6532 *J. Phys. Chem. A* **2004,** *108,* 6532−6542

New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and for Vapor Pressures

Jason D. Thompson, Christopher J. Cramer,* and Donald G. Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota,
Minneapolis, Minnesota 55455-0431

Received: January 27, 2004; In Final Form: April 28, 2004

We present a new continuum solvation model, called Solvation Model 5.43R (SM5.43R). The model is based on the generalized Born approximation for electrostatics augmented by terms that are proportional to the solvent-accessible surface areas (SASAs) of the atoms of the solute, and it is parametrized to predict the free energy of solvation of solutes containing H, C, N, O, F, P, S, Cl, and Br in water and organic solvents. The new model is an improvement over our previous solvation model, SM5.42R, in that it is based on CM3 charges rather than CM2 charges, it was trained over a larger and more diverse training set, and the choice of the value of the solvent radius, which is used to compute the SASA of the atoms of the solute, was made on a different basis than was used for SM5.42R. This paper presents parametrizations of SM5.43R using HF/6-31G(d), B3LYP/6-31G(d), mPW1PW91/6-31G(d), and mPW1PW91/6-31+G(d) to describe the electronic structure of the solute. For a data set of neutral solutes with known experimental aqueous free energies of solvation containing at most H, C, N, O, F, P, S, Cl, and Br (257 data), the mean-unsigned error (MUE, in kcal/mol), as compared to experiment, calculated by SM5.43R is respectively 0.50, 0.49, 0.50, and 0.54 using these four solute wave functions. A similar MUE is obtained for SM5.42R using HF/6-31G(d). The corresponding MUEs calculated by several other generally available continuum solvation models are approximately a factor of 2 larger than those computed by SM5.43R and SM5.42R using the same electronic structure methods. For a data set of solutes with experimental free energies of solvation in 16 organic solvents (621 data), SM5.43R and SM5.42R yield MUEs 6.3 to 7.9 times smaller than the MUEs calculated by the other continuum solvation models. The SM5.43R model is, however, universal in that it can be used in any organic solvent, as well as water. Furthermore, it allows one to analyze solvation trends in terms of local properties, and this is illustrated for acetanilide in water and diethyl ether.

Introduction

The standard-state free energy of solvation is the free energy difference associated with the transfer of a solute X from the gas-phase to a given solvent Y,[1] and it is a fundamental quantity that describes the interactions between a solute molecule and the solvent in which it is dissolved.[1−3] The free energy of solvation of a solvent molecule in a neat liquid provides the vapor pressure of a molecule, which is also known as the free energy of self-solvation,[4] and the free energy of solvation may also be used to calculate the vapor pressure of dilute solutions.[5] The free energy of solvation in two solvents provides enough information to calculate the partition coefficient of a solute between the two solvents (the free energy difference associated with partitioning is known as the free energy of transfer); not only is this a fundamental physical property in its own right, it also serves as the basis of a number of empirical correlations of biological response.[6,7] Thermodynamic relationships also can be used to relate the standard-state free energy of solvation to solubility.[8,9]

Accurate predictions of the standard-state free energy of solvation require accurate treatment of long-range mutual polarization effects and short-range interactions between the solute and the solvent,[2,3,10] where the latter include all deviations from bulk electrostatics. A number of continuum solvation models (where explicit solvent molecules are replaced with a continuous and homogeneous dielectric medium characterized by bulk-solvent descriptors) have been developed to model these effects.[2,3,10−17] We and our co-workers have developed a sequence of successively improved semiempirical continuum solvation models,[18−36] called SM*x* models in general or universal SM*x* models for those applicable to general organic solvents as well as aqueous solutions. These models account for bulk electrostatic interactions between the solute and the solvent quantum mechanically by a self-consistent-field molecular orbital calculation (Hartree-Fock or density functional theory) that employs the generalized Born (GB) approximation.[18,37−39] The GB method, which represents the solute as a space-filling cavity with atom-centered point charges surrounded by spheres of given radii, approximates the numerical solution to the Poisson equation for molecular-shaped solutes. First-solvation-shell effects (nonbulk electrostatic effects) due to interactions between the solute and solvent molecules in the near vicinity of the solute are modeled as proportional to the solvent-accessible surface area (SASA) of the atoms in the solute. Note that the SASA can be defined in several ways.[22,40,41] The constants of proportionality (called atomic surface tension functionals because they have units of energy per unit area) contain optimized parameters and depend on smooth and continuous functions of the geometry. These functions distin-

---

*To whom correspondence should be addressed

10.1021/jp0496295 CCC: $27.50

guish between the different types of chemical environments that a particular atom in the solute might encounter. In this way, our continuum solvation models can be used to calculate the free energies of solvation for arbitrary functional groups without requiring the user to assign types to atoms (as in molecular mechanics); the environmental dependence of the parameters is calculated automatically from the geometry.

In general, the interaction of the solute with the solvent depends on the geometry of the solute, and in principle, the computation of the free energy of solvation should involve an average over solute geometries. For most solutes, it is reasonable to replace this by a single calculation using either the optimized gas-phase geometry or the optimized liquid-phase geometry, and this is the procedure that is almost always employed in practice.[2,3]

In this article, we present a new universal continuum solvation model, called Solvation Model 5.43R (SM5.43R). This notation indicates the version of the geometry-dependent surface tension functionals (version 5)[24,26−30,33] used, that class IV point charges[42] are used in the GB calculation, in particular CM3 charges[43,44] (hence the "43" part of the method's name), and that gas-phase geometries are used in the parametrization (hence R, for rigid geometries); the method can also be applied using liquid-phase optimized geometries,[45,46] in which case the results would be labeled SM5.43. In the present article we use mPW1PW91[47,48]/MIDI![49,50] gas-phase geometries for all liquid-phase calculations. The SM5.43R model is an improvement over our most recent previous solvation model, SM5.42R, in that it is based on CM3 charges rather than CM2[51,52] charges. Both charge models use functions of the Mayer bond order[53−55] to map class II[42] charges, such as charges calculated from a Löwdin population analysis[56−59] of the wave function, to improved charges (as judged on the basis of accurate dipole moments). The CM3 charge model was trained on a larger training set of dipole moment data than CM2. It has also been designed to be less sensitive to variations in bond orders than CM2, especially when diffuse functions are included in the basis set, which can be particularly important for calculations employing hybrid density functional theory, where diffuse basis functions can be particularly effective for calculating relative energies.[60] For basis sets with diffuse functions, CM3 maps charges calculated from a redistributed Löwdin population analysis[61] (RLPA) of the wave function. These charges have been shown[61] to be less sensitive to the inclusion of diffuse functions in the basis set than charges calculated from a Löwdin[56−59] or a Mulliken[62] population analysis.

There are two other key differences between SM5.42R and SM5.43R. First, in SM5.42R, the solvent radius used in the calculation of the SASAs of the atoms in the solute was set to zero because that choice resulted in the smallest root-mean-square errors in predicted free energies of solvation of typical organic solutes. Because a SASA of an atom calculated with a nonzero solvent radius is more physically meaningful than a SASA calculated with no solvent radius, especially for calculating potentials of mean force (PMFs), we have chosen to use a nonzero solvent radius in SM5.43R, and the precise value of the solvent radius is based on PMF calculations. Second, SM5.43R is parametrized with a new training set. The training set is improved in two ways: (1) we use improved values for the free energy of solvation of ions, and (2) we use a larger training set of data for neutral solutes than was used to parametrize SM5.42R. We especially note that this training set includes free energies of transfer[35] for solutes for which absolute free energies of solvation are not available, and this allows us to include a greater diversity of solute functionality in the parametrization of SM5.43R as compared to SM5.42R.

Four new parametrizations of SM5.43R for solutes containing H, C, N, O, F, P, S, Cl, and Br will be presented, where each parametrization uses a particular kind of solute wave function. The four cases considered here are HF[63]/6-31G(d),[64−66] B3LYP[67−70]/6-31G(d), mPW1PW91[47,48]/6-31G(d), and mPW1PW91/6-31+G(d).[71] The B3LYP and mPW1PW91 methods are hybrid density functional theory (hybrid DFT) methods, and the mPW1PW91 electronic structure method is more accurate than B3LYP and HF for predicting energies of reaction and barrier heights.[72,73] Furthermore the 6-31+G(d) basis set contains diffuse functions, which can be particularly effective with hybrid DFT,[60] and therefore the mPW1PW91/6-31+G(d) method is of special interest for developing a solvation model for applications where one wants to use the same method for calculating relative energies in both the gas phase and the liquid phase.

We will evaluate SM5.43R by comparing its ability to predict free energies of solvation to SM5.42R and to several popular and generally available continuum solvation models, namely the conductor-like[74,75] version (C-PCM) of the polarizable continuum model[76] (PCM), the dielectric version[76−78] of PCM (D-PCM), and the integral equation formalism[79−81] version of PCM (IEF-PCM). We use both the formulation of C-PCM proposed by Barone and Cossi,[82] as implemented in Gaussian 98[83] and also the reformulation of C-PCM proposed by Cossi and co-workers,[16] as implemented in Gaussian 03.[84] These two methods will be denoted C-PCM/98 and C-PCM/03 when one uses the default settings of Gaussian 98 and Gaussian 03, respectively. For D-PCM and IEF-PCM, we use the formulation and implementation proposed by Cossi and co-workers,[85] as implemented in Gaussian 03.[84] They will be denoted as D-PCM/03 and IEF-PCM/03 throughout this paper. All PCM methods used here represent the solute as a cavity made up of a set of interlocking spheres. The cavity is built by the united-atom-for-Hartree−Fock (UAHF) method,[86] which has been optimized for HF/6-31G(d) and is the recommended method for predicting free energies of solvation according to the Gaussian 03 manual.[84]

In the sections that follow, a brief introduction to the SM5.43R model and the details of the parameter optimization for SM5.43R will be presented, the results of the four parameter optimizations will be given, and comparisons to the C-PCM, D-PCM, IEF-PCM, and SM5.42R models will be presented. Finally, some concluding remarks will be given.

Theory

In the SM5.43R continuum solvation model, the solvent is taken to be a continuous and homogeneous dielectric medium characterized by the solvent's bulk dielectric constant, ϵ. The solvent is further characterized by several solvent descriptors (described below), which are used to account for the contributions of the first solvation shell to the free energy of solvation. The standard-state free energy of solvation, $\Delta G_{\mathrm{S}}^{\mathrm{o}}$, of a solute in a liquid solvent is written in SM5.43R as

$$\Delta G_{\mathrm{S}}^{\mathrm{o}} = \Delta E + G_{\mathrm{P}} + G_{\mathrm{CDS}} + \Delta G_{\mathrm{conc}}^{\mathrm{o}} \qquad (1)$$

where G_{P} is the electronic polarization energy from mutual polarization of the solute and the solvent, ΔE is the change in the solute's internal electronic energy when the solute is placed in the solvent, G_{CDS} is a semiempirical term that accounts for all interactions except bulk electrostatics, and $\Delta G_{\mathrm{conc}}^{\mathrm{o}}$ accounts for the concentration change (if any) between the gas-phase and the liquid-phase standard states. Since we use the same concentrations (1 mol/L) in both phases, $\Delta G^{\text{o}}_{\text{conc}}$ is zero.[1] The sum of the first two terms in eq 1 is the net effect of the bulk-electrostatic interactions between the solute and the solvent, and they are calculated from a self-consistent molecular orbital calculation employing the GB method. The GB method takes as input the dielectric constant of the solvent, a set empirical Coulomb radii of the atoms in the solute, and the partial atomic charges of the solute (computed self-consistently from the CM3 charge model in this case). The Coulomb radii depend on atomic numbers Z and are called ρ_Z. Because the GB approximation represents the solute charge distribution as a set of partial atomic charges at the nuclei, it does not suffer from outlying charge error[14,87,88] (volume polarization[89,90]) due to solute charge penetration outside the electrostatic cavity, and this makes the method more stable than most other methods. The price one pays for this, in principle, is a less realistic description of the solute charge density; however, the use of class IV charges ameliorates this problem since they are designed to represent the accurate charge distribution in terms of atomic monopoles.

The third term on the right-hand side of eq 1, G_{CDS}, is a semiempirical term that accounts for short-range interactions (also known as first-shell solvation effects) between the solute and the solvent, that is, for the difference between the total free energy of solvation and that estimated from bulk electrostatics. This includes the difference from the bulk electrostatics of the interactions of the solute with solvent molecules in its near vicinity and the resultant changes in solvent−solvent interactions. The G_{CDS} term also accounts for the free energy cost associated with making a cavity in the solvent to accommodate the placement of the solute into the solvent, attractive dispersion interactions between the solute and the solvent and the change in solvent−solvent dispersion interactions when the solute is inserted, solvent-structural rearrangement of the solvent when the solute is placed in the solvent, deviations of the effective dielectric constant from the bulk one in the region near the solute, deviations of the true solute−solvent interface from the model one defined by the empirical Coulomb radii, short-range exchange repulsion forces between the solute and the solvent, neglect of charge transfer between the solute and the solvent, and systematic errors (when present) in the GB approximation and in the ability of the partial atomic charges to represent the true solute charge distribution. These effects are modeled as proportional to the solvent-accessible surface area[40,41] (SASA) of the atoms in the solute. The van der Waals radii used in the computation of the SASAs are the standard values of Bondi.[91]

For organic solvents, the G_{CDS} term in our universal solvation models depends on several solvent properties as well as ϵ; these properties are used to further characterize the nature of the solvent in the near vicinity of the solute.[26,29,30] The G_{CDS} terms depend on the following six solvent descriptors: the refractive index (at the wavelength of the Na D line), n, Abraham's[92−96] hydrogen bond acidity parameter, α (which Abraham denotes as $\sum \alpha_2$), Abraham's[92−96] hydrogen bond basicity parameter, β (which Abraham denotes as $\sum \beta_2$), the reduced surface tension, γ, which equals $\gamma_{\text{m}}/\gamma^{\text{o}}$ where γ_{m} is the macroscopic surface tension at a liquid−air interface at 298 K and γ^{o} is 1 cal mol^{-1} Å^{-2}, the square of the fraction of nonhydrogenic solvent atoms that are aromatic carbon atoms (carbon aromaticity),ϕ^2, and the square of the fraction of nonhydrogenic solvent atoms that are F, Cl, or Br (electronegative halogenicity), ψ^2. Provided that these descriptors are available for a given solvent (either from experiment or from some theoretical model, for example, a fragment model[97] or a molecular properties model[93,98,99]), the standard-state free energy of solvation of a given solute in that solvent can be calculated by universal SM*x* continuum solvation models.[100]

The G_{CDS} term is a sum of two terms

$$G_{\text{CDS}} = G^{[1]}_{\text{CDS}} + G^{[2]}_{\text{CDS}} \tag{2}$$

where

$$G^{[1]}_{\text{CDS}} = \sum_{\delta=1}^{3} S_\delta \sum_k A_k(\mathbf{R},\{r_k + r_{\text{S}}\}) \sum_j \sigma^{[1]}{}_{Z_k j\delta} f_{Z_{kj}}(\{Z_{k'},R_{k'k''}\}) \tag{3}$$

and where the index δ goes over the solvent descriptors n, α, and β, S_δ is the value of each of these three descriptors, the index k goes over all atoms in the solute, A_k is the solvent-accessible surface area (SASA) of atom k, which depends on the geometry, $\mathbf{R}$, of the solute and on a set of van der Waals solute radii r_k augmented by a solvent radius r_{S}, Z_k is the atomic number of atom k, the index j goes over the geometry-dependent functions $f_{Z_{kj}}$, which are used to distinguish between different types of functional groups in the solute that may be near atom k (their functional forms are described in detail elsewhere[24,26−30,33]), $R_{k'k''}$ is the interatomic distance between atoms $k'$and k'', and $\sigma^{[1]}{}_{Z_k j\delta}$ is an atomic surface tension coefficient (a parameter to optimize). In eq 3, $\{x\}$ denotes the set of all the quantities x. The second term in eq 2 is

$$G^{[2]}_{\text{CDS}} = \sum_{\delta=4}^{7} S_\delta\, \sigma^{[2]}_\delta \sum_k A_k(\mathbf{R},\{r_k + r_{\text{S}}\}) \tag{4}$$

where here the index δ goes over the solvent descriptors γ, β^2, ϕ^2, and ψ^2, and $\sigma^{[2]}_\delta$ is a molecular surface tension coefficient (this optimized parameter does not depend on atomic number). Note that in previous work[24,31−35] and in the present work, water has its own set of atomic surface tension coefficients, so neither eq 4 nor the solvent descriptors in eq 3 are needed for water, and the set of $\{\sigma^{[1]}_{Zj\delta}\}$ are denoted as $\{\sigma^{[1]}_{Zj,\text{water}}\}$. Thus, for water

$$G_{\text{CDS}} = \sum_k A_k(\mathbf{R},\{r_k + r_{\text{S}}\}) \sum_j \sigma^{[1]}_{Z_k j,\text{water}} f_{Z_k j}(\{Z_{k'},R_{k'k''}\}) \tag{5}$$

We note that the solvent dependence is entirely contained in the S_δ values; r_{S} is independent of solvent.

Parameter Optimization for SM5.43R. The parameters are optimized for a temperature of 298 K.

We optimized three types of parameters in SM5.43R: (i) the solvent radius used to compute the SASA in eqs 3−5, (ii) the empirical Coulomb radii for H, C, N, O, F, P, S, Cl, and Br used in the GB method, and (iii) the atomic and molecular surface tension coefficients, $\sigma^{[1]}_{Zj\delta}$, $\sigma^{[2]}_\delta$, and $\sigma^{[1]}_{Zj,\text{water}}$. The optimization is not based solely on least-squares fitting; it also involves trying to make the parameters correspond to the correct physics. In the rest of this section, we will detail the methods we used to determine each of these types of parameters, beginning with the determination of the solvent radius.

The solvent radius enters in the calculation of the SASAs, where it is added to the atomic van der Waals radii.[101] The van der Waals radii are given the values assigned by Bondi.[91] Various values for the solvent radii have been used in previous SM*x*-type ($x = 1-5.42$R) continuum solvation models. In SM1-SM3,[18−21,101] which were parametrized only for water, a solvent radius of 1.4 Å was used, a value that was justified elsewhere.[102−105] The SM4 model,[22,23] which was first developed

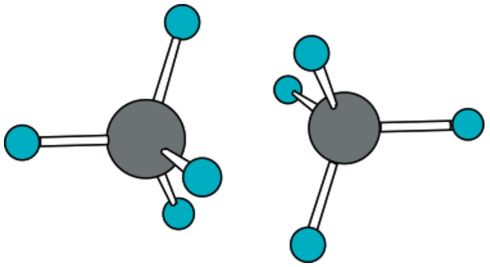

Figure 1. Methane dimer in a 3,3-staggered conformation. The system has D_{3d} symmetry.

for *n*-hexadecane and then later extended to general alkane solvents, used a solvent radius value of 2.0 Å in eq 3 and a solvent radius of 4.9 Å in eq 4. In SM5.4,[24−26,29] which was parametrized for water solvent, chloroform solvent, and then general organic solvents, a solvent radius of 1.7 Å (the average of 1.4 Å and 2.0 Å) was used in eq 3 and a solvent radius of 3.4 Å was used in eq 4. Finally, in SM5.0R,[28,33] SM5.2R,[30] SM5.42R,[31,32,34,36] and SM5CR,[35] which were parametrized for water and organic solvents, the solvent radius was set to zero because this choice yielded smaller root-mean square errors in predicted solvation energies than nonzero values of the solvent radius did. In this case, the SASA of an atom in the solute becomes the exposed area of the atom (i.e., the area of the atom-centered van der Waals sphere of the atom that does not overlap with any of the other atom-centered spheres of the atoms in the solute). Although a solvent radius of zero leads to a better fit to experimental solvation data, a nonzero solvent radius might be more physically meaningful because if the solvent radius actually corresponds to the size of a solvent molecule the resulting SASA is proportional to the number of solvent molecules in the first solvation shell.

In this work, we have chosen the value of the solvent radius so that SM5.43R yields a potential of mean force for the methane dimer in water that agrees as well as possible with the best one currently, which we take to be the PMF calculated by Shinto et al.[106] using an explicit solvent model. Because the dominant contribution to the free energy of solvation in this system is from the G_{CDS} term, it is an ideal system to study the effects of using various nonzero values of the solvent radius on the PMF. Note that the PMF, $W(R)$, computed from SM5.43 (the R in SM5.43R is dropped because the geometry of the solute in a PMF calculation does not necessarily correspond to the optimized gas-phase geometry) is

$$W(R) = \Delta G_{\mathrm{S}}^{\mathrm{o}}(R) + V(R) \tag{6}$$

where R is the coordinate specifying the geometry of the solute, $V(R)$ is the gas-phase potential energy of the solute at the coordinate value R, and $\Delta G_{\mathrm{S}}^{\mathrm{o}}(R)$ is the standard-state free energy of solvation of the solute at geometry R.

To compute $W(R)$ for the methane dimer, we first optimized the geometry of the methane dimer in a 3,3-staggered conformation (this is shown in Figure 1) using MP2[107]/MG3S.[60,108] For various values of the solvent radius, we computed $\Delta G_{\mathrm{S}}^{\mathrm{o}}(R)$ for various C−C internuclear distances, keeping all other internal coordinates fixed. For $V(R)$, we used the experimental effective spherical potential for the methane dimer of Reid et al.[109] We then define

$$\Delta W(R) = W(R) - W(R = \infty) \tag{7}$$

The well depth and well location of the resulting $\Delta W(R)$ are in Table 1. Table 1 shows that the continuum solvation model predicts the location of the minimum quite well for any value of r_{S} in the range of 0.3−0.5 Å. Furthermore, using a solvent radius of 0.4 Å gives excellent agreement of the well depth with the most reliable value.

TABLE 1: Location and Depth of the Well in the Potential of Mean Force of Methane Dimer Calculated by Several Methods

calculation	r_{e} (Å)	D_{e} (kcal/mol)[b]
Shinto et al.[a]	3.8	−0.66
SM5.43 (r_S = 0.3 Å)[c]	3.8	−0.38
SM5.43 (r_S = 0.4 Å)[c]	3.8	−0.65
SM5.43 (r_S = 0.5 Å)[c]	3.9	−0.84

[a] See ref 106. [b] Energy relative to the two methane molecules infinitely separated. [c] Using mPW1PW91/6-31G(d).

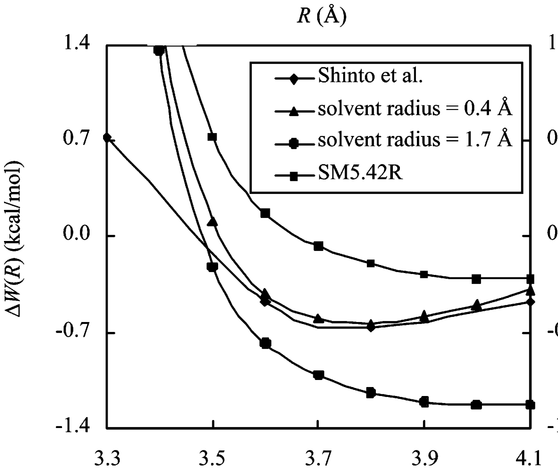

Figure 2. Potential of mean force, $W(R)$, of the methane dimer calculated by (a) Shinto et al., (b) SM5.42 using HF/6-31G(d), (c) SM5.43 using mPW1PW91/6-31G(d) and with a solvent radius of 0.4 Å, and (c) by SM5.43R with a solvent radius of 1.7 Å. In all cases, $W(R)$ is normalized to zero for $R = \infty$.

Figure 2 shows the potential of mean force of the methane dimer calculated by Shinto et al., by SM54.2 (where the solvent radius is zero), and by SM5.43 using a solvent radius of 0.4 or 1.7 Å. When a solvent radius of 0 Å is used, the shape of the well is too broad and the well is not deep enough (not shown in the figure). When a solvent radius of 1.7 Å is used, the shape of the well is also too broad and the well is too deep. Therefore, based on the results presented in Table 1 and Figure 2, we have chosen to use a solvent radius of 0.4 Å for SM5.43R and SM5.43 for both water and organic solvents. Although this value for the solvent radius is small, it leads to a more accurate PMF for methane dimer in water than when a larger solvent radius is used. Furthermore, our experience is that errors in predicted free energies of solvation increase as the value of the solvent radius increases. Thus a value of 0.4 Å is a good compromise between obtaining accurate free energies of solvation and maintaining a physically meaningful definition of the solvent-accessible surface area.

To determine the Coulomb radii and the atomic and molecular surface tension parameters, we used a training set of molecules containing H, C, N, O, F, P, S, Cl, and Br. To build this training set, we began with the training set[35] used to develop the SM5CR continuum solvation model. The SM5CR training set is more diverse than the SM5.42R one, as already mentioned in the Introduction. When the SM5CR training set was published,[35] it was stated to have 2141 absolute free energies of solvation for 278 neutral solutes containing H, C, N, O, F, P, S, Cl, Br, and I in 90 organic solvents and water, 76 transfer free energies of solvation between water and 12 organic solvents (which are a

TABLE 2: Mean-Unsigned Error (MUE) of the Aqueous Free Energy of Solvation (in kcal/mol) of Various Solute Classes Calculated by Various Continuum Solvation Models Using HF/6-31G(d)

solute class	no. data	SM5.42R	C-PCM/98	C-PCM/03	D-PCM/03	IEF-PCM/03	SM5.43R
neutral H, C, N, O, F[a]	171	0.57	0.82	0.79	1.21	0.76	0.51
Cl, Br, S, and P neutrals[b]	86	0.49	1.60	1.64	1.69	1.64	0.48
all neutrals	257	0.54	1.08	1.07	1.37	1.06	0.50
all H, C, N, O, F ions[a]	32	5.20	5.11	7.66	9.18	7.63	4.91
all Cl, Br, P, S ions[b]	15	4.03	3.42	2.16	13.12	2.18	4.10
all ions	47	4.83	4.57	5.90	10.44	5.89	4.65

[a] Solutes containing at most the five listed elements. [b] Solutes containing at least one of these elements plus, in most cases, elements from the previous row.

TABLE 3: Mean-Unsigned Error (MUE) of the Aqueous Free Energy of Solvation (in kcal/mol) of Various Solute Classes Calculated by Various Continuum Solvation Models Using B3LYP/6-31G(d)

		B3LYP				mPW1PW91	
solute class	no. data	6-31G(d)/ C-PCM/98	6-31G(d)/ D-PCM/03	6-31G(d)/ IEF-PCM/03	6-31G(d)/ SM5.43R	6-31G(d)/ SM5.43R	6-31+G(d)/ SM5.43R
neutral H, C, N, O, F[a]	171	0.86	0.89	0.88	0.50	0.52	0.54
Cl, Br, S, and P neutrals[b]	86	1.88	1.83	1.96	0.47	0.47	0.52
all neutrals	257	1.20	1.21	1.24	0.49	0.50	0.54
all H, C, N, O, F ions[a]	32	5.48	9.89	7.83	5.06	5.03	5.09
all Cl, Br, P, S ions[b]	15	3.30	13.10	2.36	3.70	3.65	3.78
all ions	47	4.78	10.91	6.08	4.63	4.59	4.67

[a] Solutes containing at most the five listed elements. [b] Solutes containing at least one of these elements plus, in most cases, elements from the previous row.

subset of the 90 organic solvents) for 54 additional neutral solutes (i.e., these 54 solutes do not have experimental absolute free energies of solvation in the training set) containing H, C, N, O, F, P, S, Cl, Br, and I, and 49 aqueous free energies of solvation for 49 ionic solutes also containing H, C, N, O, F, P, S, Cl, Br, and I. As a first step, we corrected two errors in this training set. First, the solute tetrafluoromethane was counted twice as a solute, once with experimental free energies of solvation in water and octanol and once with an experimental free energy of solvation in water. Second, the solutes 3,5-dimethylpyridine and 4-ethylpyridine appeared twice in the training set, once as solutes with experimental free energies of solvation in water and *n*-hexadecane and once as solutes with experimental transfer free energies of solvation (so these two solutes were being counted as solutes with absolute free energy of solvation data and as two additional solutes with free energy of transfer data). Then, we eliminated all compounds containing iodine and the ionic solute, PH_2^-, which is considered atypical of the kinds of phosphorus compounds that one generally encounters.

We then added experimental free energy of self-solvation data, which were assembled from vapor pressure data in previous work.[4] We removed all iodine-containing compounds and self-solvation free energy data that were already present in the training set, which yielded 71 new data.

Next we added the experimental absolute free energies of solvation and free energies of transfer for nitromethane and γ-butryolactone that are available from the Medchem database[110] (54 new absolute free energy data and 4 new transfer free energy data).

Finally, we examined the accuracy of certain solvation data for neutrals and ions. For neutrals, we re-evaluated the data in our training set that were taken from Rohrschneider,[111] which had been re-evaluated by Carr and co-workers.[112] For each new free energy of solvation measurement made by Carr and co-workers that differed from Rohrschneider's measurement by 0.20 kcal/mol or more, we used the average of the two measurements. These changes are summarized in Table S-1 of the Supporting Information.

For ions, we changed all of the data. The experimental aqueous free energies of solvation of ions depend on the value that one employs for the aqueous free energy of solvation of the proton.[113,114] In previous SM*x* models, we used the previously accepted value of −259.50 kcal/mol.[115−121] For SM5.43R, we use the value determined from experiments of Tissandier and co-workers,[122] which is −263.98 kcal/mol. Note that this change makes the experimental aqueous free energies of solvation of anions more positive by 4.48 kcal/mol and the experimental aqueous free energies of solvation of cations more negative by 4.48 kcal/mol. For comparison, we note that this change yields −105.5 kcal/mol for the free energy of hydration of hydroxide ion, which compares well to a recent first principles determination of −106.4 kcal/mol (using our choice of standard states).[123]

With the modifications described above, the training set has 2237 absolute free energies of solvation of 295 solutes containing H, C, N, O, F, P, S, Cl, and Br in 90 organic solvents and water, 77 transfer free energies of solvation between 12 organic solvents and water for 51 additional neutral solutes, 2 transfer free energies of solvation for 3,5-dimethylpyridine and 4-ethylpyridine, and 47 aqueous free energies of solvation for 47 ionic solutes.

Further details of the optimization of the parameters are summarized in the Supporting Information, which also gives full tables of parameters.

Results and Discussion

We first examine the accuracy of SM5.43R for predicting aqueous free energies of solvation of neutral and ionic solutes. We will present SM5.43R parametrizations using four levels of electronic structure theory: HF/6-31G(d), B3LYP/6-31G(d), mPW1PW91/6-31G(d), and mPW1PW91/6-31+G(d). To place the results in perspective and to assess the relative performance of different approaches, we will compare these four models to SM5.42R/HF/6-31G(d),[34] C-PCM/98[82] with HF/6-31G(d) and B3LYP/6-31G(d), C-PCM/03[16] with HF/6-31G(d), D-PCM/03[76−78] with HF/6-31G(d) and B3LYP/6-31G(d), and IEF-PCM/03[79−81] with HF/6-31G(d) and B3LYP/6-31G(d).

TABLE 4: Mean-Unsigned Error (MUE) of the Free Energy of Solvation (in kcal/mol) of Solutes in the Indicated Solvent Calculated by Various Continuum Solvation Models Using HF/6-31G(d)

solute class	no. data[a]	SM5.42R	C-PCM/98	C-PCM/03	D-PCM/03	IEF-PCM/03	SM5.43R
acetonitrile	7	0.44	4.64	4.61	4.56	4.64	0.41
aniline	9	0.42	7.00	6.93	7.06	7.23	0.52
benzene	68	0.48	4.48	4.52	4.83	4.89	0.63
carbon tetrachloride	72	0.47	4.23	4.27	4.53	4.56	0.48
chlorobenzene	37	0.60	3.45	3.46	3.62	3.72	0.54
chloroform	96	0.58	4.12	4.19	4.38	4.43	0.54
cyclohexane	83	0.37	2.08	2.15	2.35	2.36	0.41
dichloroethane	38	0.49	3.28	3.34	3.42	3.51	0.44
diethyl ether	62	0.51	2.70	2.84	3.03	3.20	0.60
dimethyl sulfoxide	7	0.96	2.86	2.82	2.82	2.87	0.75
ethanol	8	1.47	2.54	2.84	3.37	2.76	1.34
heptane	60	0.33	1.95	2.02	2.27	2.30	0.34
methylene chloride	11	0.48	1.80	2.78	2.85	2.90	0.48
nitromethane	7	0.78	2.29	2.23	2.22	2.30	0.72
tetrahydrofuran	7	0.32	2.61	2.55	2.62	2.72	0.31
toluene	49	0.31	3.09	3.11	3.40	3.47	0.41
all 16 solvents	621	0.48	3.31	3.38	3.60	3.65	0.50
all 74 other solvents	1359	0.47	na[b]	na[b]	na[b]	na[b]	0.51
self-solvation energies[c]	76 (16)[d]	0.54 (0.60)	(3.61)	(3.77)	(3.86)	(3.93)	0.54 (0.56)

[a] Number of experimental data in this solvent. [b] Not available. [c] Standard-state free energy of solvation of a solute in a pure solution of the solute. [d] The PCM-type continuum solvation models are explicitly defined for 16 of the 76 solvents used to compute the self-solvation energies, and the MUEs for these 16 solvents are given in parentheses.

TABLE 5: Mean-Unsigned Error (MUE) of the Free Energy of Solvation (in kcal/mol) of Solutes in the Indicated Solvent Calculated by Various Continuum Solvation Models

		B3LYP				mPW1PW91	
solute class	no. data[a]	6-31G(d)/ C-PCM/98	6-31G(d)/ D-PCM/03	6-31G(d)/ IEF-PCM/03	6-31G(d)/ SM5.43R	6-31G(d)/ SM5.43R	6-31+G(d)/ SM5.43R
acetonitrile	7	5.22	5.13	5.22	0.37	0.41	0.51
aniline	9	7.55	7.54	7.73	0.50	0.50	0.57
benzene	68	4.81	5.09	5.15	0.62	0.63	0.63
carbon tetrachloride	72	4.53	4.76	4.80	0.48	0.49	0.48
chlorobenzene	37	3.90	4.01	4.12	0.51	0.55	0.55
chloroform	96	4.56	4.75	4.82	0.53	0.55	0.52
cyclohexane	83	2.30	2.52	2.53	0.40	0.41	0.43
dichloroethane	38	3.85	3.94	4.05	0.43	0.44	0.51
diethyl ether	62	3.26	3.49	3.68	0.62	0.63	0.62
dimethyl sulfoxide	7	3.42	3.31	3.43	0.57	0.61	0.57
ethanol	8	2.17	2.45	1.87	1.30	1.35	1.28
heptane	60	2.21	2.46	2.49	0.33	0.34	0.37
methylene chloride	11	2.01	3.16	3.26	0.45	0.47	0.42
nitromethane	7	2.83	2.64	2.84	0.67	0.71	0.79
tetrahydrofuran	7	3.12	3.10	3.20	0.32	0.33	0.32
toluene	49	3.43	3.67	3.75	0.40	0.41	0.43
all 16 solvents	621	3.68	3.89	3.96	0.49	0.51	0.51
all 74 other solvents	1359	na[b]	na[b]	na[b]	0.50	0.52	0.53
self-solvation energies[c]	76 (16)[d]	(3.94)	(4.14)	(4.20)	0.50 (0.49)	0.53 (0.51)	0.56 (0.55)

[a] Number of experimental data in this solvent. [b] Not available. [c] Standard-state free energy of solvation of a solute in a pure solution of the solute. [d] The PCM-type continuum solvation models are explicitly defined for 16 of the 76 solvents used to compute the self-solvation energies and the MUEs for these 16 solvents are given in parentheses.

Tables 2 and 3 give the mean-unsigned errors (MUEs), as compared to experiment, of the aqueous free energies of solvation for various solute classes of the aqueous training set (i.e., the subset of the full training set containing only absolute aqueous free energies of solvation for neutral and ionic solutes). Over all 257 neutral solutes, the MUEs calculated by SM5.42R (0.54 kcal/mol) and SM5.43R (0.49−0.50 kcal/mol) with the 6-31G(d) basis set are approximately a factor of 2 smaller than the MUE (1.07−1.37 kcal/mol) calculated by the PCM methods. For ionic solutes, the C-PCM methods have MUEs of 4.6−5.9 kcal/mol, D-PCM/03 has MUEs of 10.4−10.9 kcal/mol, IEF-PCM has MUEs of 5.9−6.1 kcal/mol, SM5.42R has a MUE of 4.8 kcal/mol, and SM5.43R has MUEs of 4.6−4.7 kcal/mol. Table 3 also shows that the MUEs of the neutrals calculated by SM5.43R/mPW1PW91/6-31+G(d) are comparable to the SM5.43R methods without diffuse functions; this confirms the success of the redistributed Löwdin population analysis[61] that was designed to make population analysis more realistic in the presence of diffuse functions and the success of the CM3 charge model that was designed to be less sensitive to basis set fluctuations in calculated bond orders.[43,44]

We next investigate the accuracy of SM5.43R for predicting free energies of solvation for solutes in organic solvents. Our organic training set of data consists of free energies of solvation in 90 organic solvents. The PCM methods are explicitly defined for 16 of these solvents in *Gaussian 98* and *Gaussian 03* (i.e., the user does not need to supply any data other than the name for these solvents). Tables 4 and 5 give the MUE of the free energy of solvation in these 16 solvents. They also give the total MUE of the free energy of solvation calculated by SM5.42R and SM5.43R in the other 74 organic solvents and for the 76 free energies of self-solvation (note that the PCM methods are explicitly defined for 16 of these 76 solvents). For most solvents, SM5.42R and SM5.43R yield similar MUEs to the corresponding MUEs of the aqueous free energy of solvation. Over the remaining solvents (the last row in Table

TABLE 6: Mean-Unsigned Error (MUE) of the Free Energy of Transfer (in kcal/mol) of Solutes in the Training Set Used in This Work between Two Organic Solvents

		HF	HF	B3LYP	mPW1PW91	
		6-31G(d)	6-31G(d)	6-31G(d)	6-31G(d)	6-31+G(d)
	no. data	SM5.42R	SM5.43R	SM5.43R	SM5.43R	SM5.43R
all H, C, N, O data	22228	0.46	0.47	0.46	0.47	0.46
all F, Cl, Br, S, P data	730	0.52	0.56	0.55	0.55	0.57
all data	22958	0.47	0.47	0.46	0.47	0.46

TABLE 7: Mean-Unsigned Error (MUE) of the Free Energy of Transfer (in kcal/mol) of Solutes in the Training Set Used in This Work between Water and Organic Solvents

		HF	HF	B3LYP	mPW1PW91	
		6-31G(d)	6-31G(d)	6-31G(d)	6-31G(d)	6-31+G(d)
	no. data	SM5.42R	SM5.43R	SM5.43R	SM5.43R	SM5.43R
all H, C, N, O data	1689	0.50	0.50	0.49	0.50	0.51
all F, Cl, Br S, P data	236	0.54	0.62	0.61	0.62	0.58
all data	1925	0.50	0.52	0.51	0.52	0.52

4), the MUE is 0.54 kcal/mol from both SM5.42R and SM5.43R. Both SM5.43R and SM5.42R perform significantly better than the PCM methods for predicting free energies of solvation in organic solvents. With the exception of ethanol solvent, the MUEs of the two SM*x* methods are, on average, a factor of 6−7 smaller than the corresponding MUEs calculated by C-PCM/98, C-PCM/03, D-PCM/03, and IEF-PCM/03. For ethanol solvent, the errors from SM5.42R and SM5.43R are only about a factor of 1.5−2 smaller than the PCM methods. This is in part because there are only eight experimental data in ethanol, and of those eight data, the solutes γ-butryolactone and butanone are large outliers, with errors of approximately 2.0−3.0 and 1.5 kcal/mol, respectively. If there were more data available in ethanol, these two outliers would probably not dominate the MUE as much as they currently do. Note that in the 13 other organic solvents in our training set that are alcohols, we obtain MUEs of approximately 0.65 kcal/mol over 330 data.

We also examined the ability of SM5.43R to predict free energies of transfer for solutes between two organic solvents and between water and an organic solvent. The training set used to parametrize SM5.43R has 2237 experimental absolute free energies of solvation for solutes in 90 organic solvents and water, and we used this to construct a test set of experimental free energies of transfer between two organic solvents (22958 data) and between water and an organic solvent (1925 data), i.e., for each solute with experimental absolute solvation energies in N solvents, where N is greater than one, we obtained $0.5N(N - 1)$ free energies of transfer. Table 6 gives the MUEs of the free energies of transfer between two organic solvents calculated by SM5.42R and the SM5.43R methods for various subsets of the 22958-data test set, and Table 7 gives the MUEs of the free energies of transfer between water and an organic solvent for various subsets of the 1925-data test set. Comparison of the MUEs given in these tables to the MUEs given in Tables 4 and 5 shows that SM5.43R predicts free energies of transfer between two solvents with the same accuracy as it predicts absolute free energies of solvation in organic solvents. We also investigated the option of including these 22958 organic/organic and 1925 water/organic transfer free energy data to our training set. We found, however, that the MUEs over these data did not significantly improve when they were included in the training set, and so we did not include these data in the final parameter optimization.

Because the 22958-data and 1925-data test sets are not independent of the SM5.43R training set, we carried out additional calculations designed to test how well the model can predict data not used for training. For this purpose, we created a smaller training set of data from the SM5.43R training set by randomly removing approximately 25% of the data from the actual training set (see below), and then we parametrized SM5.43R with the B3LYP/6-31G(d) method using the smaller training set, and tested the new method using the data removed from the original training set. This smaller training set was obtained by first placing each solute data point into solute class files, where a solute class specifies some common characteristic of a set of solutes, e.g., alcohols, esters, amines, etc. The solutes, which are specified by an alphanumeric name in our training set, were placed in alphanumeric order in the appropriate solute class file. For each solute class, N_{Rm} random integer numbers were generated, where N_{Rm} was determined by one of the following formulas:

$$N_{\mathrm{Rm}} = \mathrm{INT}\left(\frac{N_{\mathrm{dat}}}{4}\right) + 1, \text{ if } N_{\mathrm{dat}}(\mathrm{mod}\ 4) > 1 \tag{8}$$

or

$$N_{\mathrm{Rm}} = \mathrm{INT}\left(\frac{N_{\mathrm{dat}}}{4}\right), \text{ if } N_{\mathrm{dat}}(\mathrm{mod}\ 4) \leq 1 \tag{9}$$

where INT returns the integer portion of its argument, $N_{\mathrm{dat}}(\mathrm{mod}\ 4)$ implies “modulo 4”, and N_{dat} is the number of data in a given solute class. Each random number was then assigned to an entry in a given solute class file, and all of these entries were removed.

Table 8 gives the mean-unsigned errors (MUEs), as compared to experiment, of the free energy of solvation in water and organic solvents for various solute classes using SM5.43R/B3LYP/6-31G(d) parametrized against the entire SM5.43R training set (these columns are labeled “full”) and SM5.43R/B3LYP/6-31G(d) parametrized against the smaller training set (these columns are labeled “partial”). The third and fourth columns in Table 8 correspond to MUEs of the free energies of solvation of solutes from the smaller training set of data, and the sixth and seventh columns correspond to MUEs of the free energies of solvation of the test-set solutes that were randomly removed. Columns three and four in Table 8 show that SM5.43R trained on the actual training set (SM5.43R/full) or on a smaller subset of it (SM5.43R/partial) yield very similar MUEs. The SM5.43R/partial method predicts free energies of solvation of solutes not used to train the model with similar accuracy as SM5.43R/full, which was trained on these solutes (see columns six and seven). In the rest of the paper, we consider only final models parametrized against all the data.

TABLE 8: Mean-Unsigned Error (MUE) of the Free Energy of Solvation (in kcal/mol) of Various Solute Classes in Water and Organic Solvents Using SM5.43R/B3LYP/6-31G(d) Parameterized against the Entire SM5.43R Training Set (Full) and SM5.43R/B3LYP/6-31G(d) Parameterized against the Training Set Obtained by Randomly Removing 25% of the Data from the SM5.43R Training Set (Partial)

solute class	no. data[a]	full[b]	partial[c]	no. data[d]	full[b]	partial[c]
In Water						
neutral H, C, N, O, F	130	0.51	0.50	41	0.45	0.51
Cl, Br, S, and P neutrals	68	0.50	0.50	18	0.35	0.37
all aqueous neutral data	198	0.51	0.50	59	0.42	0.47
In Organic Solvents						
all H, C, N, O absolute free energies	1296	0.46	0.46	437	0.47	0.48
al H, C, N, O transfer free energies	29	0.42	0.41	8	0.39	0.51
all H, C, N, O data	1325	0.46	0.46	445	0.47	0.48
F, Cl, Br, S absolute free energies	157	0.61	0.59	53	0.63	0.68
F, Cl, Br, S transfer free energies	25	0.57	0.59	8	0.37	0.39
H, C, N, O, F, S, Cl data	182	0.61	0.59	61	0.60	0.64
all P absolute free energies	27	1.52	1.50	10	1.18	1.25
all P transfer free energies	7	0.50	0.60	2	0.98	0.98
all P data	34	1.31	1.32	12	1.14	1.20
all absolute free energies	1480	0.50	0.49	500	0.50	0.52
all transfer free energies	61	0.49	0.50	18	0.45	0.51
all organic data	1541	0.50	0.49	518	0.50	0.52

[a] Number of data in a given solute class obtained by randomly removing 25% of the data from the SM5.43R training set. [b] MUE of a given solute class using SM5.43R/B3LYP/6-31G(d) parametrized against the entire SM5.43R training set. [c] MUE of a given solute class using SM5.43R/B3LYP/6-31G(d) parametrized against the training set obtained by randomly removing 25% of the data from the SM5.43R training set. [d] Number of data in a given solute class that was randomly removed from the SM5.43R training set.

TABLE 9: Summary of the Free Energy of Transfer of Acetanilide from Water to Diethyl Ether Calculated by SM5.43R and C-PCM/98[a]

contribution	SM5.43R	C-PCM/98
In Water		
ΔE	2.23	1.86
G_P	−11.21	−10.19
ΔG_{EP}	−8.98	−8.33
G_{cav}		19.53
G_{disp}		−20.59
G_{rep}		3.48
G_{CDS}	0.30	2.42[b]
ΔG_S^o	−8.68[c]	−5.91[d]
In Diethyl Ether		
ΔE	1.15	0.62
G_P	−7.74	−5.03
ΔG_{EP}	−6.59	−4.41
G_{cav}		15.21
G_{disp}		−16.70
G_{rep}		2.62
G_{CDS}	−3.56	1.13[b]
ΔG_S^o	−10.15[c]	−3.28[d]
free energy of transfer	−1.47[e]	2.63[e]

[a] All quantities in this table are in units of kcal/mol. [b] In PCM methods, G_{CDS} is replaced by $G_{nonelect} \equiv G_{cav} + G_{disp} + G_{rep}$. [c] $\Delta G_S^o = \Delta G_{EP} + G_{CDS}$. [d] $\Delta G_S^o = \Delta G_{EP} + G_{nonelect}$. [e] The experimental value of the free energy of transfer is −0.74 kcal/mol.

As an example of our predictions for ions, we note that for OH^- in water, SM5.43R yields −107.0 kcal/mol with either mPW1PW91/6-31G(d) or B3LYP/6-31G(d), in good agreement with the experimental value of −105.5 kcal/mol discussed in the parameter optimization section. This is within the range of reliability for the first-principles determination,[123] which is based on single configurations of the solute−solvent clusters, whereas the continuum model is effectively averaged over solvent configurations. The mean unsigned errors for ions should be considered in light of the large magnitudes of the ionic solvation energies and the large uncertainties in the experimental values. We estimate a *typical* uncertainty in experimental ionic free energies of solvation of 4−5 kcal/mol, in comparison to 0.2 kcal/mol for a *typical* experimental neutral value.

We conclude this section with a detailed examination of the calculation of the standard-state transfer free energy of solvation of acetanilide from water to diethyl ether calculated by SM5.43R and C-PCM/98 using the B3LYP/6-31G(d) method. This is an interesting test case because acetanilide exhibits a relatively broad range of functionality. The results of these calculations are summarized in Table 9. In Table 9, the contributions to the standard-state free energies of solvation of acetanilide in water and in diethyl ether from SM5.43R and C-PCM/98 are given. The calculated standard-state aqueous free energy of solvation of acetanilide is −8.68 kcal/mol from SM5.43R and −5.91 kcal/mol from C-PCM/98. The calculated standard-state free energy of solvation of acetanilide in diethyl ether is −10.15 kcal/mol from SM5.43R and −3.28 kcal/mol from C-PCM/98. The resulting standard-state free energy of transfer of acetanilide from water to diethyl ether is −1.47 kcal/mol from SM5.43R and 2.63 kcal/mol from C-PCM/98; the experimental value is −0.74 kcal/mol.

With SM5.43R, the contributions to ΔG_{EP} and G_{CDS} can be partitioned into contributions from each of the atoms in the solute.[31,46,124,125] The contribution to ΔG_{EP} from atom k, denoted as ΔG_{EP}^k, is computed as

$$\Delta G_{EP}^k = \left(\frac{\Delta G_{EP}}{G_P}\right) G_P^k \tag{10}$$

where ΔG_{EP} is the sum of the first two terms on the right-hand side of eq 1 and G_P^k is the generalized Born contribution to G_P from atom k. Furthermore, the atomic contributions to ΔG_{EP} can be computed using the gas-phase-optimized wave function; that is, we can compute the effect of the dielectric medium on the solute without allowing the solute's wave function to relax.[124,125] In this case, the prefactor in front of G_P^k in eq 10 is unity because ΔE (see eq 1) is zero. Figure 3a shows the gas-phase CM3 partial atomic charges of the atoms in acetanilide and the contribution to ΔG_{EP} from each atom in acetanilide using the gas-phase wave function (and a dielectric constant of 78.3 for water). Figure 3b shows the liquid-phase CM3 charges, the atom-by-atom contributions to ΔG_{EP} in water solvent, and the atom-by-atom contributions to G_{CDS} in water. Figure 4a shows the same data as Figure 3a, except that the contribution to ΔG_{EP} is computed using the dielectric constant for diethyl ether (which is 4.24). Figure 4b shows the same data as Figure 3b, except

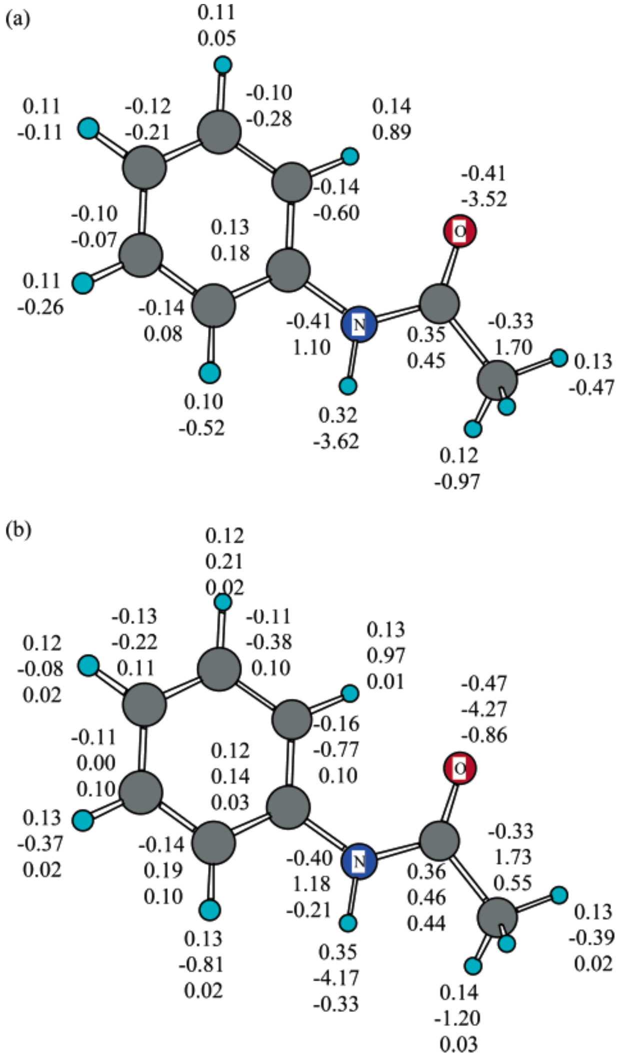

Figure 3. (a) Gas-phase CM3 partial atomic charges and atomic contributions to G_P using the gas-phase wave function for acetanilide in water solvent. There are two numbers next to each symmetry-unique atom in acetanilide. The first number is the gas-phase CM3 partial atomic charge and the second number is the contribution to ΔG_{EP} using the unrelaxed, gas-phase wave function and the dielectric constant for water. (b) CM3 partial atomic charges, atomic contributions to G_P, and atomic contributions to G_{CDS} for acetanilide in water solvent. There are three numbers next to each symmetry-unique atom in acetanilide. The first number is the CM3 partial atomic charge of the atom in water. The second and third numbers are the contributions to ΔG_{EP} and G_{CDS} from the atoms in acetanilide in water solvent.

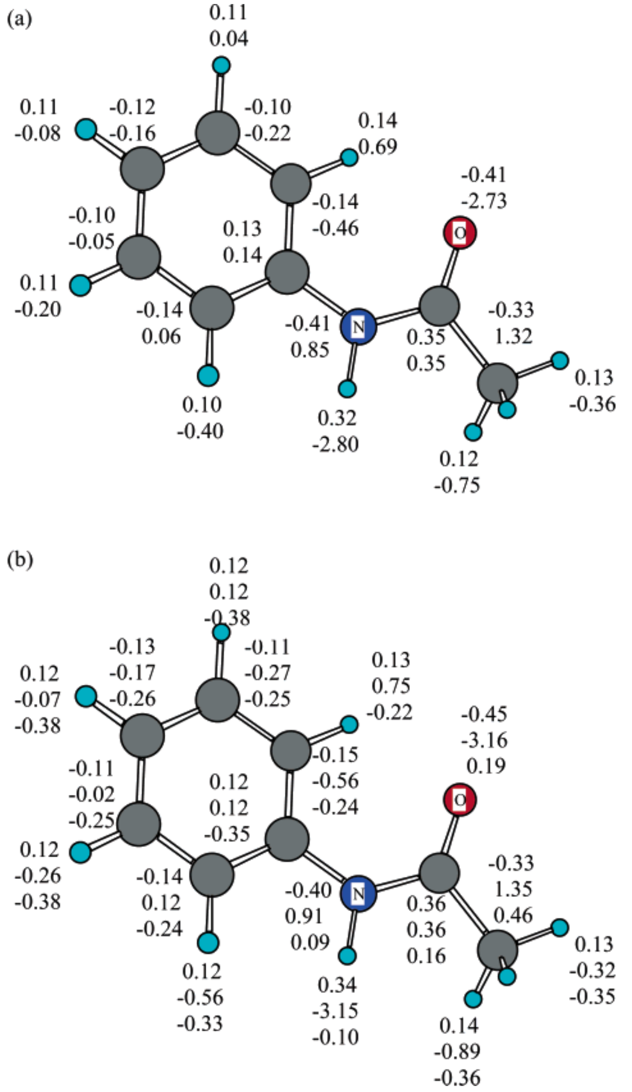

Figure 4. (a) Same as Figure 3a, but for diethyl ether solvent. (b) Same as Figure 3b, but for diethyl ether solvent.

that the solvent is diethyl ether. Figure 3, parts a and b, shows that, upon solvation in water, the CM3 charges on the carbon atoms in the benzene ring and the oxygen atom become more negative, and the charge on the nitrogen atom and the carbonyl carbon become more positive. The changes in the atomic contributions to ΔG_{EP} are most significant for the hydrogen bonded to the nitrogen and for the oxygen atom. The contributions to G_{CDS} are small for most atoms in acetanilide in water; the carbon atom in the methyl group and the carbonyl carbon atom yield positive contributions to G_{CDS}, which are partially canceled out by the contribution to G_{CDS} from the hydrogen bonded to the nitrogen atom. Figure 4, parts a and b, shows that the changes in the CM3 charges upon solvation in diethyl ether are similar to the changes in the charges upon solvation in water. However, due to the smaller dielectric constant, the changes in the contributions to ΔG_{EP} are smaller in diethyl ether than they are in water (the values of G_P and ΔG_{EP} are −11.21 and −8.98 kcal/mol, respectively, in water and −7.74 and −6.59 kcal/mol, respectively, in diethyl ether). The contributions to G_{CDS} are larger in diethyl ether than they are in water, particularly for the carbon atoms and hydrogen atoms in the benzene ring. The predicted value of G_{CDS} is 0.30 kcal/mol in water and −3.56 kcal/mol in diethyl ether. Thus, the free energy calculations from SM5.43R show that the partitioning of acetanilide between water and diethyl ether are due to a competition between dominant bulk-electrostatic interactions between water and acetanilide and bulk- and nonbulk-electrostatic interactions between diethyl ether and acetanilide.

Concluding Remarks

We have presented a new continuum solvation model based on the generalized Born approximation and atomic surface tensions that provide empirical corrections to the electrostatics and are proportional to the SASAs of the atoms of the solute. The new method, called SM5.43R, is as accurate as the SM5.42R solvation model for predicting free energies of solvation of organic solutes in water and organic solvents. The SM5.43R solvation model uses a solvent radius of 0.4 Å, and we know (from the SM5.42R parametrizations) that using a nonzero value for the radius results in slightly increased errors in predicted free energies of solvation. However, using a solvent radius of 0.4 Å leads to a more accurate potential of mean force for the methane dimer in water than when using a solvent radius of zero. Moreover, because SM5.43R is trained on a larger, more diverse, and more accurate training set of data than SM5.42R and uses a charge model that was parametrized in a more stable way against a broader training set, it should be a more robust model than SM5.42R. Both SM5.42R and SM5.43R predict more accurate aqueous free energies of solvation of neutral solutes than C-PCM/98, C-PCM/03, D-PCM/03, and IEF-PCM/03 by about a factor of 2 or more. For the free energies of solvation of organic solutes in various organic solvents, SM5.43R and SM5.42R are more accurate than both the PCM models by even larger margins, typically by a factor of 6 to seven. The success of the SM5.43R model is very encouraging, and we plan to carry out SM5.43R parametrizations for several other levels of theory and basis sets in subsequent work.

Availability of SM5.43R. The SM54.3R parameters for HF/6-31G(d) and B3LYP/6-31G(d) are available in the GAMESSPLUS[126] program, which is an add-on module to the GAMESS[127] program. The parameters for these two cases and also for the two mPW1PW91 cases are also available in the HONDOPLUS[128,129] program. Note also that geometry optimizations can be carried out in liquid solution (in this case, the "R" is dropped from the notation) using analytical free energy gradients.[45] In addition, numerical Hessian calculations based on analytic free energy gradients can be carried out with either of these two programs. For more information on GAMESSPLUS and HONDOPLUS, see http://comp.chem.umn.edu/software.

The new methods are also available for *Gaussian* users in a new program called SMXGAUSS.[130] This program can read a Gaussian output file corresponding to a gas-phase calculation of a given solute and carry a out single-point calculation with SM5.43R or a geometry optimization or a Hessian calculation with SM5.43. In addition, for users who have a *Gaussian 03* executable, SMXGAUSS can be used in conjunction with the "External" option introduced in *Gaussian 03*. This allows for liquid-phase geometry optimizations with the powerful optimizers available in *Gaussian*. For more information on SMXGAUSS, see http://comp.chem.umn.edu/smxgauss.

Although the present article is restricted to mPW1PW91/MIDI! gas-phase geometries, all of these programs allow liquid-phase optimizations with analytic gradients of both the electrostatic and CDS terms for all of the combinations of electronic structure method and basis set for which parameters are available.

Acknowledgment. This work was supported by a Department of Defense Multidisciplinary University Research Initiative (MURI) grant managed by the Army Research Office.

Supporting Information Available: Further details of the optimization of the parameters as well as full tables of parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

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

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

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

(4) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2000**, *104*, 4726.

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

(6) Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.

(7) Reynolds, C. H. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 738.

(8) Grant, D. J.; Higuchi, T. *Solubility Behavior of Organic Compounds*; John Wiley & Sons: New York, 1990.

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

(10) Rivail, J.-L.; Rinaldi, D. In *Computational Chemistry, Review of Current Trends*; Leszczynski, J., Ed.; World Scientific: New York, 1996; Vol. 1, pp 139−174.

(11) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074.

(12) Amovilli, C.; Barone, V.; Cammi, R.; Cancès, E.; Cossi, M.; Mennucci, B.; Pornelli, C. S.; Tomasi, J. *Adv. Quantum Chem.* **1998**, *32*, 227.

(13) Tomasi, J.; Cammi, R.; Mennucci, B.; Cappelli, C.; Corni, S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697.

(14) Chipman, D. M. *J. Chem. Phys.* **2002**, *116*, 10129.

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

(16) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(17) Luque, F. J.; Curutchet, C.; Muñoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I.; Morreale, A.; Gelpi, J. L.; Orozco, M. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3827.

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

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

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

(21) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 629.

(22) Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1995**, *117*, 1057.

(23) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1995**, *99*, 7137.

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

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

(26) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.

(27) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 2061.

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

(29) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85.

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

(31) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. D. *J. Chem. Phys.* **1998**, *109*, 9117. Errata: **1999**, *111*, 5624.

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

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

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

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

(36) Winget, P.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2002**, *106*, 5160.

(37) Daudel, R. *Quantum Theory of Chemical Reactivity*; Reidel: Dordecht, The Netherlands, 1973.

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

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

(40) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(41) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

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

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

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

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

(46) Chuang, Y.-Y.; Radhakrishnan, M. L.; Fast, P. L.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 4893.

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

(48) Perdew, J. P. In *Electronic Structure of Solids '91*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; pp 11−20.

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

(50) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192.

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

(52) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *110*, 724.

(53) Mayer, I. *Chem. Phys. Lett.* **1983**, *97*, 270.

(54) Mayer, I. *Chem. Phys. Lett.* **1985**, *97*, 210.

(55) Mayer, I. *Int. J. Quantum Chem.* **1986**, *29*, 73.

(56) Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.

(57) Golebiewski, A.; Rzecowska, E. *Acta Phys. Pol.* **1974**, *45*, 563.

(58) Baker, J. *Theor. Chim. Acta* **1985**, *68*, 221.

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

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

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

(62) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(63) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.

(64) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(65) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(66) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.

(67) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(68) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(69) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(70) Stephens, P. J.; Delvin, J. F.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

(71) Clark, T.; Chandrasekhar, J.; Sptiznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.

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

(73) Lynch, B. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2003**, in press.

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

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

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

(77) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(78) Cossi, M.; Barone, R.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.

(79) Cancès, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

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

(81) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

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

(83) 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.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; 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.

(84) 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.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision A.1; Gaussian, Inc.: Pittsburgh, PA, 2003.

(85) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43.

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

(87) Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1995**, *16*, 1449.

(88) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622.

(89) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1998**, *109*, 10543.

(90) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 1611.

(91) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.

(92) 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.

(93) 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; pp 83−134.

(94) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.

(95) Abraham, M. H. *J. Phys. Org. Chem.* **1993**, *6*, 660.

(96) Abraham, M. H.; Weatherby, P. K. *J. Pharm. Sci.* **1994**, *83*, 1087.

(97) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 835.

(98) Taft, R. W.; Murray, J. S. In *Quantitative Treatment of Solute/Solvent Interactions*; Theoretical and Computational Chemistry 1; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 55−82.

(99) Murray, J. S.; Politzer, P. In *Quantitative Treatment of Solute/Solvent Interactions*; Theoretical and Computational Chemistry 1; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 243−289.

(100) Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvent Descriptor Database*, Version July 9, 1999; University of Minnesota: Minneapolis, MN, 1999.

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

(102) Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. *J. Phys. Chem.* **1975**, *79*, 2239.

(103) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. *Science* **1985**, *229*, 834.

(104) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.

(105) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci., USA* **1987**, *84*, 3086.

(106) Shinto, H.; Morisada, S.; Miyahara, M.; Higashitani, K. *J. Chem. Eng. Jpn.* **2003**, *36*, 57.

(107) Møller, C. M.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

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

(109) Reid, B. P.; O'Loughlin, M. J.; Sparks, R. K. *J. Chem. Phys.* **1986**, *83*, 5656.

(110) Leo, A. J. *Pomona Medchem Database*; Biobyte Corporation: Claremont, CA, 1994.

(111) Rohrschneider, L. *Anal. Chem.* **1973**, *45*, 1241.

(112) Park, J. H.; Hussam, A.; Couasnon, P.; Fritz, D.; Carr, P. W. *Anal. Chem.* **1987**, *59*, 1970.

(113) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.

(114) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

(115) Noyes, R. M. *J. Am. Chem. Soc.* **1962**, *84*, 513.

(116) Noyes, R. M. *J. Am. Chem. Soc.* **1964**, *86*, 971.

(117) Rossinsky, D. R. *Chem. Rev.* **1965**, *65*, 467.

(118) Miller, W. A.; Watts, D. W. *J. Am. Chem. Soc.* **1967**, *89*, 6051.

(119) Nosik, A. J. *Annu. Rev. Phys. Chem.* **1978**, *29*, 189.

(120) Farrell, J. F.; McTigue, P. *J. Electroanal. Chem.* **1982**, *139*, 37.

(121) Ress, H.; Heller, A. *J. Phys. Chem.* **1985**, *89*, 4207.

(122) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Jr., T. R. T. *J. Phys. Chem. A* **1998**, *102*, 7787.

(123) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 9734.

(124) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74. Errata: **1993**, *202*, 567.

(125) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1992**, *114*, 8794.

(126) Pu, J.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Liotard, D. A.; Rinaldi, D.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS, version 4.2; University of Minnesota: Minneapolis, MN, 2004.

(127) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

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

(129) Nakamura, H.; Xidos, J. D.; Thompson, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Lynch, B. J.; Volobuev, Y.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. HONDOPLUS, version 4.5; University of Minnesota: Minneapolis, MN, 2004.

(130) Thompson, J. D.; Lynch, B. J.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Zhu, T.; Volobuev, Y.; Dupuis, M.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. SMXGAUSS, version 1.0; University of Minnesota: Minneapolis, MN, 2004.