.9 <sup>e</sup>	-79.8	-66.7
methylamine	$CH_3NH_3^+$	-23.1 <sup>e</sup>	-82.4	-65.0
pyridine	$C_5H_5NH^+$	-18.3 <sup>e</sup>	-67.2	-54.6
hydrochloric acid	$Cl^-$	-12.5 <sup>d</sup>	-62.7	-55.9
hydrobromic acid	$Br^-$	-10.9 <sup>d</sup>	-57.8	-52.6
2,2,2-trifluoroethanol	$CF_3CH_2O^-$	-11.3	-56.1	-50.6
trifluoroacetamide	$CF_3NHCO^-$	-5.9	-49.2	-49.1
trifluoroacetic acid	$CF_3CO_2^-$	-5.7	-45.0	-45.0
2-chlorobenzoic acid	$C_6H_4ClCO_2^-$	-6.4	-53.6	-52.9
4-chlorobenzoic acid	$C_6H_4ClCO_2^-$	-7.0	-52.6	-51.3
dichloroacetic acid	$CHCl_2CO_2^-$	-5.6	-49.2	-49.3

<sup>a</sup> All entries (in kcal/mol) are given for 298.15 K. <sup>b</sup> Solvent–solute binding free energies (BE), calculated in this work at the B97-1/MG3S level of theory, unless indicated otherwise. <sup>c</sup> Taken from Tables S1–S3 in the Supporting Information (part II). <sup>d</sup> Experimental energies.<sup>143</sup> <sup>e</sup> B97-1/MG3S energies.<sup>25</sup>

energy of the cluster. The binding free energies were calculated at the B97-1<sup>140</sup>/MG3S<sup>141</sup> level of theory using *Gaussian 03*,<sup>142</sup> except for the clusters of  $Cl^-$  and  $Br^-$ , for which the experimental binding free energies were available in standard reference data.<sup>143</sup> The molecular geometries for all of the clusters were optimized, and conformational analysis was carried out by calculating harmonic frequencies to verify the nature of minima and by searching to find the global minimum conformations in the gas phase. The primary reason for adding the clustered-ion data was to increase the number of data to achieve a more robust fit. The resulting solvation free energies of clusters are given in Table 3. The molecular structures of a few typical clusters are depicted in Figure 1. The Cartesian coordinates corresponding to the B97-1/MG3S optimized global minima for all of the nonaqueous clusters used in this work are given in the Supporting Information.





**Figure 1.** Clusters of selected ions with acetonitrile and DMSO.

**Molecular Geometries of Solutes.** All computed solvation free energies in this article are based on rigid, gas-phase geometries. The molecular geometries of all unclustered neutral and ionic solutes are optimized at the *m*PW1PW/MIDI<sup>50,51,144</sup> level of electronic structure theory.

The use of gas-phase optimized geometries allows us to save computational time without any significant loss of accuracy in the SM8 parametrization. Indeed, because most solutes considered here prefer similar conformations and structures in the gas phase and solution the difference in solvation free energy between using gas-phase geometries and using liquid-phase geometries is smaller than the mean error of the model in many cases. Having obtained the parameters with such a training set, they can be used more broadly in further applications, for instance for the liquid-phase geometry optimization when the solute's geometry is expected to change significantly upon passing a solute from the gas phase to solution.

## 5. Parametrization

The first step was to parametrize the intrinsic Coulomb radii. The intrinsic Coulomb radii for aqueous solution are frozen at the values that were optimized for the SM6 model in previous work.<sup>18</sup> The intrinsic Coulomb radii for nonaqueous

**Table 4.** Solvent Acidity and Basicity Descriptors for the Four Solvents with Ionic Data<sup>a</sup>

solvent	$\alpha$	$\beta$
acetonitrile	0.07	0.32
DMSO	0	0.88
methanol	0.43	0.47
water	0.82	0.35

<sup>a</sup>  $\alpha$  is Abraham's<sup>80–83</sup> hydrogen bond acidity parameter (which Abraham denotes as  $\Sigma\alpha_2$ ), and  $\beta$  is Abraham's hydrogen bond basicity parameter (which Abraham denotes as  $\Sigma\beta_2$ ).

solution were optimized using the solvation free energies of ions in acetonitrile, DMSO, and methanol in Tables 2 and S1–S3 in part II of the Supporting Information. A number of schemes were tested in which the intrinsic Coulomb radii of various sets of atoms were optimized as functions of the solvent hydrogen-bond acidity  $\alpha$  and hydrogen-bond basicity  $\beta$ . The values of these solvent descriptors for the four solvents including water in which we have ionic data are listed in Table 4. After considerable trial and error we concluded that there was no reason to change any of the intrinsic Coulomb radii from their water values in methanol and no reason to change the carbon or nitrogen intrinsic Coulomb radii in any solvent. We also found that for the three nonaqueous solvents tested, making the Coulomb radii functions of  $\beta$  had little effect on the overall accuracy of the model. Thus we settled on the scheme

$$\rho_Z = \begin{cases} \rho_Z(\text{water}) & \alpha \geq 0.43 \\ \rho_Z(\text{water}) + a(0.43 - \alpha) & \alpha < 0.43 \end{cases} \quad (17)$$

with  $a$  as a parameter, and we constrained  $a$  to zero for  $Z = 6$  and 7. There is not enough data to optimize  $a$  for  $Z = 15$ , so it was set equal to the value for  $Z = 16$ . The optimized radii are given in Table 5 where they are compared to some previous SMx intrinsic Coulomb radii<sup>18,26,30–32,34</sup> and to the van der Waals radii of Bondi. (The other columns in this table will be explained below.) The free energies of solvation for ions in acetonitrile and DMSO calculated by SM8 and SM7 are compared to the corresponding reference data in Tables S4 and S5 in part II of the Supporting Information.

A technical point should be mentioned here. The optimum intrinsic Coulomb radii actually depend slightly on the atomic surface tensions (whereas the atomic surface tensions depend strongly on the intrinsic Coulomb radii). Thus we optimized the intrinsic Coulomb radii for ions with SM7 atomic surface tensions,<sup>26</sup> then determined a first round of SM8 surface tensions using neutral data, then reoptimized the intrinsic Coulomb radii for ions with these atomic and molecular surface tensions, and then found the final atomic surface tensions by a final round of calculations on neutrals.

To begin the parametrization, of the atomic surface tensions,  $\Delta G_{EP}$  values were calculated for all of the solutes in the SM8 universal model training set for which solvation free energies are available (total of 2346 calculations). The  $\Delta G_{EP}$  values for all of the solutes in the training set for which transfer free energies are available, in water, and in the organic solvent to which the transfer free energy refers, were

**Table 5.** Intrinsic Coulomb Radii (Å) of Various Models and Bondi's van der Waals Radii (Å)<sup>a</sup>

atom	Z	SM8			SM6/SM7 <sup>b</sup>	SM5.43 <sup>c</sup>	SM5.42 <sup>d</sup>	C-PCM GAMESS	PB Jaguar	GCOSMO NWChem	Bondi <sup>e</sup>
		<i>a</i>	$\rho_Z(\text{water})$	$\rho_Z(\text{DMSO})$							
H	1	-0.52	1.02	0.80	1.02	0.79	0.91	1.44	1.15	1.20	1.20
C	6	0	1.57	1.57	1.57	1.81	1.78	2.04	1.90	1.50	1.70
N	7	0	1.61	1.61	1.61	1.66	1.92	1.92	1.60	1.50	1.55
O	8	1.54	1.52	2.18	1.52	1.63	1.60	1.80	1.60	1.40	1.52
F	9	2.69	1.47	2.63	1.47	1.58	1.50	1.62	1.68	1.35	1.47
Si	14	0	2.10	2.10	2.10	2.10	2.10	2.40	2.15	1.17	2.10
P	15	0.77	1.80	2.13	1.80	2.01	2.27	2.28	2.07	1.80	1.80
S	16	0.77	2.12	2.45	2.12	2.22	1.98	2.22	1.90	1.75	1.80
Cl	17	1.42	2.02	2.63	2.02	2.28	2.13	2.17	1.97	1.70	1.75
Br	35	0.59	2.60	2.85	2.60	2.38	2.31	2.34	2.10	1.80	1.85

<sup>a</sup> The SM8 parameters *a* (eq 17) were optimized for H, O, F, S, Cl, and Br. The value of *a*(P) was fixed at the *a*(S) value. The SMx model radii are compared to the default values of radii used in the Conductor-like Polarizable Continuum Model (C-PCM/GAMESS) as implemented in GAMESS, in Jaguar's Poisson–Boltzmann self-consistent reaction field solver (PB/Jaguar), and in the generalized Conductor-like screening model implemented in NWChem (GCOSMO/NWChem). <sup>b</sup> References 18 and 26. <sup>c</sup> Reference 34. <sup>d</sup> References 30–32 and 75. <sup>e</sup> Reference 79.

also calculated (a total of 286 calculations). A locally modified version<sup>145</sup> of the Gaussian 03<sup>142</sup> electronic structure package was used to carry out the above calculations. The above calculations also gave the computed COT functions for each molecule in the training set as well as the SASAs for each atom in each molecule in the training set (the COT functions and SASAs are independent of the solvent).

Optimizing the parameters for nonaqueous solvents,  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ ,  $\tilde{\sigma}_i^{[\beta]}$ ,  $\tilde{\sigma}_i^{[\gamma]}$ ,  $\tilde{\sigma}_i^{[\phi^2]}$ ,  $\tilde{\sigma}_i^{[\psi^2]}$ , and  $\tilde{\sigma}_i^{[\beta^2]}$ , and the parameters for water,  $\tilde{\sigma}_i^{[\text{water}]}$ , involves minimizing the following error function

$$\chi = \sum_{j=1}^4 \sum_{J=1}^{2489} |\Delta G_S^0(\text{expt}, J) - \Delta G_{\text{EP}}(j, J) - G_{\text{CDS}}(j, J)| \quad (18)$$

where the first summation is over four levels of electronic structure theory (in particular, mPW1PW<sup>19</sup> with four different basis sets: MIDI!6D,<sup>50,51</sup> 6-31G(d),<sup>20</sup> 6-31+G(d),<sup>20</sup> and 6-31+G(d,p)<sup>20</sup>), and the second summation is over all data points in the neutral training set (2346 solvation free energies plus 143 transfer free energies), and  $\Delta G_S^0(\text{expt}, J)$  is the experimental solvation or transfer free energy. For solvation free energies,  $\Delta G_{\text{EP}}(j, J)$  and  $G_{\text{CDS}}(j, J)$  can be calculated directly with the solvation model, whereas for transfer free energies, two separate solvation model calculations are required, that is,

$$\Delta G_{\text{EP, transfer}} = \Delta G_{\text{EP, organic}} - \Delta G_{\text{EP, water}} \quad (19)$$

$$G_{\text{CDS, transfer}} = G_{\text{CDS, organic}} - G_{\text{CDS, water}} \quad (20)$$

where  $\Delta G_{\text{EP, organic}}$  and  $\Delta G_{\text{EP, water}}$  are calculated in the same way, except that different values are used for the dielectric constant in eq 2, and  $G_{\text{CDS, organic}}$  and  $G_{\text{CDS, water}}$  are computed using eqs 5 and 10, respectively. Note that because transfer free energies depend on both the aqueous solvation free energy and the solvation free energy in the organic solvent (eq 13), the parameters for nonaqueous solvents and the parameters for water must be optimized simultaneously.

The optimization of the above parameters was carried out in three stages. First, the  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ ,  $\tilde{\sigma}_i^{[\beta]}$ , and  $\tilde{\sigma}_i^{[\text{water}]}$  parameters for atoms involving at most H, C, N, and O and the

$\tilde{\sigma}_i^{[\gamma]}$ ,  $\tilde{\sigma}_i^{[\phi^2]}$ ,  $\tilde{\sigma}_i^{[\psi^2]}$ , and  $\tilde{\sigma}_i^{[\beta^2]}$  parameters were optimized against data for molecules containing H, C, N, and/or O. Next, these parameters were frozen, then the  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ ,  $\tilde{\sigma}_i^{[\beta]}$ , and  $\tilde{\sigma}_i^{[\text{water}]}$  parameters for atoms involving the elements F, S, Cl, and Br were optimized against data for molecules containing H, C, N, and/or O, plus F, S, Cl, and/or Br. Finally, these parameters were frozen, and then the  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ ,  $\tilde{\sigma}_i^{[\beta]}$ , and  $\tilde{\sigma}_i^{[\text{water}]}$  parameters for atoms involving P or Si were optimized against molecules containing P or Si.

SM8 uses the same functional forms for the atomic surface tensions as SM6, which contains 25 different  $\tilde{\sigma}_i^{[\text{water}]}$  values. Thus, SM8 contains these 25 parameters for water, plus 75  $\tilde{\sigma}_i^{[n]}$ ,  $\tilde{\sigma}_i^{[\alpha]}$ ,  $\tilde{\sigma}_i^{[\beta]}$  parameters, the 4  $\tilde{\sigma}_i^{[\gamma]}$ ,  $\tilde{\sigma}_i^{[\phi^2]}$ ,  $\tilde{\sigma}_i^{[\psi^2]}$ , and  $\tilde{\sigma}_i^{[\beta^2]}$  parameters, and the 4  $\tilde{\sigma}_{\text{Si}}^{[\text{water}]}$ ,  $\tilde{\sigma}_{\text{Si}}^{[n]}$ ,  $\tilde{\sigma}_{\text{Si}}^{[\alpha]}$ , and  $\tilde{\sigma}_{\text{Si}}^{[\beta]}$  parameters for silicon that were introduced as part of this work (108 parameters in all). However, as demonstrated by the performance of previous universal SMx models, it is not necessary (or desirable) to use all of these parameters. In an earlier paper,<sup>27</sup> a set of rules was adopted for determining which parameters to include, which are as follows: (1) If a parameter affects less than two different solutes, it is set to zero. (2) If using a parameter does not improve the mean unsigned error for the affected solutes by at least 0.1 kcal/mol, the parameter is set to zero. (3) Any surface tension coefficient that is not set to zero by either of these rules is retained.

For SM8 we instead used a different approach. In order to determine which parameters to retain in the SM8 parametrization, we used an approach based on statistical significance. First, all 108 parameters were optimized using the three-step procedure outlined above. Any parameter with a value greater than 1000 cal mol<sup>-1</sup> Å<sup>-2</sup> was removed, then the remaining parameters were reoptimized. In addition to being very large in value, all of the parameters that were removed in this stage of the optimization also had low values for the statistical significance. Next, the parameter with the least amount of statistical significance was removed, and then all of the remaining parameters were reoptimized. This step was repeated, until the statistical significance associated with each parameter was greater or equal to 95%. The only

**Table 6.** Surface Tension Model Parameters for SM8<sup>a</sup>

<i>i</i>	$\tilde{\sigma}_i^{[\text{water}]}$	$\tilde{\sigma}_i^{[\eta]}$	$\tilde{\sigma}_i^{[\alpha]}$	$\tilde{\sigma}_i^{[\beta]}$
H	58.93	22.02		
C	91.53	59.79	19.30	75.66
H, C	−81.35	−66.35		
C, C	−70.57	−63.62		−54.83
O	−97.68	−20.89	71.43	−142.42
H, O	−123.51	−78.77		
O, C	164.72	−11.64	127.68	134.04
O, O	86.92		122.12	−58.17
N	47.91	57.33	−120.41	
H, N	−118.50	−50.59		
C, N		−89.40	297.26	−105.62
N, C	−52.45	−3.88	−48.74	
N, C(2)	−194.27		−453.99	
N, C(3)	69.80			161.49
O, N	190.74			243.64
F	32.13			
Cl		−25.74		
Br	−19.92	−37.89		
S	−38.08	−49.29		
O, P	163.44		271.82	
Si		−72.87		
$\tilde{\sigma}^{[\nu]} = 0.21^b$ $\tilde{\sigma}^{[\phi^2]} = -2.79^b$ $\tilde{\sigma}^{[\psi^2]} = -8.46^b$ $\tilde{\sigma}^{[\psi^2]} = 2.51^b$				

<sup>a</sup> The units of  $\tilde{\sigma}$  are cal mol<sup>−1</sup> Å<sup>−2</sup>. The 54 CDS coefficients were optimized by fitting theoretical electrostatic solvation energies to the corresponding reference data that contained 274 aqueous free energies, 2072 nonaqueous free energies, and 143 transfer free energies. <sup>b</sup> These quantities are multiplied by the total solvent accessible surface area of a solute molecule.

exceptions were made for the two parameters  $\tilde{\sigma}_{\text{Si}}^{[\eta]}$  and  $\tilde{\sigma}_{\text{Br}}^{[\text{water}]}$ , which were retained despite having statistical significances of 93% and 91%, respectively. In all, only 54 of the original 108 parameters were retained, compared to the 75 parameters that are used by SM5.43. The final set of parameters obtained using the procedure described above is listed in Table 6. Note that in this table, only 21 types of surface tension parameters are listed, even though there are 26 possible types. This is because for  $\tilde{\sigma}_{\text{C,C(2)}}$ ,  $\tilde{\sigma}_{\text{H,S}}$ ,  $\tilde{\sigma}_{\text{S,S}}$ ,  $\tilde{\sigma}_{\text{P}}$ , and  $\tilde{\sigma}_{\text{S,P}}$  the final  $\tilde{\sigma}_i^{[\text{water}]}$ ,  $\tilde{\sigma}_i^{[\eta]}$ ,  $\sigma_i^{[\alpha]}$ , and  $\sigma_i^{[\beta]}$  values that result from following the above procedure are all equal to zero.

Thus we vary only 64 independent parameters, consisting of the 10 parameters *a* in eq 17 (see also Table 5) and 54 CDS parameters in eqs 5 and 10 (Table 6). These parameters are fit to 2730 reference free energies, which consist of 2346 solvation free energies for 318 neutral solutes in 90 nonaqueous solvents and water, 143 transfer free energies for 93 neutral solutes between water and 15 organic solvents, and 241 ionic free energies for 220 bare ions and 21 clustered ions in acetonitrile, dimethyl sulfoxide, and methanol. Thus there are more than 42 data per parameter, and this large ratio contributes to the robustness of the SM8 model. We refer the reader to results of the cross-validation procedure performed for the earlier SM5.43 model (that uses more parameters than SM8 and thus is formally less robust than SM8) by random removal of 25% of the data from the SM5.43 training set.<sup>34</sup> In these tests, the solvation free energies of solutes not used to train the SM5.43 model were predicted with only slightly increased mean unsigned errors

(0.47 vs 0.42 kcal/mol for aqueous and 0.52 vs 0.50 kcal/mol for organic neutral data).

## 6. Performance for Neutrals

Table 7 gives a breakdown of the errors in calculated aqueous solvation free energies by solute class. In Tables 8 and 9, the errors are broken down by solute class for calculated solvation free energies in nonaqueous solvents and for calculated transfer free energies, respectively. These tables show the results not only for the four levels of electronic structure theory used in the parametrization but also for two other density functional levels, namely B3LYP<sup>146–148</sup>/6-31G-(d) and M06-2X<sup>149</sup>/6-31G(d). Tables 7–9 show generally good agreement with experiment across both solute classes and electronic structure levels. The mean errors for electronic structure levels not included in the parametrization are not systematically worse (and in fact are often better) than those for the four levels used in parametrization.

Table 10 provides an overall summary of the performance of SM8 for neutral data and a comparison to SM7 and SM5.43 (which is called SM5.43R in the tables because in the SM5.43 model, the convention had been to append “R” if a gas-phase geometry was used; we will continue to say SM5.43 in the text). The performance of SM8 is quite comparable to the performance of SM5.43 and SM7 for neutral solvation.

Table 10 also shows (in parentheses) the mean unsigned errors we obtain if the sum over *j* in eq 18 is restricted to a single term, yielding a new set of atomic surface tensions for each level of electronic structure theory. The results are typically better by only 0.01–0.04 kcal/mol. We are willing to accept the slightly larger errors obtained with the catholic parameters, and so we are not publishing the individually optimized parameters.

The SM8 errors in solvation and transfer free energies for a few selected solvents, for which we have the most abundant solute data, are listed in Table 11. The errors over all solvents are given in the Supporting Information. The errors are small, lying within the uncertainty of experimental data. We found no statistically significant correlation between the errors and solvent values of hydrogen bond acidity or basicity parameters or other solvent properties.

## 7. Performance for Ions in Acetonitrile, DMSO, Methanol, and Water

Using the solvation free energies listed in Tables 2 and S1–S3 in part II of the Supporting Information as well as the earlier SM6 selectively clustered aqueous ion set, the performance of SM8 was tested for predicting solvation free energies of ions. Note that the data in Table 3 were used in parametrization but are not included in the error analyses of this section. See Tables S4 and S5 in part II of the Supporting Information for the error analyses including the clusters.

Table 12 shows the mean unsigned errors in SM8 solvation free energies of ions in the four solvents with three density functionals and four basis sets. As was mentioned above for neutral solutes, the overall performance of the SM8 model for ions only slightly depends on the change of electronic



**Table 7.** Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using SM8, by Solute Class<sup>a</sup>

solute class	N	mPW1PW				B3LYP	M06-2X
		MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	6-31G(d)	6-31G(d)
H <sub>2</sub> ,NH <sub>3</sub> ,H <sub>2</sub> O,(H <sub>2</sub> O) <sub>2</sub>	4	1.07	1.59	1.81	1.63	1.43	1.37
unbranched alkanes	8	0.88	0.88	1.05	1.01	0.82	0.83
branched alkanes	5	0.82	0.79	0.91	0.90	0.73	0.74
cycloalkanes	5	0.74	0.74	0.81	0.78	0.63	0.66
alkenes	9	0.43	0.45	0.59	0.53	0.35	0.36
alkynes	5	0.44	0.36	0.29	0.39	0.59	0.53
arenes	8	0.27	0.31	0.80	0.60	0.37	0.24
alcohols	12	0.63	0.41	0.40	0.46	0.50	0.59
phenols	4	0.94	0.53	0.47	0.58	0.98	0.79
ethers	12	0.45	0.46	0.57	0.61	0.46	0.62
aldehydes	6	0.74	0.46	0.39	0.45	0.36	0.38
ketones	12	0.50	0.36	0.56	0.57	0.28	0.52
carboxylic acids	5	0.45	0.44	1.07	1.04	0.59	0.86
esters	13	0.60	0.43	0.41	0.40	0.26	0.15
peroxides	3	0.43	0.14	0.26	0.22	0.13	0.12
bifunctional H,C,O compounds	5	0.76	0.45	0.40	0.47	0.62	0.94
aliphatic amines	15	0.75	0.61	0.57	0.59	0.60	0.60
anilines	7	0.61	0.41	0.26	0.53	0.92	0.79
aromatic N-heterocycles (1 N)	10	0.15	0.20	0.54	0.47	0.53	0.66
aromatic N-heterocycles (2 Ns)	2	1.29	0.43	0.45	0.47	0.35	0.34
nitriles	4	0.65	0.34	1.07	1.05	0.39	1.03
hydrazines	3	1.22	0.99	0.97	0.89	0.86	0.85
bifunctional H,C,N compounds	3	0.46	0.55	0.39	0.83	0.69	0.63
amides	4	1.00	0.71	0.99	1.05	0.85	1.12
ureas	2	0.52	0.41	1.04	0.52	0.29	0.41
thymines (uracils)	1	1.18	1.76	0.23	0.77	1.61	0.68
nitrohydrocarbons	7	0.74	0.30	0.42	0.45	0.32	0.41
bifunctional H,C,N,O compounds	3	0.72	0.30	0.16	0.22	0.04	0.23
fluoroalkanes	5	1.00	0.58	0.32	0.31	0.55	0.29
fluoroarenes	1	0.11	0.28	0.80	0.71	0.00	0.08
chloroalkanes	13	0.30	0.31	0.46	0.53	0.28	0.27
chloroalkenes	6	0.57	0.59	0.44	0.52	0.66	0.70
chloroarenes	8	0.29	0.41	0.68	0.85	0.22	0.20
bromoalkanes	9	0.18	0.17	0.35	0.35	0.16	0.15
bromoalkenes	1	0.14	0.06	0.09	0.06	0.17	0.23
bromoarenes	4	0.27	0.41	0.65	0.53	0.16	0.22
multihalogen hydrocarbons	12	0.48	0.29	0.31	0.33	0.28	0.27
halogenated bifunctional compounds	9	1.39	1.12	1.77	1.77	1.18	1.48
thiols	4	0.71	0.60	0.50	0.43	0.74	0.72
sulfides	5	0.79	0.80	0.59	0.51	0.92	0.94
disulfides	2	0.16	0.09	0.49	0.65	0.20	0.22
sulfur heterocycles	1	0.24	0.24	0.27	0.11	0.50	0.41
halogenated sulfur compounds	2	1.26	1.70	0.84	1.08	1.48	1.71
phosphorus compounds	14	1.21	1.48	1.55	1.58	1.50	1.62
silicon compounds	1	0.33	0.23	0.27	0.27	0.31	0.17
all neutral data	274	0.63	0.55	0.66	0.67	0.57	0.62

<sup>a</sup> All the solvation free energies were obtained using the SM8 model parameters. *N* is the number of data in a given solute class.

structure level. Again, the errors for B3LYP/6-31G(d) and M06-2X/6-31G(d), which were not included in the SM8 parametrization, are not systematically larger than for those levels of theory that were used in parametrization.

Experimental solvation free energies  $\Delta G_s^o$  for neutral solutes in the SM8 training set vary from  $-14.1$  kcal/mol (for chrysene in hexadecane) to  $4.3$  kcal/mol (for octafluoropropane in water) with the average value (averaged over all 2346 data) equal to  $-4.8$  kcal/mol. Experimental solvation free energies  $\Delta G_s^o$  for ions in acetonitrile, DMSO, metha-

nol, and water vary from  $-110.3$  kcal/mol (for aqueous H<sub>3</sub>O<sup>+</sup>) to  $-36.0$  kcal/mol (for 2,4-dinitrophenoxide anion in acetonitrile) with the average value (averaged over all 332 data) equal to  $-65.0$  kcal/mol. The SM8 model predicts these average values of  $\Delta G_s^o$  quite precisely: indeed,  $-4.9$  kcal/mol for neutrals and  $-66.0$  kcal/mol for ions. The average SM8 values of  $\Delta G_{ENP}$  are  $-2.1$  kcal/mol for neutral solutes and  $-64.3$  kcal/mol for ions. (Recall that  $\Delta G_{ENP}$  is approximated as  $\Delta G_{EP}$  in the present article. All SM8 results given in this paragraph are calculated at the mPW1PW/6-

**Table 8.** Mean Unsigned Errors (kcal/mol) in Nonaqueous Solvation Free Energies Calculated Using SM8, by Solute Class<sup>a</sup>

solute class	N	mPW1PW				B3LYP	M06-2X
		MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	6-31G(d)	6-31G(d)
H <sub>2</sub> , NH <sub>3</sub> , H <sub>2</sub> O, (H <sub>2</sub> O) <sub>2</sub>	29	1.60	1.81	2.00	2.03	1.73	1.72
unbranched alkanes	85	0.45	0.45	0.41	0.42	0.45	0.45
branched alkanes	7	0.39	0.41	0.39	0.39	0.41	0.41
cycloalkanes	13	0.47	0.47	0.49	0.48	0.42	0.43
alkenes	18	0.41	0.40	0.36	0.37	0.42	0.42
alkynes	9	0.52	0.47	0.39	0.46	0.57	0.55
arenes	134	0.44	0.50	0.83	0.71	0.35	0.38
alcohols	272	0.38	0.39	0.39	0.39	0.38	0.38
phenols	109	0.75	0.59	0.52	0.59	0.81	0.72
ethers	87	0.68	0.67	0.67	0.68	0.69	0.71
aldehydes	32	0.69	0.68	0.76	0.77	0.63	0.60
ketones	195	0.86	0.77	0.75	0.80	0.70	0.51
carboxylic acids	120	0.51	0.54	0.75	0.73	0.58	0.68
esters, including lactones <sup>b</sup>	243	0.44	0.42	0.48	0.50	0.44	0.47
peroxides	17	0.58	0.60	0.60	0.60	0.58	0.59
bifunctional H,C,O compounds	24	1.37	1.22	1.07	1.10	1.30	1.42
aliphatic amines	154	0.43	0.41	0.40	0.40	0.42	0.43
anilines	61	0.38	0.36	0.37	0.36	0.48	0.43
aromatic N-heterocycles (1 N)	52	0.62	0.61	0.59	0.60	0.64	0.65
aromatic N-heterocycles (2 Ns)	8	0.46	0.58	0.84	1.15	0.81	0.94
nitriles	20	0.70	0.58	0.75	0.75	0.54	0.51
hydrazines	5	1.30	1.26	1.26	1.24	1.26	1.27
bifunctional H,C,N compounds	2	0.79	1.02	0.77	0.81	0.80	0.80
amides	26	0.69	0.60	0.65	0.67	0.70	0.83
ureas	7	1.14	0.85	0.59	0.93	1.02	1.10
lactams	4	0.69	0.79	0.84	0.89	0.88	0.95
thymines (uracils)	1	0.67	0.97	0.19	0.22	0.78	0.42
nitrohydrocarbons	86	0.77	0.56	0.43	0.44	0.56	0.53
bifunctional H,C,N,O compounds	3	0.72	0.67	0.83	0.97	0.77	0.80
fluoroalkanes	5	0.86	0.72	0.54	0.54	0.70	0.61
fluoroarenes	11	0.54	0.60	0.76	0.73	0.55	0.57
chloroalkanes	26	0.44	0.51	0.59	0.63	0.48	0.43
chloroalkenes	15	0.69	0.72	0.58	0.62	0.75	0.77
chloroarenes	31	0.31	0.33	0.42	0.39	0.34	0.32
bromoalkanes	21	0.55	0.54	0.61	0.61	0.51	0.48
bromoalkenes	2	0.09	0.08	0.18	0.16	0.10	0.11
bromoarenes	16	0.41	0.49	0.63	0.58	0.35	0.39
multihalogen hydrocarbons	14	0.46	0.39	0.42	0.42	0.39	0.35
halogenated bifunctional compounds	37	1.14	0.99	1.13	1.18	1.14	1.11
thiols	10	0.30	0.23	0.18	0.17	0.35	0.31
sulfides	13	0.88	0.91	0.94	0.92	0.87	0.89
disulfides	4	0.38	0.41	0.37	0.36	0.47	0.47
sulfurheterocycles	4	0.63	0.63	0.29	0.39	0.80	0.74
sulfoxides	1	0.33	0.92	0.92	0.96	0.76	0.40
phosphorus compounds	37	1.43	1.63	1.61	1.63	1.67	1.73
silicon compounds	2	1.96	1.53	1.92	1.92	1.46	1.57
all neutral data	2072	0.60	0.57	0.61	0.62	0.58	0.57

<sup>a</sup> All the solvation free energies were obtained using the SM8 model parameters. *N* is the number of data in a given solute class. <sup>b</sup> Five lactones and 238 other esters.

31G(d)/CM4 level.) The  $\Delta G_{EP}$  results indicate that the average signed value of non-bulk-electrostatic contributions in the free energy of solvation  $\Delta G_S^0$  is approximately the same for both neutral and ionic solutes ( $-2.8$  kcal/mol for neutrals and  $-1.7$  kcal/mol for ions; the smaller absolute value for ions is understandable in that the typical ion in our data set is smaller than the typical neutral solute). The average absolute value of  $G_{CDS}$ , i.e.  $\langle |G_{CDS}| \rangle$ , is 3.0 kcal/

mol for neutrals and 2.3 kcal/mol for ions, whereas  $\langle |\Delta G_S^0| \rangle$  is 5.0 kcal/mol for neutrals and 66.0 kcal/mol for ions, and  $\langle |\Delta G_{EP}| \rangle$  is the same as  $|\langle \Delta G_{EP} \rangle|$  since  $\Delta G_{EP}$  is intrinsically negative. In the generalized Born approximation,  $\Delta G_{EP}$  need not be negative (as it should be), but there are only 7 positive values (out of 2346 neutral solvation free energies in our SM8 data set), and the largest—hexadecane in hexadecane—is only +0.15 kcal/mol.

**Table 9.** Mean Unsigned Errors (kcal/mol) in Transfer Free Energies between Water and Organic Solvents Calculated Using SM8, by Solute Class<sup>a</sup>

solute class	N	mPW1PW				B3LYP	M06-2X
		MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	6-31G(d)	6-31G(d)
lactones	10	1.27	1.03	0.96	0.96	1.05	0.87
aromatic N-heterocycles	6	0.43	0.44	0.39	0.34	0.37	0.33
bifunctional H,C,N compounds	2	0.79	0.91	0.74	0.74	0.89	0.81
amides	13	0.96	0.79	0.98	0.77	0.88	1.14
ureas	11	0.30	0.35	0.32	0.27	0.34	0.27
lactams	4	1.66	1.60	1.72	1.73	1.62	1.74
thymines and uracils	12	0.78	0.98	0.70	0.73	1.03	0.60
bifunctional H,C,N,O compounds	5	0.45	0.60	0.47	0.46	0.64	0.53
halogenated bifunctional compounds	39	0.77	0.83	0.69	0.66	0.84	0.65
sulfur compounds (with no P)	19	0.42	0.47	0.54	0.53	0.47	0.49
phosphorus compounds	9	0.69	1.15	0.55	0.55	1.21	1.44
silicon compounds	13	0.82	0.81	0.76	0.78	0.80	0.81
all neutral data	143	0.74	0.78	0.70	0.66	0.80	0.74

<sup>a</sup> All the solvation free energies were obtained using the SM8 model parameters. *N* is the number of data in a given solute class.

**Table 10.** Mean Unsigned Errors (kcal/mol) in Solvation Free Energies of Neutral Solutes Using CM4

model <sup>a</sup>	DFT method	basis	MUE		
			aqueous <sup>b</sup>	organic <sup>c</sup>	transfer <sup>d</sup>
SM8-CM4	mPW1PW	MIDI!6D	0.63 (0.58)	0.60 (0.60)	0.74 (0.70)
SM8-CM4	mPW1PW	6-31G(d)	0.55 (0.55)	0.57 (0.57)	0.78 (0.68)
SM8-CM4	mPW1PW	6-31+G(d)	0.66 (0.63)	0.61 (0.60)	0.70 (0.69)
SM8-CM4	mPW1PW	6-31+G(d,p)	0.67 (0.63)	0.62 (0.61)	0.66 (0.68)
SM8-CM4	B3LYP	6-31G(d)	0.57 (0.55)	0.58 (0.57)	0.80 (0.70)
SM8-CM4	M06-2X	6-31G(d)	0.62 (0.55)	0.57 (0.54)	0.74 (0.68)
SM7 <sup>e</sup>	mPW1PW	6-31G(d)	0.53	0.61	0.70
SM5.43R <sup>f</sup>	mPW1PW	6-31G(d)	0.55	0.61	1.02

<sup>a</sup> The CDS contributions in the SM8 free energies of solvation were found using the 54 parameters ( $\tilde{\sigma}_i$ ) of the SM8 model presented in Table 6 with the exception of the numbers in the parentheses that were obtained by optimization of the 54 parameters  $\tilde{\sigma}_i$  for each specified basis set and density functional. <sup>b</sup> Two hundred seventy-four data. <sup>c</sup> Two thousand seventy-two data. <sup>d</sup> One hundred forty-three data. <sup>e</sup> The SM7 model<sup>26</sup> uses 56 parameters  $\tilde{\sigma}_i$  and the CM4 charge model. <sup>f</sup> The SM5.43R model<sup>34</sup> uses 75 parameters  $\tilde{\sigma}_i$  and the CM3 charge model.

## 8. Performance of Other Continuum Models

In addition to SM8, SM7, and SM5.43, we tested the performance of several other implicit solvent models, in particular five models that serve as default solvation models in five popular quantum-chemical program packages: (1) the Integral Equation Formalism Polarizable Continuum Model<sup>150,151</sup> of *Gaussian 03*,<sup>142</sup> namely IEF-PCM/G03;<sup>152–155</sup> (2) the dielectric version<sup>150,151,156</sup> of PCM (D-PCM/G98) as implemented in *Gaussian 98*;<sup>157</sup> (3) the Conductor-like PCM model<sup>150,151,158–164</sup> in *GAMESS* (C-PCM/GAMESS);<sup>165–167</sup> (4) *Jaguar*'s Poisson–Boltzmann (PB) self-consistent reaction field solver (PB/*Jaguar*);<sup>168–170</sup> and (5) the Generalized Conductor-like Screening Model (GCOSMO) as implemented in *NWChem* (GCOSMO/*NWChem*).<sup>171</sup>

For nonelectrostatic contributions, we accept the defaults of these program packages. Thus the *Gaussian* PCM calculations include not only electrostatics but also cavitation, dispersion, and repulsion, as explained in the original references.<sup>154,156</sup> In contrast, the default in *GAMESS*<sup>165–167</sup> and *NWChem*<sup>171</sup> is to only include electrostatics. In *Jaguar*,<sup>170</sup> the default involves only electrostatics for the nonaqueous

**Table 11.** Errors (kcal/mol) in Solvation and Transfer Free Energies for Neutrals Calculated at the SM8/mPW1PW/6-31G(d) Level of Theory, by Solvent<sup>a</sup>

solvent	N	$\alpha$	$\beta$	MSE <sup>b</sup>	MUE <sup>c</sup>
benzene	75	0.00	0.14	0.21	0.65
carbon tetrachloride	78	0.00	0.00	−0.45	0.62
chlorobenzene	38	0.00	0.07	−0.36	0.54
chloroform	105	0.15	0.02	0.11	0.77
cyclohexane	91	0.00	0.00	0.05	0.49
decane	39	0.00	0.00	−0.18	0.41
dichloroethane	38	0.10	0.11	0.30	0.59
diethyl ether	67	0.00	0.41	0.05	0.71
heptane	66	0.00	0.00	−0.02	0.36
hexadecane	190	0.00	0.00	−0.03	0.50
hexane	59	0.00	0.00	−0.08	0.49
isooctane	32	0.00	0.00	−0.37	0.48
octane	38	0.00	0.00	−0.09	0.39
octanol	206	0.37	0.48	−0.10	0.66
octanol–water transfer	90	0.37	0.48	0.04	0.65
toluene	51	0.00	0.14	0.08	0.45
water	274	0.82	0.35	−0.06	0.55
xylene	48	0.00	0.16	0.11	0.45

<sup>a</sup> *N* is the number of neutral solute data in a given solvent. Only the solvents with *N* > 30 are listed here. See the Supporting Information for all data.  $\alpha$  is Abraham's<sup>80–83</sup> hydrogen bond acidity parameter (which Abraham denotes as  $\Sigma\alpha_2$ ), and  $\beta$  is Abraham's hydrogen bond basicity parameter (which Abraham denotes as  $\Sigma\beta_2$ ).

<sup>b</sup> Mean signed error. <sup>c</sup> Mean unsigned error.

solvents but both electrostatics and nonelectrostatic terms<sup>169</sup> for the aqueous model.

There are various ways to implement the Conductor-like Screening Model (COSMO) algorithm,<sup>33,160,161,172–175</sup> and the various later implementations should not be confused with the original COSMO method of Klamt and Schüürmann<sup>160</sup> or with the Conductor-like Screening Model for Real Solvents (COSMO-RS)<sup>176,177</sup> that provides a current enhanced version of the COSMO method.<sup>160</sup> Analysis of the performance of the original COSMO method<sup>160</sup> or COSMO-RS<sup>176</sup> or the performance of GCOSMO with the radii optimized by Stefanovich and Truong<sup>174</sup> is beyond the scope of the present study. By GCOSMO/*NWChem* we refer to the default implementation of the COSMO method in the



**Table 12.** Mean Unsigned Errors (kcal/mol) in Ionic Solvation Free Energies Calculated Using SM8<sup>a</sup>

		mPW1PW				B3LYP	M06-2X
solute class	N	MIDI!6D	6-31G(d)	6-31+G(d)	6-31+G(d,p)	6-31G(d)	6-31G(d)
Acetonitrile							
H,C,N,O cations <sup>b</sup>	36	6.3	6.4	6.7	6.6	6.5	6.5
S cations <sup>c</sup>	3	15.9	16.1	16.6	16.6	16.2	16.1
all cations	39	7.0	7.2	7.4	7.4	7.2	7.2
H,C,N,O anions <sup>b</sup>	19	3.5	4.3	5.5	5.5	3.9	4.3
F,Cl,Br,S anions <sup>c</sup>	11	2.6	3.2	4.3	4.4	3.0	3.3
all anions	30	3.1	3.9	5.1	5.1	3.6	3.9
all ions	69	5.3	5.8	6.4	6.4	5.6	5.8
DMSO							
H,C,N,O cations <sup>b</sup>	4	1.8	1.7	2.0	2.2	1.7	1.8
all cations	4	1.8	1.7	2.0	2.2	1.7	1.8
H,C,N,O anions <sup>b</sup>	52	7.8	8.3	8.9	8.8	7.9	8.0
F,Cl,Br,S anions <sup>c</sup>	15	3.2	3.9	4.6	4.7	3.7	4.0
all anions	67	6.8	7.3	7.9	7.9	6.9	7.1
all ions	71	6.5	7.0	7.6	7.6	6.6	6.8
Methanol							
H,C,N,O cations <sup>b</sup>	26	2.0	2.0	2.3	2.3	2.1	2.1
Cl,Br cations <sup>c</sup>	3	0.3	0.7	1.3	1.7	0.4	0.7
all cations	29	1.8	1.9	2.2	2.3	1.9	2.0
H,C,N,O anions <sup>b</sup>	36	2.1	2.3	3.4	3.5	2.2	2.3
F,Cl,Br anions <sup>c</sup>	15	2.0	2.3	4.3	4.3	2.1	2.4
all anions	51	2.1	2.3	3.6	3.7	2.2	2.4
all ions	80	2.0	2.2	3.1	3.2	2.1	2.2
Water <sup>d</sup>							
H,C,N,O cations <sup>b</sup>	48	2.8	2.7	3.4	3.4	2.7	2.7
Cl,S cations <sup>c</sup>	4	2.3	2.3	2.7	2.9	2.1	2.4
all cations	52	2.7	2.7	3.3	3.4	2.7	2.7
H,C,N,O anions <sup>b</sup>	43	4.9	4.0	3.5	3.4	4.3	4.3
F,Cl,Br,S anions <sup>c</sup>	17	3.4	3.0	2.5	2.5	3.1	3.1
all anions	60	4.5	3.7	3.2	3.2	4.0	3.9
all ions	112	3.7	3.2	3.3	3.3	3.4	3.4

<sup>a</sup> The solvation free energies obtained using the SM8 model parameters. <sup>b</sup> Ions containing no elements other than H, C, N, or O. <sup>c</sup> Ions containing any of the listed elements in addition to H, C, N, or O. <sup>d</sup> One hundred twelve selectively clustered ions from the SM6 model training set as defined in ref 18.

NWChem-version 4.7 computer package.<sup>171</sup> This implementation uses the atomic radii in Table 5 with the generalized COSMO<sup>161,172,173</sup> (GCOSMO) dielectric screening factor for the conductor-like surface charge. Note that the radii used by default in NWChem, as given in Table 5, differ from the values given by the NWChem manual. Note also that the default radius of the silicon atom (1.17 Å) is equal to the covalent Si radius given in Table 7-13 in ref 178 that is much smaller than Bondi's van der Waals atomic radius for Si,  $R(\text{Si}) = 2.10 \text{ Å}$ .<sup>79</sup> Nevertheless we used the default radii given in Table 5, and we also accepted all other program defaults.

The radii used for IEF-PCM/G03, D-PCM/G98, C-PCM/GAMESS, and PB/Jaguar electrostatic calculations also require further discussion. The PCM methods in Gaussian were tested both with atomic radii and group radii; in the latter case one treats certain groups consisting of an atom and its covalently attached hydrogens as a pseudoatom (called a united atom) in forming the cavity. We used three different schemes for assigning atomic or group radii in the IEF-PCM/G03 calculation. First we used the United-Atom Hartree-Fock (UAHF) scheme<sup>179</sup> that is the recommended

method for predicting solvation free energies with PCM according to the Gaussian 03 manual.<sup>142</sup> We also tested IEF-PCM with the UA0 and Bondi schemes. (We note that although UAHF is the recommended scheme in the Gaussian 03 manual for using with the Hartree-Fock method or DFT, UA0 is the default scheme.) With the UA0 scheme (also called the "united atoms topological model") in Gaussian 03, one sometimes needs to use the "sphereonh=N" option to place an individual sphere on a hydrogen that Gaussian recognizes as having more than one bond. In particular this is required for the anion of 2-hydroxybenzoic acid and for all of the solute-water clusters used in the set of selectively clustered ions. The D-PCM/G98 model was tested with UAHF group radii. The values of intrinsic atomic radii used for cavity construction with C-PCM/GAMESS and PB/Jaguar are listed in Table 5. The Jaguar program uses the atomic radii of Table 5 only for calculation of nonaqueous solvation free energies, whereas for calculation of aqueous solvation energies it employs atomic radii that depend on typing certain functional groups in a solute molecule.<sup>169</sup> The boundary between the solute and solvent used by PB/Jaguar is the so-called molecular surface,<sup>168</sup> which depends on the

**Table 13.** Errors (kcal/mol) in Ionic Solvation Free Energies Calculated Using Various Solvent Models<sup>a</sup>

Table 10. Errors (kcal/mol) in the Solvation Free Energies Calculated Using Various Solvent Models												
solute class	N	SM8	SM7	SM5.43R	IEF-PCM/G03				D-PCM/G98 UAHF <sup>+</sup>	C-PCM** GAMESS	PB** Jaguar	GCOSMO** NWChem
					UA0	UAHF	Bondi	UAHF <sup>+</sup>				
Acetonitrile												
MSE (cations)	39	5.1	6.6	4.2	18.7	24.2	12.7	23.8	23.4	14.6	7.3	−3.2
MSE (anions)	30	−3.9	−13.7	−10.1	2.2	1.0	−9.1	−1.0	−1.4	−6.9	−12.4	−22.7
MUE (cations)	39	7.2	6.6	5.6	18.7	24.2	12.7	23.8	23.4	14.6	7.3	4.6
MUE (anions)	30	3.9	13.7	10.1	3.4	2.7	9.1	2.0	2.4	6.9	12.4	22.7
DMSO												
MSE (cations)	4	−1.3	5.0	0.4	15.7	23.7	6.9	23.5	23.5	15.0	5.0	0.4
MSE (anions)	67	−7.0	−14.3	−10.1	−3.5	−1.5	−12.2	−3.1	−2.4	−6.4	−13.2 <sup>b</sup>	−22.1
MUE (cations)	4	1.7	5.0	2.1	15.7	23.7	6.9	23.5	23.5	15.0	5.0	1.4
MUE (anions)	67	7.3	14.3	10.1	5.0	4.9	12.2	5.7	4.5	6.6	13.2 <sup>b</sup>	22.1
Methanol												
MSE (cations)	29	−1.0	−1.2	−4.0	7.5	5.0	0.3	4.3	4.5	8.0	0.1	−10.3
MSE (anions)	51	−1.5	−1.4	3.0	7.3	1.7	−0.4	−1.6	−4.3	6.2	0.5	−9.7
MUE (cations)	29	1.9	1.9	4.5	7.5	5.1	2.0	4.5	4.6	8.0	2.1	10.3
MUE (anions)	51	2.3	2.2	3.6	7.8	3.0	3.3	2.5	4.3	6.4	3.2	10.0
Water <sup>c</sup>												
MSE (cations)	52	1.0	1.0	−0.2	10.9	5.8	2.8	5.3	5.8	7.7	2.4	−10.8
MSE (anions)	60	1.8	2.0	6.4	13.7	6.2	3.8	4.1	4.5	8.9	3.0	−6.9 <sup>d</sup>
MUE (cations)	52	2.7	2.8	5.2	10.9	6.2	3.7	5.7	6.1	7.7	3.1	11.0
MUE (anions)	60	3.7	3.8	6.7	13.7	10.7	5.5	8.9	5.4	8.9	4.8	7.0 <sup>d</sup>

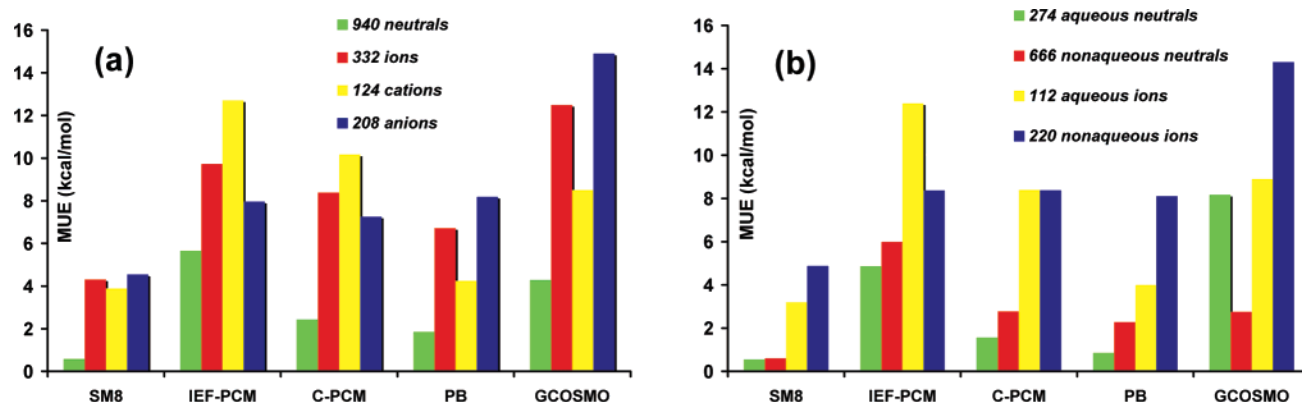
<sup>a</sup> N is the number of data in a given solute class. MSE/MUE refers to mean signed/unsigned error. The SMx models are described in the text. IEF-PCM/G03 was used with the following methods for assigning atomic or group radii: the united-atom universal force field topological model (UA0), the united-atom Hartree–Fock method (UAHF), and the Bondi atomic radii (Bondi). D-PCM/G98 is the dielectric version of PCM implemented in *Gaussian 98* with using the UAHF radii. The calculations were performed at the mPW1PW/6-31G(d) level of theory, except for the calculations marked by the asterisks: they used the Hartree–Fock method (\*) and B3LYP (\*\*). <sup>b</sup> No data were obtained for 3-hydroxybenzoic acid (anion). The total count is reduced to  $N - 1$ . <sup>c</sup> One hundred twelve selectively clustered ions from the SM6 model training set as defined in ref 18. <sup>d</sup> No data were obtained for hydroperoxyl radical (anion). The total count is reduced to  $N - 1$ .

solvent probe radius, which was set, following the instructions in the manual,<sup>170</sup> to 1.4 Å for water and calculated from the solvent density by assuming a packing fraction of 0.5 for other solvents. The values listed in Table 5 for C-PCM/GAMESS are the radii used in the C-PCM/GAMESS electrostatic calculations. Note that the user would have to input values a factor of 1.2 smaller than those in the table since in the cavity construction algorithm, *GAMESS* multiplies the input values by 1.2.

To test all of the continuum models we used the same data on 332 ions in acetonitrile, DMSO, methanol, and water as described above. However, the test of neutrals was performed only for those 17 solvents (including acetonitrile, DMSO, and water) that are available for IEF-PCM/G03 in *Gaussian 03* (actually *Gaussian 03* supports 21 solvents, but we have neutral data in only 17 of them; see Table 1). Note that, for example, methanol is available in *Gaussian 03*, but we have no neutral data for this solvent. All the calculations for these comparisons were carried out with the 6-31G(d) basis. The mPW1PW density functional<sup>19</sup> was used with the SM8, SM7, SM5.43, and IEF-PCM/G03 models. However mPW1PW was not available with C-PCM/GAMESS, PB/Jaguar, and GCOSMO/NWChem. In these three cases we employed B3LYP<sup>146–148</sup> instead of mPW1PW. The IEF-PCM/G03 and D-PCM/G98 models with the UAHF scheme for assigning atomic radii were also tested using the Hartree–Fock method because the parameters contained in the UAHF model were originally optimized for the HF/6-31G(d) level of theory.<sup>179</sup>

Before turning to the results, we comment on the standard states used by the various program packages. All programs tested use a gas-phase standard state of 1 mol/L, and all results presented in the paper use this standard state.

Table 13 shows the mean signed and unsigned errors between calculated and experimental solvation free energies of anions and cations for each of the implicit solvent models mentioned above, in acetonitrile, DMSO, methanol, and water (see the Supporting Information for more details). For ions in methanol and water, SM8 and SM7 give nearly identical average errors, and both are more accurate than the other models tested. PB/Jaguar also gives quite accurate (but still inferior to SM8) predictions for ions in methanol and water. For acetonitrile and DMSO, the performance of each of these models is highly dependent on whether the solute is an anion or cation. For SM7 and SM5.43, the calculated solvation free energies in acetonitrile and DMSO are significantly more accurate for cations than for anions. The opposite occurs when IEF-PCM/G03 or D-PCM/G98 (that yields similar results to those from IEF-PCM/G03) is used with either UA0 or UAHF radii; any of these models is able to predict solvation free energies of anions in acetonitrile and DMSO fairly accurately, whereas for cations, any of these PCM models gives mean unsigned errors of over 15 kcal/mol! One can observe that in many cases the Bondi scheme in conjunction with IEF-PCM/G03 can provide more accurate predictions than the united atom models. The GCOSMO/NWChem model gives errors for cations in acetonitrile and DMSO that are smaller than



**Figure 2.** Mean unsigned errors (MUEs) in solvation free energies of neutral and ionic solutes calculated using SM8 and other continuum models including IEF-PCM/G03 with the UA0 model for assigning atomic or group radii, C-PCM/GAMESS, PB/Jaguar, and GCOSMO/NWChem. B3LYP was used with *GAMESS*, *Jaguar*, and *NWChem*, because mPW1PW was unavailable. The calculation was done only for 18 solvents, which are available for IEF-PCM/G03, including acetonitrile (ions and neutral solutes), DMSO (ions and neutral solutes), methanol (only ions), water (ions and neutral solutes), and an additional 14 organic solvents from the SM8 neutral training set (Table 1; see also footnotes in Tables 13 and 14). (a) MUEs are given for ions and neutrals in all of the 18 solvents. (b) MUEs for solutes in aqueous solutions are compared to MUEs for solutes in nonaqueous solutions.

**Table 14.** Errors (kcal/mol) in Solvation Free Energies Calculated Using Various Solvent Models<sup>a</sup>

method	aqueous neutrals <sup>b</sup>		organic neutrals <sup>c</sup>		ions <sup>d</sup>	
	MSE	MUE	MSE	MUE	MSE	MUE
SM8	-0.06	0.55	-0.02	0.61	-1.02	4.31
SM7	-0.07	0.53	-0.11	0.59	-3.09	6.59
SM5.43R	0.00	0.55	0.10	0.67	-1.18	6.60
IEF-PCM/UA0	4.86	4.87	5.94	5.99	7.45	9.73
IEF-PCM/UAHF	0.61	1.18	3.88	3.94	5.63	8.15
C-PCM/GAMESS	-0.65	1.57	2.62	2.78	4.45	8.39
PB/Jaguar	0.22	0.86	1.69	2.28	-1.86 <sup>e</sup>	6.72 <sup>e</sup>
GCOSMO	-8.17 <sup>f</sup>	8.17 <sup>f</sup>	-2.12	2.76	-12.21 <sup>g</sup>	12.49 <sup>g</sup>
3PM	0.00	2.65	0.00	1.49	0.00	8.60

<sup>a</sup> MSE/MUE refers to mean signed/unsigned error. IEF-PCM/G03 was used with the united-atom universal force field topological model (UA0) and the united-atom Hartree–Fock model (UAHF) for assigning group radii. The SMx and IEF-PCM calculations were performed at the mPW1PW/6-31G(d) level; the other calculations were performed at the B3LYP/6-31G(d) level. 3PM refers to the three-parameter model described in the text. <sup>b</sup> Two hundred seventy-four neutral data, unless indicated otherwise. <sup>c</sup> Six hundred sixty-six neutral data in 16 non-aqueous solvents available with IEF-PCM/G03. <sup>d</sup> Three hundred thirty-two data (unless indicated otherwise) for ions in acetonitrile, DMSO, methanol, and water. <sup>e</sup> No data were obtained for 3-hydroxybenzoic acid (anion) in DMSO. The total count is reduced to 331. <sup>f</sup> No data were obtained for 11 phosphorus-containing compounds and tetramethylsilane. The total count is reduced to 262. <sup>g</sup> No data were obtained for hydroperoxyl radical (anion) in water. The total count is reduced to 331.

those obtained by SM8, but the errors for anions exceed 20 kcal/mol. Similarly, SM7 and SM5.43 tend to overestimate the solvation free energies of anions in acetonitrile and DMSO. Thus SM8 has a much better average performance in acetonitrile than any other model in Table 13.

The errors in predicting the solvation free energies of neutral solutes by different models are listed in Table 14. We limited our calculations on neutral solutes only to testing the following models at the DFT/6-31G(d) level of theory: IEF-PCM/G03/UA0, IEF-PCM/G03/UAHF, C-PCM/

**Table 15.** Mean Unsigned Errors (kcal/mol) in Solvation Free Energies Calculated with the mPW1PW Density Functional and with Class II and Class IV Partial Atomic Charges<sup>a</sup>

model	basis	neutral data			
		aqueous <sup>b</sup>	organic <sup>c</sup>	transfer <sup>d</sup>	ions <sup>e</sup>
SM8-LPA	MIDI!6D	1.26	0.83	0.81	4.21
SM8-LPA	6-31G(d)	1.93	1.40	0.78	5.60
SM8-RLPA	6-31+G(d)	2.18	1.51	0.93	6.33
SM8-RLPA	6-31+G(d,p)	1.39	0.89	0.90	5.54
SM8-CM4	MIDI!6D	0.63	0.60	0.74	4.20
SM8-CM4	6-31G(d)	0.55	0.57	0.78	4.31
SM8-CM4	6-31+G(d)	0.66	0.61	0.70	4.79
SM8-CM4	6-31+G(d,p)	0.67	0.62	0.66	4.81

<sup>a</sup> LPA denotes Löwdin population analysis, and RLPA denotes redistributed Löwdin population analysis; population analysis yields class II charges as defined in ref 41. CM4 denotes charge model 4, which yields class IV charges, also defined in ref 41. <sup>b</sup> Two hundred seventy-four data. <sup>c</sup> Two thousand seventy-two data. <sup>d</sup> One hundred forty-three data. <sup>e</sup> Three hundred thirty-two data.

GAMESS, PB/Jaguar, and GCOSMO/NWChem. Again, the SMx models provide much more accurate predictions of experimental free energies of solvation than any of these models. In particular, GCOSMO/NWChem gives an unacceptably large overestimate of aqueous neutral data (up to 8 kcal/mol on average). The most accurate non-SMx model tested is PB/Jaguar. However even in this case the error for the neutral solutes in organic solvents is 4.5 times larger than obtained with the SM8 model. Figure 2 complements the analysis of various continuum solvation models presented in Tables 13 and 14 and shows again that the newly developed SM8 solvation model significantly outperforms the most popular implicit solvent models that are widely used in modeling condensed media.

We close this section by evaluating one more solvation model, which we call the three-parameter model (3PM). The 3PM predicts that all neutral solvation free energies in aqueous solution are -2.99 kcal/mol, all neutral solvation



**Table 16.** Mean Unsigned Errors (kcal/mol) in Solvation Free Energies Calculated Using SMx and Non-SMx Implicit Solvent Models<sup>a</sup>

solute class	N	SM8	SM7	SM5.43R	IEF-PCM/G03				D-PCM/G98 UAHF <sup>+</sup>	C-PCM <sup>**</sup> GAMESS	PB <sup>**</sup> Jaguar	GCOSMO <sup>**</sup> NWChem
					UA0	UAHF	Bondi	UAHF <sup>+</sup>				
all neutrals	940	0.59	0.57	0.64	5.66	3.14				2.43	1.86	4.29 <sup>b</sup>
all ions	332	4.31	6.59	6.60	9.73	8.15	7.08	7.67	7.15	8.39	6.72 <sup>c</sup>	12.49 <sup>d</sup>
all cations	124	3.90	10.19	5.06	12.71	12.17	6.24	11.69	11.75	10.18	4.25	8.51
all anions	208	4.55	8.22	7.53	7.97	5.79	7.64	5.30	4.41	7.26	8.19 <sup>c</sup>	14.90 <sup>d</sup>
aqueous neutrals	274	0.55	0.53	0.55	4.87	1.18				1.57	0.86	8.17 <sup>b</sup>
nonaqueous neutrals	666	0.61	0.59	0.67	5.99	3.94				2.78	2.28	2.76
aqueous ions	112	3.24	3.31	6.00	12.43	8.61	4.64	7.43	5.73	8.36	4.03	8.85 <sup>d</sup>
nonaqueous ions	220	4.88	8.26	6.90	8.37	7.93	8.34	7.81	7.89	8.38	8.11 <sup>c</sup>	14.31

<sup>a</sup> N is the number of data in a given solute class. The SMx models are described in the text. IEF-PCM/G03 was used with the following methods for assigning atomic or group radii: the united-atom universal force field topological model (UA0), the united-atom Hartree–Fock model (UAHF), and the Bondi atomic radii (Bondi). D-PCM/G98 is the dielectric version of PCM implemented in *Gaussian 98* with using the UAHF radii. The calculations were performed at the mPW1PW/6-31G(d) level of theory, except for the calculations marked by the asterisks: they used the Hartree–Fock method (\*) and B3LYP (\*\*). <sup>b</sup> No data were obtained for 11 phosphorus-containing compounds and tetramethylsilane. The total count is reduced to  $N - 12$ . <sup>c</sup> No data were obtained for 3-hydroxybenzoic acid (anion). The total count is reduced to  $N - 1$ . <sup>d</sup> No data were obtained for hydroperoxyl radical (anion). The total count is reduced to  $N - 1$ .

free energies in organic solvents are  $-5.38$  kcal/mol, and all ionic solvation free energies are  $-65.0$  kcal/mol; these are the average experimental values averaged over 274, 666, 332 data, respectively (Table 14). The mean unsigned error of the 3PM is 2.7 kcal/mol for neutrals in water, 1.5 kcal/mol for neutrals in nonaqueous solvents, and 8.6 kcal/mol for ions. Comparison to the MUE column in Table 14 shows, somewhat disappointingly, that the non-SMx models outperform the 3PM in only 6 out of 15 possible cases.

## 9. Using Other Charge Models

Although in this work the performance of SM8 has only been illustrated for six electronic structure levels, experience with SM6<sup>18</sup> and the MPWX series<sup>34,39</sup> (where X denotes a fraction of Hartree–Fock exchange) shows that the SMx models can be used with any density functional or with the Hartree–Fock approximation as long as one uses class IV charges. SM8 can also be used with other kinds of charges. One can expect the most reliable results if the user validates that the charge model chosen gives partial atomic charges that are reasonably similar to CM4 charges. Table 15 and Tables S8–S11 in part II of the Supporting Information show examples of using SM8 with other charge models; the results are less accurate than with CM4 charges, but even with the less accurate class II charges based on Löwdin or redistributed population analysis the average errors of SM8 are smaller than the errors of many non-SMx model listed in Table 14.

## 10. Summary and Concluding Remarks

Using experimental  $pK_a$  values in acetonitrile, DMSO, and methanol, experimental gas-phase acidities, accurate values for the absolute solvation free energy of the proton, and solvation free energies of neutral solutes that were computed using SM7, a data set of single-ion solvation free energies in the three solvents above was assembled. Using these data and data assembled previously for solvation free energies of ions in water, solvation free energies of neutrals in water and 90 nonaqueous solvents, and transfer free energies of neutrals from water to 15 nonaqueous solvents, a new universal implicit solvent model called SM8 has been

developed for predicting solvation free energies of neutral and ionic solutes in water and in nonaqueous solvents. For nonaqueous solvents, SM8 uses a small set of solvent descriptors that characterize the properties of the solvent.

Like several previous universal SMx models, SM8 gives solvation free energies of neutral solutes that are typically within  $\sim 0.6$  kcal/mol of the experimental value, despite using fewer parameters than the earlier models. For ionic solvation the present models provide considerable improvement over all previous methods. Since new ionic data are used to obtain a physical partitioning of the solvation energy into bulk electrostatic and non-bulk-electrostatic components, and self-consistently polarized charge distributions are used to calculate the bulk electrostatic contributions, we expect that not only the solvation free energies but also the charge distributions and properties of the dissolved molecules should be well represented. Thus the present model can be used with confidence to calculate partition coefficients (e.g., Henry's Law constants, octanol/water partition coefficients, etc),<sup>54</sup> solubilities,<sup>180</sup> vapor pressures,<sup>34,181</sup> liquid-phase geometries of neutral and charged species (including transition state species),<sup>55</sup> and, when combined with gas-phase acid dissociation free energies, liquid-phase  $pK_a$  constants.<sup>182</sup> These properties can be calculated in any solvent that can be characterized by the solvent descriptors used by the SM8 model.

The key descriptors used by SM8 such as dielectric constant, Abraham's hydrogen bond acidity and basicity parameters, refractive index, and macroscopic surface tension at an air/solvent interface are tabulated in the literature for almost all possible organic solvents, and that is the primary sense in which our model is universal. However the applicability of SM8 as presented here is still limited to room temperature. (An extension to variable temperature for aqueous solutions is essentially complete,<sup>183</sup> and it will be submitted soon.). Although the SM8 training set includes solvation free energies of solutes only in pure aqueous and organic solvents but not in mixtures of solvents, the SM8 model can also be applied to complex "solvents" such as

membranes,<sup>184</sup> interfaces,<sup>185</sup> or mixtures, provided effective values for the solvent descriptors are available or can be obtained.

Table 16 summarizes the comparison of the present model to the previously published universal solvation model from our group (SM5.43), to an unpublished model based on using atomic radii optimized for water in all solvents (SM7), and to several solvation models from popular computer packages. For neutral solutes in aqueous solution, the mean unsigned error of SM8 is 0.55 kcal/mol, whereas the errors in the five non-SMx models we tested are 0.9–8.2 kcal/mol. For solvation of neutrals in nonaqueous solvents the mean unsigned error of SM8 increases to 0.61 kcal/mol, whereas the errors in the five non-SMx model we tested are 2.3–6.0 kcal/mol. For ions, SM8 gives a mean unsigned error of 4.3 kcal/mol, whereas the errors in the eight non-SMx models we tested are 6.7–12.5 kcal/mol.

**Acknowledgment.** This work was supported by the National Institutes of Health training grant for Neurophysical-computational Sciences, by the U.S. Army Research Office under the Multidisciplinary Research Program of the University Research Initiative through grant no. DAAD19-02-1-0176, by the National Science Foundation (grant CHE06-10183 and grant CHE07-04974), and by the Office of Naval Research under grant no. 00014-05-01-0538.

**Supporting Information Available:** Two thousand three hundred forty-six reference solvation free energies and 143 reference transfer energies for neutral solutes in the SM8 training set; reference free energies for 112 selectively clustered ions in water; and 220 unclustered ions and 21 ionic clusters in acetonitrile, DMSO, and methanol (part I) and reference  $pK_a$  constants, gas-phase acidity and basicity values, and solvation free energies of neutral species used for evaluation of the reference solvation free energies of the corresponding ions in acetonitrile, DMSO, and methanol; SM7 and SM8 calculated free energies of nonaqueous ions; errors in solvation and transfer free energies calculated by SM8 using the class IV CM4 charges and the class II charges based on Löwdin or redistributed Löwdin population analyses; errors in solvation free energies of neutral and ionic solutes calculated using SMx and non-SMx implicit solvent models, by solvent and by solute class; and the Cartesian coordinates corresponding to the B97-1/MG3S optimized global minima for nonaqueous clusters (part II). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.
- (2) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
- (3) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305.
- (4) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
- (5) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
- (6) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
- (7) Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-López, M. F.; Rinaldi, D.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284.
- (8) Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
- (9) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.
- (10) Giesen, D. J.; Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. In *Computational Thermochemistry*; Irikura, K., Frurip, D. J., Eds.; ACS Symposium Series 677; American Chemical Society: Washington, DC, 1998; p 285.
- (11) Hawkins, G. D.; Zhu, T.; Li, J.; Chambers, C. C.; Giesen, D. J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998; p 201.
- (12) Cramer, C. J.; Truhlar, D. G. In *Trends and Perspectives in Modern Computational Science*; Maroulis, G., Simos, T. E., Eds.; Lecture Series on Computer and Computational Sciences 6; Brill/VSP: Leiden, 2006; p 112.
- (13) Klotz, I. M.; Rosenberg, R. M. *Chemical Thermodynamics: Basic Theory and Methods*, 5th ed.; Wiley: New York, 1994; p 459.
- (14) Lewis, G. N.; Randall, M.; Pitzer, K. S.; Brewer, L. *Thermodynamics*, 2nd ed.; McGraw-Hill: New York, 1961; p 399.
- (15) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, *102*, 7787.
- (16) Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795.
- (17) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066.
- (18) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.
- (19) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (20) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- (21) Fuoss, R. M.; Accascina, F. *Electrolytic Conductance*; Interscience: New York, 1959.
- (22) Szwarc, M. *Acc. Chem. Res.* **1969**, *2*, 87.
- (23) Mayer, U. *Coord. Chem. Rev.* **1976**, *21*, 159.
- (24) Krell, M.; Symons, M. C. R.; Barthel, J. *J. Chem. Soc., Faraday Trans. 1* **1987**, *83*, 3419.
- (25) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408.
- (26) (a) Kelly, C. P. Ph.D. Thesis, University of Minnesota: Minneapolis, 2007. (b) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Unpublished.
- (27) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.
- (28) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.
- (29) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.

- (30) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117; errata: **1999**, *111*, 5624 and **2000**, *113*, 3930.
- (31) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293.
- (32) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
- (33) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340.
- (34) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532.
- (35) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 2061.
- (36) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85.
- (37) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305.
- (38) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178.
- (39) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 107.
- (40) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- (41) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87.
- (42) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.
- (43) Winget, P.; Thompson, J. D.; Xidos, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 10707.
- (44) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.
- (45) Golebiewski, A.; Rzeszowska, E. *Acta Phys. Pol., A* **1974**, *45*, 563.
- (46) Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221.
- (47) Kar, T.; Sannigrahi, A. B.; Mukherjee, D. C. *J. Mol. Struct.: THEOCHEM* **1987**, *153*, 93.
- (48) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545.
- (49) Frauenheim, T.; Seifert, G.; Elstner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. *Phys. Status Solidi B* **2000**, *217*, 41.
- (50) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281.
- (51) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192.
- (52) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.*, **2007**, *6*, 2046–2054.
- (53) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987.
- (54) 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; p 63.
- (55) Zhu, T.; Li, J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 5503.
- (56) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; Reidel: Dordrecht, 1980; p 25ff.
- (57) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; p 1.
- (58) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas Belg.* **1956**, *75*, 487.
- (59) Peradejordi, F. *Cahiers Phys.* **1963**, *17*, 393.
- (60) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.
- (61) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 6, p 1.
- (62) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. *PhysChemComm* **2002**, *5*, 117.
- (63) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.
- (64) Mayer, I. *Chem. Phys. Lett.* **1985**, *117*, 396.
- (65) Mayer, I. *Int. J. Quantum Chem.* **1986**, *29*, 73.
- (66) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291.
- (67) Brom, J. M.; Schmitz, B. J.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 6483.
- (68) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 133.
- (69) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (70) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 724.
- (71) Jorgensen, W. L.; Ulmschneider, J. P.; Tirado-Rives, J. *J. Phys. Chem. B* **2004**, *108*, 16264.
- (72) Martin, F.; Zipse, H. *J. Comput. Chem.* **2004**, *26*, 97.
- (73) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
- (74) Winget, P.; Cramer, C. J.; Truhlar, D. G. *Environ. Sci. Technol.* **2000**, *34*, 4733.
- (75) Winget, P.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 5160.
- (76) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422.
- (77) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
- (78) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
- (79) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- (80) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1989**, 699.
- (81) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.
- (82) Abraham, M. H. *J. Phys. Org. Chem.* **1993**, *6*, 660.
- (83) Abraham, M. H. In *Quantitative Treatment of Solute/Solvent Interactions*; Theoretical and Computational Chemistry Series Vol. 1; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; p 83.
- (84) *Physical/Chemical Property Database (PHYSPROP)*; SRC Environmental Science Center: Syracuse, NY, 1994.



- (85) Leo, A. J. *Masterfile from MedChem Software*; BioByte Corp.: Claremont, CA, 1994.
- (86) Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. G.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database version 3.0*; University of Minnesota: Minneapolis, MN 55455-0431, 2007.
- (87) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, 27, 413.
- (88) 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.
- (89) Chantooni, M. K., Jr.; Kolthoff, I. M. *J. Am. Chem. Soc.* **1968**, 90, 3005.
- (90) Coetsee, J. F.; Padmanabhan, G. R. *J. Am. Chem. Soc.* **1965**, 87, 5005.
- (91) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1973**, 95, 8539.
- (92) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1973**, 95, 4768.
- (93) Kolthoff, I. M.; Chantooni, M. K., Jr.; Bhowmik, S. *Anal. Chem.* **1967**, 39, 1627.
- (94) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1968**, 90, 3320.
- (95) Beltrame, P.; Gelli, G.; Loi, A. *Gazz. Chim. Ital.* **1980**, 110, 491.
- (96) Chantooni, M. K., Jr.; Kolthoff, I. M. *J. Phys. Chem.* **1974**, 78, 839.
- (97) Coetsee, J. F.; Padmanabhan, G. R. *J. Phys. Chem.* **1965**, 69, 3193.
- (98) Jasinski, T.; El-Harakany, A. A.; Halaka, F. G.; Sadek, H. *Croat. Chem. Acta* **1978**, 51, 1.
- (99) Kolthoff, I. M.; Chantooni, M. K., Jr.; Bhowmik, S. *J. Am. Chem. Soc.* **1966**, 88, 5430.
- (100) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Phys. Chem.* **1966**, 70, 856.
- (101) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1970**, 92, 7025.
- (102) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1971**, 93, 3843.
- (103) Chantooni, M. K., Jr.; Kolthoff, I. M. *J. Phys. Chem.* **1976**, 80, 1306.
- (104) Chantooni, M. K., Jr.; Kolthoff, I. M. *J. Phys. Chem.* **1975**, 79, 1176.
- (105) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1969**, 91, 4621.
- (106) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1975**, 97, 1376.
- (107) Kolthoff, I. M.; Chantooni, M. K., Jr.; Bhowmik, S. *J. Am. Chem. Soc.* **1968**, 90, 23.
- (108) Ludwig, M.; Pytela, O.; Vecera, M. *Collect. Czech. Chem. Commun.* **1984**, 49, 2593.
- (109) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1976**, 98, 5063.
- (110) Chantooni, M. K., Jr.; Kolthoff, I. M. *Anal. Chem.* **1979**, 51, 133.
- (111) Kolthoff, I. M.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1965**, 87, 4428.
- (112) Kolthoff, I. M.; Bruckenstein, S.; Chantooni, M. K., Jr. *J. Am. Chem. Soc.* **1961**, 83, 3927.
- (113) Kolthoff, I. M.; Chantooni, M. K., Jr.; Bhowmik, S. *Anal. Chem.* **1967**, 39, 315.
- (114) Bordwell, F. G. *Acc. Chem. Res.* **1988**, 21, 456.
- (115) Jasinski, T.; Stefaniuk, K. *Chem. Anal. (Warsaw)* **1965**, 10, 211.
- (116) Ritchie, C. D.; Uschold, R. E. *J. Am. Chem. Soc.* **1968**, 90, 2821.
- (117) Matthews, W. S.; Bares, J. E.; Bartmess, J. E.; Bordwell, F. G.; Cornforth, F. J.; Drucker, G. E.; Margolin, Z.; McCallum, R. J.; McCollum, G. J.; Vanier, N. R. *J. Am. Chem. Soc.* **1975**, 97, 7006.
- (118) Bordwell, F. G.; Algrim, D. J. *J. Am. Chem. Soc.* **1988**, 110, 2964.
- (119) Clare, B. W.; Cook, D.; Ko, E. C. F.; Mac, Y. C.; Parker, A. J. *J. Am. Chem. Soc.* **1966**, 88, 1911.
- (120) Courtot-Coupez, J.; Le Demez, M. *Bull. Soc. Chim. Fr.* **1969**, 1033.
- (121) Kolthoff, I. M.; Reddy, T. B. *Inorg. Chem.* **1962**, 1, 189.
- (122) Olmstead, W. N.; Margolin, Z.; Bordwell, F. G. *J. Org. Chem.* **1980**, 45, 3295.
- (123) Ritchie, C. D.; Uschold, R. E. *J. Am. Chem. Soc.* **1967**, 89, 1721.
- (124) Bordwell, F. G.; McCallum, R. J.; Olmstead, W. N. *J. Org. Chem.* **1984**, 49, 1424.
- (125) Rived, F.; Rosés, M.; Bosch, E. *Anal. Chim. Acta* **1998**, 374, 309.
- (126) Rochester, C. H. *J. Chem. Soc. B* **1967**, 33.
- (127) Ritchie, C. D.; Heffley, P. D. *J. Am. Chem. Soc.* **1965**, 87, 5402.
- (128) Juillard, J. *Bull. Soc. Chim. Fr.* **1966**, 1727.
- (129) Juillard, J.; Dondon, M.-L. *Bull. Soc. Chim. Fr.* **1963**, 2535.
- (130) Kononov, O. M. *Zh. Fiz. Khim.* **1965**, 39, 693.
- (131) Bolton, P. D.; Rochester, C. H.; Rossall, B. *Trans. Faraday Soc.* **1970**, 66, 1348.
- (132) Juillard, J. Ph.D. Thesis, University of Clermont-Ferrand: Clermont-Ferrand, France, 1967.
- (133) Shedlovsky, T.; Kay, R. L. *J. Phys. Chem.* **1956**, 60, 151.
- (134) Kolthoff, I. M.; Lingane, J. J.; Larson, W. D. *J. Am. Chem. Soc.* **1938**, 60, 2512.
- (135) Leung, C. S.; Grunwald, E. *J. Phys. Chem.* **1970**, 74, 696.
- (136) Charlot, G.; Tremillon, B. *Chemical Reactions in Solvents and Melts*, 1st Engl. ed.; Pergamon Press: New York, 1969; p 278.
- (137) Izmailov, N. A.; Chernyi, V. S.; Spivak, L. L. *Zh. Fiz. Khim.* **1963**, 37, 822.
- (138) Mason, R. B.; Kilpatrick, M. *J. Am. Chem. Soc.* **1937**, 59, 572.
- (139) Rochester, C. H. *Trans. Faraday Soc.* **1966**, 62, 355.

- (140) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. *J. Chem. Phys.* **1998**, *109*, 6264.
- (141) Fast, P. L.; Sánchez, M. L.; Truhlar, D. G. *Chem. Phys. Lett.* **1999**, *306*, 407.
- (142) 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.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, 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.; Bakken, V.; 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, Revisions C.01, C.02, and D.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (143) 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.
- (144) Thompson, J. D.; Winget, P.; Truhlar, D. G. *PhysChemComm* **2001**, *16*, 1.
- (145) Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM, version 2007-beta*; University of Minnesota: Minneapolis, MN, 55455-0431, 2007.
- (146) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (147) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (148) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (149) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* In press.
- (150) Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
- (151) Miertuš, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
- (152) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.
- (153) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.
- (154) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253.
- (155) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43.
- (156) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.
- (157) 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.; 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*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (158) Li, H.; Pomelli, C. S.; Jensen, J. H. *Theor. Chem. Acc.* **2003**, *109*, 71.
- (159) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449.
- (160) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
- (161) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253.
- (162) Baldrige, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622.
- (163) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.
- (164) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.
- (165) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.
- (166) Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; p 1167.
- (167) *GAMESS computer package, version 7 SEP 2006 (R6)*; Iowa State University: Ames, IA, 2006. <http://www.msg.ameslab.gov/GAMESS/GAMESS.html> (accessed Feb 2007).
- (168) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875.
- (169) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.
- (170) *Jaguar 6.5, Release 112*; Schrödinger, Inc.: Portland, OR, 2005.
- (171) Bylaska, E. J.; de Jong, W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Aprà, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; Zhao, Y.; Fan, P.-D.; Harrison, R. J.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Nooijen, M.; Brown, E.; Cisneros, G.; Fann, G. I.; Früchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, R. A.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyal, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Pollack, L.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7*; Pacific Northwest National Laboratory: Richland, WA, 2006.

- (172) Truong, T. N.; Stefanovich, E. V. *J. Phys. Chem.* **1995**, *99*, 14700.
- (173) Truong, T. N.; Stefanovich, E. V. *J. Chem. Phys.* **1995**, *103*, 3709.
- (174) Stefanovich, E. V.; Truong, T. N. *Chem. Phys. Lett.* **1995**, *244*, 65.
- (175) Truong, T. N.; Nguyen, U. N.; Stefanovich, E. V. *Int. J. Quantum Chem.* **1996**, *60*, 1615.
- (176) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224.
- (177) Klamt, A.; Eckert, F. *Fluid Phase Equilib.* **2000**, *172*, 43.
- (178) Pauling, L. *The Nature of The Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (179) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
- (180) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661.
- (181) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2000**, *104*, 4726.
- (182) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493.
- (183) Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. To be published.
- (184) Chambers, C. C.; Giesen, D. J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G.; Vaes, W. H. J. In *Rational Drug Design*; Truhlar, D. G., Howe, W. J., Hopfinger, A. J., Blaney, J. M., Dammkoehler, R. A., Eds.; Springer: New York, 1999; p 51.
- (185) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2004**, *108*, 12882.

CT7001418