Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions

Aleksandr V. Marenich, Christopher J. Cramer,* and Donald G. Truhlar*

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

We present a new continuum solvation model based on the quantum mechanical charge density of a solute molecule interacting with a continuum description of the solvent. The model is called SMD, where the "D" stands for "density" to denote that the full solute electron density is used without defining partial atomic charges. "Continuum" denotes that the solvent is not represented explicitly but rather as a dielectric medium with surface tension at the solute–solvent boundary. SMD is a universal solvation model, where "universal" denotes its applicability to any charged or uncharged solute 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). The model separates the observable solvation free energy into two main components. The first component is the bulk electrostatic contribution arising from a self-consistent reaction field treatment that involves the solution of the nonhomogeneous Poisson equation for electrostatics in terms of the integral-equation-formalism polarizable continuum model (IEF-PCM). The cavities for the bulk electrostatic calculation are defined by superpositions of nuclear-centered spheres. The second component is called the cavity–dispersion-solvent-structure term and is the contribution arising from short-range interactions between the solute and solvent molecules in the first solvation shell. This contribution is a sum of terms that are proportional (with geometry-dependent proportionality constants called atomic surface tensions) to the solvent-accessible surface areas of the individual atoms of the solute. The SMD model has been parametrized with a training set of 2821 solvation data including 112 aqueous ionic solvation free energies, 220 solvation free energies for 166 ions in acetonitrile, methanol, and dimethyl sulfoxide, 2346 solvation free energies for 318 neutral solutes in 91 solvents (90 nonaqueous organic solvents and water), and 143 transfer free energies for 93 neutral solutes between water and 15 organic solvents. The elements present in the solutes are H, C, N, O, F, Si, P, S, Cl, and Br. The SMD model employs a single set of parameters (intrinsic atomic Coulomb radii and atomic surface tension coefficients) optimized over six electronic structure methods: M05-2X/MIDI!6D, M05-2X/6-31G*, M05-2X/6-31+G**, M05-2X/cc-pVTZ, B3LYP/6-31G*, and HF/6-31G*. Although the SMD model has been parametrized using the IEF-PCM protocol for bulk electrostatics, it may also be employed with other algorithms for solving the nonhomogeneous Poisson equation for continuum solvation calculations in which the solute is represented by its electron density in real space. This includes, for example, the conductor-like screening algorithm. With the 6-31G* basis set, the SMD model achieves mean unsigned errors of 0.6–1.0 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 4 kcal/mol on average for ions with either Gaussian03 or GAMESS.

1. Introduction

By invoking the continuum approximation for the solvent, the electronic structure problem for a molecule in a liquid is reduced to the size of the solute of interest. Continuum solvation models (sometimes called implicit solvation or implicit solvent models) represent a solvated molecule at an atomic level of detail inside a molecule-sized (and usually molecule-shaped) electrostatic cavity surrounded by a dielectric medium that represents the solvent. In the electrostatic theory of dielectric media, the medium is associated with a relative permittivity, which is a scalar function of position for isotropic nonhomogeneous media. In practice, the relative permittivity is usually equated to the bulk solvent static dielectric constant outside the solute cavity and to a smaller value inside the cavity. The smaller value is usually taken as unity (that is, the relative permittivity

of vacuum), which is appropriate if the solute polarization is treated explicitly. Although models that treat the solute by unpolarized molecular mechanics are sometimes used, the present article deals with models that treat the solute as polarizable and quantum mechanical. The charge distribution of the solute (charge density) induces polarization in the surrounding dielectric medium, and the self-consistently determined interaction between the solute charge distribution and the electric polarization field of the solvent, when adjusted for the energetic cost of polarizing the solute and the solvent, constitutes what is called the electrostatic (or bulk electrostatic) contribution to the free energy of solvation. The electric potential due to the polarized dielectric continuum and the polarization of the solute, called the reaction field, equals the total potential minus the electrostatic potential[1] of the gas-phase solute molecule.

The total electric potential satisfies the nonhomogeneous Poisson equation (NPE) for electrostatics. Therefore, the reaction

* Corresponding authors. E-mail: cramer@umn.edu (C.J.C.); truhlar@umn.edu (D.G.T.).

field can be obtained self-consistently by numerical integration of the NPE coupled to the quantum mechanical electron density of the solute molecule. NPE solvers that employ the continuous charge density (without approximating it by distributed point charges or multipoles) will be further called density-based solvation models. An example of such models is the polarizable continuum model (PCM).[2,3] Various PCM formulations include the integral-equation-formalism protocol (IEF-PCM),[4-7] the dielectric version of PCM (D-PCM),[2,3,8] and the conductor-like screening algorithm.[9-15] It has been shown[16] that the IEF-PCM formalism is equivalent to the SS(V)PE method of Chipman.[17] Other implicit solvation models solve the NPE using alternative representations of the continuous density, for example, single- or multicenter multipolar expansions.[18-20] An alternative continuum model, the generalized Born (GB) approximation,[21-27] does not start with the NPE, but instead employs a starting point based on Coulomb's law and represents the solute as a collection of point charges (a distributed monopole approximation), located at the nuclear positions.

Comprehensive analyses of various methods for treating solvation are beyond the scope of the present article but can be found elsewhere.[28-30] Nevertheless, it is worth adding a few qualitative remarks here. Both NPE solvers and Coulomb's law models have uncertainties (discussed below) due to the necessity to augment the estimate of the effect of bulk electrostatics with nonelectrostatic and nonbulk contributions, and they also have uncertainties in the bulk electrostatic part itself. Current methods based on the NPE have uncertainties due to the definition, size, and shape of the solute cavity, the portion of the solute charge that may lie outside the cavity, and the assumed way in which the permittivity changes at and near the solute–solvent boundary.[31] On the other hand, the GB approximation and other approximations based on Coulomb's law may have different but equally serious deficiencies. For instance, in contrast to the NPE solvers, the GB approximation simplifies the treatment of charge distributions by replacing the continuous charge density of the solute by a set of atom-centered partial charges for all stages of the calculation. This eliminates the incorrect treatment of outlying charge but at the cost of incorrectly eliminating the outlying charge itself.[31] The GB method also introduces errors due to the fact that certain types of charge distributions may be poorly described by atom-centered monopoles, for example, the local charge distributions around atoms with lone electron pairs.[32,33] The weakness of approximations based on Coulomb's law from the standpoint of the Kirkwood distributed monopole model[18] was discussed by Grycuk.[34]

Because the GB approach is based on partial atomic charges, its accuracy for a particular level of electronic structure theory may depend on whether meaningful partial charges can be computed for that level of theory, but the ability to compute such partial atomic charges is not guaranteed for all possible theory levels and basis sets. For this reason, density-based solvation models are deemed to be less sensitive to the choice of basis set, unless class IV charge mappings[35] are available to obtain reasonable partial atomic charges for a given basis set.

Continuum models remove the difficulties associated with the statistical sampling of solvent configuration space, but they do so at the cost of losing molecular-level detail. Indeed, the assumption that the electrostatic interactions of the solute and the surrounding solvent do not depend on the molecular structure of the solvent and that the dielectric response of the medium is uniform and linear at all positions outside the space that defines the solute may not provide a completely valid description of the solvent in the first solvation shell. This assumption is particularly poor when strong, specific interactions between a solute and one or more first-shell solvent molecules are present, for example strong hydrogen bonding, $\pi-\pi$ stacking interactions,[36] or monatomic ions.[37] Therefore, reliable calculations of solutes in solution must take into account not only bulk electrostatics (that is long-range electrostatic polarization effects) but also shorter-range polarization effects and shorter-range nonelectrostatic effects such as cavitation, dispersion, and solvent structural effects (CDS), the latter including both hydrogen bonding and exchange repulsion effects.[28-30] On the other hand, there is no thermodynamically unique way to separate the bulk electrostatic contribution to the free energy of solvation from the non-bulk-electrostatic one;[38] only their sum is a state function.[39,40] Therefore, continuum solvation models differ from one another in the manner in which the bulk-electrostatic term and the non-bulk-electrostatic term are defined. A partition of solvation free energy into electrostatic and nonelectrostatic contributions may be associated with a particular implicit path for thermodynamic integration, and some paths may have more predictive power for modeling than others do. The validity of a model can be judged by the usefulness of the whole model in predicting and correlating experimental observables, but not by any supposed rigor in the electrostatic or nonelectrostatic parts of the formulation.[31]

One particularly relevant issue in regard to the definition of the bulk-electrostatic and non-bulk-electrostatic contributions to the free energy of solvation 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. If we assume that the solute cavity is approximately independent of charge (see next paragraph), then 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. A key issue is that the division of solvation effects into electrostatic and nonelectrostatic contributions is not well-defined, so the "electrostatic contribution" should more properly be labeled the bulk-electrostatic contribution assumed by the particular model, and that is what we mean even when we use more concise language.

The cavity for the bulk-electrostatic calculation represents a solvated molecule surrounded by a dielectric medium that represents the solvent; it is defined by a superposition of nuclear-centered spheres. We define the electrostatic contribution to the free energy of solvation by using systematically optimized atomic radii. Such radii are sometimes called intrinsic atomic Coulomb radii. It has been our usual practice to optimize such radii in calculations on ions, then fix these parameters and optimize the nonelectrostatic terms on data for neutrals.[41] In particular, this way of proceeding has been used in the development of our most recent solvation model, SM8,[42] and it will also be used here. We should emphasize, however, that the most physical definition of a solute cavity would surely depend on more details of the solute than just its geometry and the atomic numbers of its constituent atoms, and in fact some solvation models use cavities in which the radii depend on charges and/or atomic environments or the positions of isodensity contours. However, the precise definition of a solute cavity is uncertain, and actually the very concept of a solute cavity is

an approximate one. As a consequence, in designing SM8 and the new solvation model presented here, we keep the cavity definition as simple as possible, thereby achieving a clear and systematic definition of the bulk electrostatic contribution of the particular model, and we attempt to make up for the intrinsic uncertainty in cavity definition and its variation with chemical environment and functionality by the use of semiempirical surface tensions that account for short-range nonbulk electrostatics and also account for shorter-range nonelectrostatic effects such as cavitation, dispersion, and solvent structural effects.

The SM8 model was parametrized using a large training set composed of 2842 solvation data for neutrals in 90 organic solvents and water and for single cations and anions in water, acetonitrile, dimethyl sulfoxide, and methanol. Achieving mean unsigned errors of 0.5−0.8 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 2−7 kcal/mol for ions, the SM8 model is one of the most accurate continuum solvation models presently available for predicting free energies of solvation for molecular solutes.[42,43]

The recent SM8 continuum solvation model[42,43] represents the culmination of many years of earlier SMx models with $x = 1−7$.[44−56] All SMx ($x = 1−8$) solvation models except SM5.0 and SM5C[53] are generalized Born models based on discrete partial atomic charges, whereas SM5.0 has no explicit electrostatics and the SM5C solvation model was based on the continuous electronic density in terms of the conductor-like screening model (COSMO) protocol.[9] The SM8 model employs class IV charge models, in particular, Charge Model 4 (CM4)[55] and Charge Model 4M (CM4M).[57] 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,[58] Löwdin,[59] or redistributed Löwdin[60] population analyses. This allows one to shift the focus of the modeling effort away from the description of the solute toward the various components of the solvation process.[42] In addition, CM4 and CM4M charges yield more accurate long-range electrostatic potentials than population-analysis charges, and this makes the solvation models based on such charges more physical. The charge model parameters are basis-set dependent, but the CM4 and CM4M models have so far been parametrized only for 10 basis sets.[57] SM8 can also be used with any other charge model, but the user must validate that the partial atomic charges obtained by the model are similar to those obtained by CM4 and CM4M.

The SM8 model does not require the user to assign molecular-mechanics types to an atom or group; all parameters are unique and continuous functions of geometry and are independent of total charge, local charge, and user-assigned atomic environment. The SM8 model may therefore be used straightforwardly at transition states and along reaction paths.

Another feature of SM8 is that nonelectrostatic effects due to cavity (C) formation, dispersion (D) interactions, and changes in solvent structure (S) are included in terms of empirical atomic surface tensions according to algorithms[61] previously developed. The contribution of the C, D, and S effects is labeled CDS, and the CDS contribution to the free energy of solvation is a sum of terms that are proportional (with geometry-dependent proportionality constants called atomic surface tensions) to the solvent-accessible surface areas (SASA) of the individual atoms of the solute. The CDS terms are parametrized to include all of the deviations of the electrostatics from the assumed bulk model, such as the inexactness of the solute charge model and the inexactness of the solvent permittivity model including assumed values for intrinsic Coulomb radii and uncertainties in the

treatment of solute charge outside the cavity.[31] Since the non-bulk-electrostatic effects are primarily associated with physical effects in the first solvation shell and the inexactness and uncertainties of the bulk electrostatic model in the first solvation shell, they can be modeled in terms of short-range analytic forms such as surface tensions associated with the solvent-accessible surface.[44,61] This strategy has been used in the SM8 parametrization, and it has led to more accurate models than those for which the electrostatic and nonelectrostatic terms were determined separately.[31,42]

In the present study, we extend the SM8 parametrization strategy[42] to the development of a new continuum solvation model called SMD. In contrast to SM8, which employs the GB approximation for bulk electrostatics and represents the solute molecule as a collection of partial atomic charges in a cavity, the new SMD model is based on the polarized continuous quantum mechanical charge density of the solute (the "D" in the name stands for "density"). The SMD bulk electrostatic contribution to the free energy of solvation arises from a self-consistent reaction field treatment that involves solution of the NPE by the IEF-PCM protocol.[4−7] In contrast to SM8, whose accuracy is susceptible to the possible unavailability of reliable partial atomic charges, the SMD model does not utilize a discrete representation of the solute's charge density, and, therefore, it can be used effectively in those cases where the SM8 model may not be accurate enough due to the unavailability of a charge model for a particular electronic structure method and/or basis set. The SM8 model was developed only for density functional theory or Hartree−Fock theory whereas the SMD model is more portable, and it can be used with any electronic structure method for which the PCM algorithm is available.

The SMD model employs a single set of parameters (intrinsic atomic Coulomb radii and atomic surface tension coefficients) optimized for the IEF-PCM algorithm. However, the SMD parameters may also be used with other algorithms for solving the NPE in which the solute is represented by its electron density in real space. This includes, for example, the conductor-like screening algorithm called, in various implementations, which sometimes differ in details, COSMO, generalized COSMO (GCOSMO), COSab, conductor-like PCM (CPCM or C-PCM), or CD-COSMO.[9−15] Here we use C-PCM to label the implementations called C-PCM in Gaussian03[62] and GAMESS,[63−65] and we use COSMO to label the implementation called COSMO in NWChem.[66] C-PCM and COSMO solve the nonhomogeneous Poisson equation for an infinite dielectric constant (corresponding to the solvent being a conductor rather than a dielectric) but with scaled dielectric boundary conditions in order to approximate the result for a finite dielectric constant; this is a better approximation at high dielectric constant (e.g., 80) than at low dielectric constant (e.g., below 10).

Others have also combined the IEF-PCM algorithm (or the earlier but closely related PCM model) for computation of the polarization free energy with various approaches to compute CDS-type contributions.[67−73] In addition to representation of the details of the new SMD model, we highlight algorithmic and philosophical differences between SMD and prior strategies for parametrizing IEF-PCM or PCM.

2. Description of the SMD Universal Model

The standard-state free energy of solvation, that is the standard-state free energy of transfer from the gas phase to the condensed phase, may be partitioned according to

$$\Delta G_S^\circ = \Delta G_{ENP} + G_{CDS} + \Delta G_{conc}^\circ \qquad (1)$$

The ENP subscript in eq 1 denotes the electronic (E), nuclear (N), and polarization (P) components of the free energy. The nuclear relaxation component of the ENP term is equal to the difference between the gas-phase total energy calculated at the gas-phase equilibrium structure and the *gas*-phase total energy calculated at the *liquid*-phase equilibrium structure. If the geometry is assumed to be the same in the gas phase and in the liquid phase, then the ENP term becomes just an electronic polarization (EP) term. Because all calculations reported here are based on gas-phase geometries, the nuclear component in ΔG_{ENP} is assumed to be zero in this article, although not in the model in general. The CDS subscript in eq 1 emphasizes that the corresponding term is nominally associated with the free energy change associated with solvent cavitation (C), changes in dispersion (D) energy, and possible changes in local solvent structure (S). The final term in eq 1 accounts for the concentration change between the gas-phase standard state and the liquid-phase standard state. Since here the same concentration (1 mol/L) is used in both the gaseous and solution phases, ΔG_{conc}° is zero.[74] If we instead used a gas-phase standard state of 1 atm, ΔG_{conc}° would be 1.89 kcal/mol.

The ΔG_{EP} contribution to the total solvation free energy is calculated from a molecular orbital self-consistent reaction field calculation where the reaction field is defined by solution of the NPE as discussed below.

Bulk Electrostatics Formalism. In the electrostatic theory of dielectric media (i.e., nonconducting media), the medium is associated with a relative permittivity ε, which is a scalar constant for isotropic homogeneous media and a scalar function of position for isotropic nonhomogeneous media. The charges of the medium (called bound charges) do not appear explicitly, and the charge density per unit volume of explicitly treated objects in the medium is called the free charge density ρ_f. In a linear isotropic homogeneous medium, the scalar electric potential Φ satisfies Poisson's equation

$$\varepsilon \nabla^2 \Phi = -4\pi\rho_f \qquad (2)$$

For a linear isotropic nonhomogeneous medium, eq 2 is replaced by[75]

$$\nabla \cdot (\varepsilon \nabla \Phi) = -4\pi\rho_f \qquad (3)$$

which is called the nonhomogeneous Poisson equation (NPE). In the application considered in this article, ρ_f is the solute charge density. One can solve the NPE coupled to the quantal description of the solute to obtain the electric potential due to the polarized dielectric continuum and the polarization of the solute; this potential, called the reaction field ϕ, equals the total potential Φ minus the electrostatic potential $\Phi^{(0)}$ of the gas-phase molecule.[1] From the reaction field one calculates the free energy change corresponding to the solvation process. If we approximate the solute as rigid, the electrostatic contribution to the free energy of solvation is given by[76]

$$\Delta G_{EP} = \langle \Psi | H^{(0)} - \frac{e}{2}\varphi | \Psi \rangle + \frac{e}{2}\sum_k Z_k \varphi_k - \\ \langle \Psi^{(0)} | H^{(0)} | \Psi^{(0)} \rangle \qquad (4)$$

where e is the atomic unit of charge, ϕ_k is the reaction field evaluated at atom k, Z_k is the atomic number of atom k, $H^{(0)}$ and $\psi^{(0)}$ are the solute electronic Hamiltonian and electronic wave function, respectively, in the gas phase, and Ψ is the polarized solute electronic wave function in solution. This equation includes the polarization of the solvent by the solute and the distortion of the solute that is induced by this polarization effect.

A key issue in all implicit solvation models is the boundary between the solute cavity and the solvent continuum. In the SMD model, the boundary is defined to enclose a superposition of nuclear-centered spheres with radii ρ_k, which are called intrinsic Coulomb radii. The values depend only on the atomic numbers of the atoms. This boundary forms a so-called solvent-accessible surface (SAS). In terms of the boundary element method[77] routinely used for solution of differential equations such as eq 3, the cavity can be expressed in a discrete way and, therefore, the reaction field at an arbitrary position **r** within the cavity can be approximated by

$$\phi(\mathbf{r}) = \sum_m \frac{q_m}{|\mathbf{r} - \mathbf{r}_m|} \qquad (5)$$

where $\mathbf{r}_m$ is the position of an element m of surface area on the solute−solvent boundary (such elements are called tessarae), and q_m is the apparent surface charge on element m. The discretization (tessellation) of the boundary surface area can be achieved in many different ways, for instance, by using a surface-building method called GEPOL,[78] modified versions[79−81] of which have been implemented in Gaussian03[62] and GAMESS.[63−65]

The SMD bulk electrostatic solvation energies ΔG_{EP} used for the SMD parametrization were calculated using Gaussian03[62] and the IEF-PCM algorithm[4−7] as implemented in Gaussian03. In such calculations, the nonelectrostatic contributions produced by IEF-PCM/Gaussian03 by default were turned off by using the keywords *nodis*, *norep*, and *nocav* in the corresponding Gaussian03 input file. The solvent radius (or probe radius) was set to zero by using the *rsolv=0* option. The solvent was specified only by the value of ε, that is, by its dielectric constant (bulk relative permittivity). We also used the *scfvac* keyword to invoke a calculation of the gas-phase total energy prior to a solvation calculation in order to compute ΔG_{EP} according to eq 4. By default IEF-PCM/Gaussian03 uses united-atom schemes[67] for assigning atomic or group radii, but the default radii were overridden by using the *modifysph* option and the corresponding SMD values for atomic radii (ρ_k). The options *surface=sas* and *noaddsph* were used to generate the solvent-accessible surface of the solute molecule defined only by the superposition of nuclear-centered spheres. Thus, the SAS was defined to precisely enclose these spheres and nothing else. In other words, we keep the cavity definition as simple as possible, thereby achieving a clear and systematic definition of the bulk electrostatic contribution of the particular model. All the NPE Gaussian03 calculations were carried out using the default tessellation algorithm (that is, GEPOL[78−81]) and the default value for *tsare* equal to 0.2 Å^2. According to the Gaussian03 manual, the option *tsare* specifies the average area of the surface elements (tesserae) generated on each atomic sphere in the SAS. The number of the tesserae is determined by the program based on the value of *tsare* and the radius of the sphere.

Cavity-Dispersion-Solvent-Structure Formalism. The G_{CDS} contribution to the free energy of solvation in eq 1 is given by

6382 *J. Phys. Chem. B, Vol. 113, No. 18, 2009*

Marenich et al.

$$G_{CDS} = \sum_{k}^{atoms} \sigma_k A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) +$$
$$\sigma^{[M]} \sum_{k}^{atoms} A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) \quad (6)$$

where σ_k and $\sigma^{[M]}$ are the atomic surface tension of atom k and the molecular surface tension, respectively, and A_k is the solvent-accessible surface area (SASA)[82,83] 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.[61] Notice that the van der Waals radii used in the SASA calculation are not the same as the intrinsic Coulomb radii used for solution of the bulk electrostatic problem; in fact we use Bondi's values[84] for the van der Waals radii plus a solvent radius that is $r_s = 0.4$ Å in the SMD model. In cases where the atomic radius is not given in Bondi's paper[84] we generally recommend that a radius of 2.0 Å be used, although if Bondi-like radii are available from other sources, they may also be used.

The atomic surface tensions are given by

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{k'}^{atoms} \tilde{\sigma}_{Z_k Z_{k'}} T_k(\{Z_{k'}, R_{kk'}\}) \quad (7)$$

where $\tilde{\sigma}_Z$ is an atomic-number-specific parameter, $\tilde{\sigma}_{ZZ'}$ is a parameter that depends on the atomic numbers of atoms k and k', and $T_k(\{Z_{k'}, R_{k,k'}\})$ is a geometry-dependent switching function called a cutoff tanh, or COT; this function is described in the Supporting Information. SMD uses the same functional forms T_k as does SM8.[42]

As in previous SM*x* universal solvation models, in SMD 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), α is Abraham's[85-88] hydrogen bond acidity parameter of the solvent (which Abraham denotes as $\sum \alpha_2$), β is Abraham's hydrogen bond basicity parameter of the solvent (which Abraham denotes as $\sum \beta_2$), and $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, and $\tilde{\sigma}_i^{[\beta]}$ are empirical parameters that depend on i. SMD also uses a molecular surface tension that is multiplied by the total SASA of the given solute (see eq 6; 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

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]}(\gamma / \gamma_0) + \tilde{\sigma}^{[\phi^2]}\phi^2 + \tilde{\sigma}^{[\psi^2]}\psi^2 + \tilde{\sigma}^{[\beta^2]}\beta^2 \quad (9)$$

where γ is the macroscopic surface tension of the solvent at air/solvent interface at 298.15 K; we express surface tension in units of cal mol^{-1} Å^{-2} (note that 1 dyn/cm = 1.439 32 cal mol^{-1} Å^{-2}), and $\gamma_0 = 1$ cal mol^{-1} Å^{-2}, ϕ^2 is the square of the fraction of solvent atoms that are aromatic carbon atoms (carbon

aromaticity), ψ^2 is the square of the fraction of solvent atoms that are F, Cl, or Br (electronegative halogenicity), β^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 SMD model may be applied to any solvent (or medium) for which the relevant macroscopic descriptors are either known or may be estimated. We note, however, that water is treated as a special solvent that is given its own set of surface tension coefficients, 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 SMD is used to compute solvation free energies in aqueous solvent, eq 6 reduces to its first term. In the case of water, the parameters $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ used in eq 7 to obtain σ_k are simply numbers that do not depend on solvent descriptors.

3. SMD Training Set

The SMD training set is essentially the same as the SM8 training set[42,89] used in the parametrization of SM8 except for two modifications that are described at the end of this section. The training set contains charged and uncharged solutes composed of H, C, N, O, F, Si, P, S, Cl, or Br. All standard-state 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.

The SMD training set contains 274 and 2072 experimental solvation free energies for neutrals in water and 90 nonaqueous solvents, respectively, and the total number of unique solutes is 318. In addition to the 2346 absolute free energies, the training set contains 143 transfer free energies for an additional 93 neutral solutes between water and 15 organic solvents. The transfer free energies were determined directly from experimental partition coefficients according to

$$\Delta G_{o/w}^{\circ} = -2.303 RT \log P_{o/w} \quad (10)$$

where $\Delta G_{o/w}^{\circ}$ 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

$$P_{o/w} = \frac{[solute]_o}{[solute]_w} \quad (11)$$

where $[solute]_o$ is the equilibrium concentration of the solute in the organic phase and $[solute]_w$ is the equilibrium concentration of the solute in the aqueous phase. Transfer free energy data are included in this training set 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.

In addition to 2489 experimental absolute and transfer free energies for neutrals in water and 90 organic solvents, the current training set contains 112 aqueous ionic solvation free energies and 220 solvation free energies for 166 ions in acetonitrile, dimethyl sulfoxide (DMSO), and methanol. Thus, the total number of solvation data in the SMD training set amounts to 2821. The single-ion solvation free energies were evaluated in previous work[55,56,90] based on the corresponding thermochemical cycle. For instance, the absolute solvation free energy of the cation BH$^+$ can be written as

New Continuum Solvation Model

J. Phys. Chem. B, Vol. 113, No. 18, 2009 **6383**

$$\Delta G_S^\circ(BH^+) = \Delta G_g^\circ(BH^+) + \Delta G_S^\circ(B) -$$
$$2.303RT\, pK_a(BH^+) + \Delta G_S^\circ(H^+) + \Delta G^{1atm \rightarrow 1M} \quad (12)$$

where $\Delta G_g^\circ(BH^+)$ is the gas-phase acidity of BH^+, which is equal to

$$\Delta G_g^\circ(BH^+) = G^\circ(B) + G^\circ(H^+) - G^\circ(BH^+) \quad (13)$$

In eq 12, $\Delta G_S^\circ(B)$ is the solvation free energy of the neutral species B, pK_a is the negative common logarithm of the solution-phase acid dissociation constant of BH^+, and $\Delta G^{1atm \rightarrow 1M} = 1.89$ kcal/mol 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, the absolute solvation free energy of the anion A^- can be written as

$$\Delta G_S^\circ(A^-) = -\Delta G_g^\circ(AH) + \Delta G_S^\circ(AH) +$$
$$2.303RT\, pK_a(AH) - \Delta G_S^\circ(H^+) - \Delta G^{1atm \rightarrow 1M} \quad (14)$$

In the above equations, we use the value of −265.9 kcal/mol for the absolute aqueous solvation free energy of the proton that corresponds to the value of −264.0 kcal/mol reported[91] by Tissandier et al. (the difference is attributed to the 1 M standard state correction, −1.89 kcal/mol). For the free energies of solvation for the proton in acetonitrile, DMSO, and methanol, we use −260.2, −273.3, and −263.5 kcal/mol, respectively. These solvation energies were evaluated in previous work based on the cluster pair approximation.[90] They are different from those recently reported by Fawcett:[92] −256.2, −270.5, and −263.8 kcal/mol for acetonitrile, DMSO, and methanol, respectively (with the 1 M standard state correction, −1.89 kcal/mol, applied to the originally reported values). This difference merits further investigation and is therefore considered further in section 6.

For aqueous ions, 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 criteria for whether to cluster an ion are as follows. First, we cluster any ionic solute in our set containing three or fewer atoms. Second, we cluster any ionic solute with one or more oxygen atoms bearing a partial atomic charge more negative than the charge on oxygen in water. Finally, we cluster any ammonium or oxonium cation. The rationale for these criteria is explained elsewhere.[55,56] In all cases, the clusters were represented by a single, lowest-energy conformation. The aqueous solvation free energies of these 31 solute−water clusters ($H_2O \cdot M^\pm$) were determined according to

$$\Delta G_S^\circ(H_2O \cdot M^\pm) = \Delta G_S^\circ(H_2O) + \Delta G_S^\circ(M^\pm) -$$
$$\Delta G_g^\circ(B.E.) + \Delta G^{1atm \rightarrow 1M} + RT \ln 55.34 \quad (15)$$

The above equation uses the aqueous solvation energies of the unclustered ions $\Delta G_S^\circ(M^\pm)$ and the gas-phase binding energies of the solute−water clusters given in previous work.[55,56] It has been reported recently[93] that the previously used equation (eq 8 in ref 55) misses the concentration correction term that is the last term in eq 15 in the present article. Thus the aqueous solvation energies of the clustered ions used in the present study have been corrected by 2.38 kcal/mol to account for this error.

TABLE 1: The 90 Nonaqueous Solvents in the SMD Neutral Training Set[a]

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

[a] Methanol is not included in this training set because there are no data on neutral solutes in methanol. The asterisk denotes the nonaqueous solvents presently available with the default solvation model implemented in Gaussian03 in addition to methanol and water. The names of 15 solvents for which we used solvent−water transfer free energies are italicized.

The SMD training set does not contain any data for ions clustered in nonaqueous solvents whereas the SM8 training set includes 21 additional data for ionic clusters in acetonitrile and DMSO.[42]

Table 1 shows the 90 organic solvents used in the SMD parametrization. Experimental values for the 2821 solvation free energies (including 2346 free energies and 143 transfer free energies for neutrals and 332 ionic free energies) are given in the Supporting Information.

4. Parametrization

As for the earlier SM8 model,[42] the SMD parametrization effort is focused on two types of parameters: (1) the atomic radii used for construction of the cavities for the bulk electrostatic calculation; (2) the parameters $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ}$ in eq 7 and the parameters $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, $\tilde{\sigma}^{[\beta^2]}$ in eq 9.

The first step is to parametrize the intrinsic Coulomb radii. It is widely recognized that the electrostatic contributions to free energies of solvation depend strongly on the model radii.[8,49,94−97] We considered a number of schemes for the optimization of Coulomb radii. This includes an optimization of a single set of radii for both water and organic solvents and possible optimizations of the radii as functions of certain solvent descriptors such as hydrogen-bond acidities α and hydrogen-bond basicities β.[85−88] The values of these solvent descriptors for the four solvents including water in which we have ionic data are listed in Table 2. The physical interpretation of the radii depending on solvent descriptors would be that first solvation shell effects depend on the solvent, and changing the

TABLE 2: Solvent Acidity and Basicity Descriptors for the Four Solvents with Ionic Data[a]

solvent	α	β
acetonitrile	0.07	0.32
DMSO	0	0.88
methanol	0.43	0.47
water	0.82	0.35

[a] α is Abraham's hydrogen bond acidity parameter (which Abraham denotes as $\sum\alpha_2$), and β is Abraham's hydrogen bond basicity parameter (which Abraham denotes as $\sum\beta_2$).

radius by an amount no larger than the width of the first solvation shell can be justified as alternative treatment of that shell.

It was our usual practice in the past to optimize the radii in calculations only on ions, then fix these parameters and optimize the non-bulk-electrostatic terms on data for neutrals.[42] For the SMD parametrization, however, we opted to include both ions and neutrals in the radii optimization followed by an optimization of the nonelectrostatic terms only on neutrals. After considerable trial and error, we adopted the following procedure to optimize the SMD radii: first, we optimized the radii for the whole training set (2821 data) by optimizing a scaling factor for the aqueous SM8 values of the radii (the optimal factor was found to be 1.176); fixing the other radii at the new values, we then reoptimized the oxygen radius using the following scheme:

$$\rho = \begin{cases} 1.52 & \alpha \geq 0.43 \\ 1.52 + 1.8(0.43 - \alpha) & \alpha < 0.43 \end{cases} \quad (16)$$

where α is the solvent's value of Abraham's hydrogen-bond acidity parameter (the parameter that he calls $\sum\alpha_2$).[85−88] The optimization was done by minimizing the root mean squared error calculated over 2821 data points. The ΔG_{EP} values used in the SMD optimization of radii were calculated using the IEF-PCM protocol for bulk electrostatics as implemented in Gaussian03,[62] the M05-2X density functional,[98] and the 6-31G* basis set.[99] Table 3 shows the SMD values of intrinsic Coulomb radii given for water ($\alpha = 0.82$) and DMSO ($\alpha = 0$). The SMD set of Coulomb radii includes the radii for H, C, N, O, F, Si, P, S, Cl, and Br. For any other atom the SMD model uses the van der Waals radius of Bondi for those atoms for which Bondi defined radii;[84] in cases where the atomic radius is not given in Bondi's paper,[84] a radius of 2.0 Å is used.

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 on ions and neutrals with SM8 atomic surface tensions,[42] then determined the SMD surface tensions using only neutral data. Employing this procedure, we also tested other sets of intrinsic Coulomb radii as an initial guess for the SMD radii by scaling them and optimizing the scaling factor and found no advantage in using these radii for the SMD parametrization. The tested sets of the radii, as given in Table 3, include Bondi's van der Waals radii,[84] the radii used by Jaguar's Poisson−Boltzmann self-consistent reaction field solver,[100] the radii reported[101] by Klamt and Eckert (K&E), and the radii reported[102] by Stefanovich and Truong (S&T). We used Bondi's values[84] for those atoms for which the other radii were not available.

To begin the parametrization of the atomic surface tensions, ΔG_{EP} values were calculated for all of the neutral solutes in the SMD training set for which solvation free energies are available (total of 2346 calculations). The ΔG_{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 also calculated (a total of 286 calculations). The ΔG_{EP} values used in the SMD optimization of surface tensions were calculated using IEF-PCM/Gaussian03 and six electronic structure methods, namely, M05-2X[98]/MIDI!6D,[103,104] M05-2X/6-31G*,[99] M05-2X/6-31+G**,[99] M05-2X/cc-pVTZ,[105] B3LYP[106−108]/6-31G*, and HF/6-31G*.

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

$$\chi = \sum_{J=1}^{2489} \left[\Delta G_S^{\circ}(\text{expt},J) - \frac{1}{6}\sum_{j=1}^{6} \Delta G_{EP}(j,J) - G_{CDS}(J) \right]^2 \quad (17)$$

where the j-summation is over six levels of electronic structure theory (in particular, the B3LYP density functional and the Hartree−Fock method, both with the 6-31G* basis set, and the M05-2X density functional with four basis sets: MIDI!6D, 6-31G*, 6-31+G**, and cc-pVTZ). The J-summation in eq 17 is over all data points in the neutral training set (2346 solvation free energies plus 143 transfer free energies), and $\Delta G_S^{\circ}(\text{expt},J)$ is the experimental standard-state solvation or transfer free energy. For solvation free energies, $\Delta G_{EP}(j,J)$ and $\Delta G_{CDS}(J)$ can be calculated directly by the solvation model, whereas for transfer free energies, two separate solvation model calculations are required; that is

$$\Delta G_{EP,\text{transfer}} = \Delta G_{EP,\text{organic}} - \Delta G_{EP,\text{water}} \quad (18)$$

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

where $\Delta G_{EP,\text{organic}}$ and $\Delta G_{EP,\text{water}}$ are calculated in the same way, except that different values are used for dielectric constant, and $G_{CDS,\text{organic}}$ and $G_{CDS,\text{water}}$ are computed using eq 6. Note that because transfer free energies depend on both the aqueous solvation free energy and the solvation free energy in the organic solvent (eq 10), 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 $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ parameters were optimized against data for molecules containing H, C, N, and/or O. Next, 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 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. SMD uses the same functional forms for the atomic surface tensions as SM8, which contains 26 different $\tilde{\sigma}_i^{[\text{water}]}$ values, plus 78 $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, $\tilde{\sigma}_i^{[\beta]}$ parameters, and the four $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ parameters (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. After an optimization of *all* 108 parameters using the three-step procedure outlined above we removed any parameter with a value greater than 1000 cal mol^{-1} Å^{-2}, then the remaining parameters were reoptimized (in addition to being very large

New Continuum Solvation Model

J. Phys. Chem. B, Vol. 113, No. 18, 2009 **6385**

TABLE 3: Intrinsic Coulomb Radii (Å) of Various Models and Bondi's van der Waals Radii (Å)

atom	Z	SMD(aq)[a]	SMD(DMSO)[a]	SM8(aq)[b]	SM8(DMSO)[b]	Bondi[c]	Jaguar[d]	K&E[e]	S&T[f]
H	1	1.20	1.20	1.02	0.80	1.20	1.15	1.30	1.17
C	6	1.85	1.85	1.57	1.57	1.70	1.90	2.00	1.87
N	7	1.89	1.89	1.61	1.61	1.55	1.60	1.83	1.93
O	8	1.52	2.29	1.52	2.18	1.52	1.60	1.72	1.58
F	9	1.73	1.73	1.47	2.63	1.47	1.68	1.72	1.28
Si	14	2.47	2.47	2.10	2.10	2.10	2.15		
P	15	2.12	2.12	1.80	2.13	1.80	2.07		2.28
S	16	2.49	2.49	2.12	2.45	1.80	1.90	2.16	2.02
Cl	17	2.38	2.38	2.02	2.63	1.75	1.97	2.05	1.75
Br	35	3.06	3.06	2.60	2.85	1.85	2.10	2.16	

[a] The SMD radius of the oxygen atom in the solute molecule is defined as a function of Abraham's hydrogen-bond acidity parameter (α) for a given solvent according to eq 16. See the text for more detail. [b] The intrinsic Coulomb radii used by the SM8 model for any solute in water and DMSO (ref 42). [c] Bondi's values of van der Waals radii (ref 84). [d] The default values of radii used in Jaguar's Poisson−Boltzmann self-consistent reaction field solver (ref 100). [e] The intrinsic Coulomb radii of Klamt and Eckert (K&E) (ref 101). [f] The intrinsic Coulomb radii of Stefanovich and Truong (S&T) (ref 102); the carbon and nitrogen radii are given here as averaged over the C_σ and $C_{\sigma\pi}$ values for carbon and over the N_{hb} and C_{nhb} values for nitrogen given in ref 102.

TABLE 4: Atomic Surface Tension Parameters (in cal mol^{-1} Å^{-2}) for SMD That Depend on Atomic Numbers[a]

i	$\tilde{\sigma}_i^{[water]}$	$\tilde{\sigma}_i^{[n]}$	$\tilde{\sigma}_i^{[\alpha]}$	$\tilde{\sigma}_i^{[\beta]}$
H	48.69			
C	129.74	58.10	48.10	32.87
H, C	−60.77	−36.37		
C, C	−72.95	−62.05		
O		−17.56	193.06	−43.79
H, O		−19.39		
O, C	68.69	−15.70	95.99	
O, O				−128.16
N		32.62		
C, N		−99.76	152.20	
N, C	−48.22		−41.00	
N, C(3)	84.10			
O, N	121.98			79.13
F	38.18			
Cl	9.82	−24.31		
Br	−8.72	−35.42		
S	−9.10	−33.17		
O, P	68.85			
Si		−18.04		

[a] Blank entries denote zero. We note that eqs 7 and 8 would allow for a very large number of parameters if all possible surface tension parameters in eqs 7 and 8 were included; however, we do not include them all. Any possible surface tension parameter that is not in this table is set equal to zero in SMD. For example, there is no surface tension on P atoms in SMD.

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%. In all, only 37 of the original 108 parameters were retained in the SMD model, compared to the 54 parameters that are used by SM8. The final set of parameters obtained using the procedure described above is listed in Tables 4 and 5. Note that the partition of solvation free energy into electrostatic and nonelectrostatic contributions used in SM8 and SMD is model-dependent. Since the SMD electrostatics and the SM8 electrostatics differ, the SMD nonelectrostatic (CDS) contribution as well as individual surface tension coefficients in SMD cannot be unambiguously compared to those in SM8.

5. SMD Performance for Neutrals

Table 6 gives a breakdown of the mean unsigned errors (MUE) in calculated aqueous solvation free energies by solute

TABLE 5: Atomic Surface Tension Parameters (in cal mol^{-1} Å^{-2}) for SMD That Do Not Depend on Atomic Numbers

$\tilde{\sigma}^{[\gamma]}$	0.35
$\tilde{\sigma}^{[\phi^2]}$	−4.19
$\tilde{\sigma}^{[\psi^2]}$	−6.68
$\tilde{\sigma}^{[\beta^2]}$	0.00

class. In Tables 7 and 8, the mean unsigned errors are broken down by solute class for calculated solvation free energies in nonaqueous solvents and for calculated transfer free energies, respectively. The corresponding mean signed errors (MSE) are given in the Supporting Information. The errors broken down by solvent name are provided in the Supporting Information as well. The SMD solvation free energies are compared to the SM8 solvation energies obtained in previous work[42] and in the present study using a locally modified version of Gaussian03 called MN-GSM.[109]

For neutral data in water, the mean unsigned error on average varies from 0.59 to 0.94 kcal/mol, the MSE is between −0.71 and 0.61 kcal/mol, and the root mean squared error (RMSE) is between 0.86 and 1.24 kcal/mol. The most accurate method among those which have been tested in this study is M05-2X/6-31G* with −0.05, 0.59, and 0.86 kcal/mol for MSE, MUE, and RMSE, respectively. The Hartree−Fock method used within the SMD protocol noticeably overpolarizes the tested aqueous neutral solutes, overestimating the corresponding hydration free energies by 0.69 kcal/mol on average (with MUE = 0.91 and RMSE = 1.24 kcal/mol). On the other hand, the SMD model used with B3LYP/6-31G* underestimates the solvation of aqueous neutrals by 0.61 kcal/mol on average (with MUE = 0.80 and RMSE = 1.08 kcal/mol). For neutral data in 90 organic solvents, the mean unsigned error on average varies from 0.64 to 0.79 kcal/mol, the MSE is in the range from −0.50 to 0.34 kcal/mol, and the RMSE is in the range from 0.86 to 1.05 kcal/mol. In both cases (water and organic solvents) the SMD model employed with a basis set containing diffuse functions (6-31+G**) tends to produce somewhat overly negative solvation energies.

The overall performance of the SMD model for neutral compounds in water (MUE = 0.59−0.94 kcal/mol, Table 6) is slightly worse in comparison with the overall performance of the earlier SM8 model, which produces MUEs of 0.55−0.67 kcal/mol averaged over the 274 aqueous data.[42] The SMD mean unsigned errors for neutral compounds in nonaqueous solvents, averaged over 2072 data (0.64−0.79 kcal/mol), are larger than those for SM8 (0.57−0.62 kcal/mol)[42] as well. For those classes

TABLE 6: Mean Unsigned Errors (kcal/mol) in Aqueous Solvation Free Energies Calculated Using SMD, by Solute Class[a]

solute class	N	M05-2X				B3LYP	HF
		MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*
H_2, NH_3, H_2O, $(H_2O)_2$	4	1.05	1.13	1.46	1.01	0.76	1.38
unbranched alkanes	8	0.11	0.15	0.28	0.35	0.08	0.06
branched alkanes	5	0.09	0.14	0.33	0.42	0.11	0.07
cycloalkanes	5	0.32	0.32	0.32	0.30	0.40	0.37
alkenes	9	0.14	0.23	0.51	0.43	0.21	0.16
alkynes	5	0.14	0.48	0.74	0.80	0.13	0.64
arenes	8	0.20	0.29	0.62	0.36	0.95	0.27
alcohols	12	0.97	0.17	0.91	0.18	0.59	0.77
phenols	4	0.51	0.48	0.95	0.36	0.86	0.87
ethers	12	0.85	0.69	0.81	0.66	1.03	0.64
aldehydes	6	1.31	0.17	1.20	0.81	0.83	1.77
ketones	12	1.28	0.19	1.32	0.96	0.52	1.79
carboxylic acids	5	1.97	0.55	0.25	0.34	1.32	1.06
esters	13	0.82	0.26	1.13	0.67	0.50	1.71
peroxides	3	0.27	1.26	1.78	0.74	0.55	1.98
bifunctional H, C, O compds	5	1.26	0.40	1.11	0.41	0.96	1.43
aliphatic amines	15	0.83	0.43	0.83	0.50	0.75	0.37
anilines	7	0.27	0.38	0.84	0.25	0.94	0.28
aromatic N-heterocycles (1 N)	10	1.04	0.53	0.40	0.26	1.35	0.37
aromatic N-heterocycles (2 Ns)	3	0.66	0.60	1.86	1.17	0.33	1.83
nitriles	4	1.68	0.64	0.17	0.19	0.95	0.25
hydrazines	3	0.85	0.91	1.66	0.75	0.64	0.97
bifunctional H, C, N compds	3	0.58	0.75	1.79	0.35	0.42	1.35
amides	4	3.08	1.36	0.52	0.88	2.42	0.53
ureas	2	2.48	0.86	0.79	0.92	2.25	0.51
nitrohydrocarbons	7	1.34	0.58	0.95	0.72	1.11	1.60
bifunctional H, C, N, O compds	3	0.43	0.48	1.87	0.50	0.41	1.15
fluoroalkanes	5	0.55	0.51	0.71	0.55	0.54	0.48
fluoroarenes	1	0.20	0.19	0.85	0.43	0.67	0.46
chloroalkanes	13	1.08	0.93	1.26	0.94	0.74	1.24
chloroalkenes	6	0.64	0.70	0.68	0.76	0.81	0.60
chloroarenes	8	0.63	0.61	1.02	0.52	0.55	0.91
bromoalkanes	9	0.09	0.53	0.95	0.74	0.23	0.66
bromoalkenes	1	0.16	0.78	1.35	0.97	0.26	0.97
bromoarenes	4	0.38	0.59	1.05	0.69	0.39	0.68
multihalogen hydrocarbons	12	0.68	0.74	1.03	0.83	0.65	0.99
halogenated bifunctional compounds	9	1.79	1.41	1.64	1.48	1.53	1.52
thiols	4	0.38	0.22	0.53	0.23	0.22	0.26
sulfides	5	0.88	0.82	0.67	0.65	1.09	0.83
disulfides	2	0.14	0.11	0.26	0.21	0.52	0.12
sulfur heterocycles	1	0.04	0.01	0.26	0.04	0.66	0.22
halogenated sulfur compds	2	0.88	0.76	1.01	1.03	1.01	1.32
phosphorus compds	14	2.05	1.64	1.56	1.89	2.00	1.76
silicon compds	1	0.90	0.71	0.45	0.42	1.04	0.75
all neutral data	274	0.88	0.59	0.94	0.68	0.80	0.91

[a] All the solvation free energies were obtained using the SMD model parameters and the IEF-PCM/Gaussian03 protocol for bulk electrostatics. *N* is the number of data in a given solute class.

of neutral compounds for which the SMD model underperforms SM8, the corresponding bulk-electrostatic contributions to the free energy of solvation produced by SMD are usually more negative than those produced by SM8, and the CDS formalism we use for both models cannot fully make up for this feature of the SMD electrostatics.

6. SMD Performance for Ions in Acetonitrile, DMSO, Methanol, and Water

Table 9 shows the mean signed errors in the SMD solvation energies of ions in the four solvents. Table 10 shows the corresponding mean unsigned errors.

Analysis of the errors on the ionic data indicates generally good agreement of the model predictions with reference data across both solute classes and electronic structure levels. The mean unsigned error averaged over all ions varies from 2.1 to 5.9 kcal/mol, the MSE is between −3.1 and 4.7 kcal/mol, and

the RMSE is between 2.8 and 8.1 kcal/mol. The mean signed errors for both anions and cations in water are positive within any method. The MSEs for ions in methanol can be positive or negative, depending on the method. The MSEs for cations in acetonitrile and DMSO are positive, and for anions they are negative for all tested methods. The errors for the sulfur-containing cations in acetonitrile are very large, but this discrepancy is mainly attributed to the H_3S^+ cation which is systematically (by ∼30 kcal/mol) undersolvated by the SMD model with respect to the experimental solvation free energy ΔG_S°(expt) = −100.2 kcal/mol.

The SMD errors for ions are substantially the same as the SM8 errors except for the errors in the SMD solvation free energies for cations in DMSO which are undersolvated on average by 8−9 kcal/mol. However, the SMD model outperforms SM8 by 3 kcal/mol on anions in the same solvent (Table 10).

New Continuum Solvation Model

J. Phys. Chem. B, Vol. 113, No. 18, 2009 **6387**

TABLE 7: Mean Unsigned Errors (kcal/mol) in Nonaqueous Solvation Free Energies Calculated Using SMD, by Solute Class[a]

		M05-2X				B3LYP	HF
solute class	*N*	MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*
H_2, NH_3, H_2O	29	0.92	1.33	1.67	1.10	1.14	1.45
unbranched alkanes	85	0.46	0.44	0.39	0.39	0.55	0.52
branched alkanes	7	0.46	0.43	0.43	0.43	0.50	0.46
cycloalkanes	13	0.43	0.43	0.44	0.42	0.45	0.44
alkenes	18	0.27	0.29	0.37	0.34	0.30	0.27
alkynes	9	0.34	0.55	0.70	0.72	0.26	0.64
arenes	134	0.65	0.70	0.83	0.72	0.50	0.67
alcohols	272	0.49	0.46	0.64	0.45	0.45	0.52
phenols	109	0.72	0.65	0.69	0.68	0.84	0.63
ethers	87	0.65	0.58	0.85	0.62	0.69	0.67
aldehydes	32	0.87	0.73	1.02	0.89	0.79	1.00
ketones	195	0.35	0.64	1.18	1.01	0.46	1.04
carboxylic acids	120	1.19	0.58	0.55	0.49	0.83	0.61
esters, including lactones[b]	243	0.69	0.51	0.58	0.53	0.59	0.55
peroxides	17	0.89	0.82	0.90	0.78	0.84	0.92
bifunctional H, C, O compds	24	1.46	1.09	1.13	0.98	1.43	0.96
aliphatic amines	154	0.54	0.49	0.66	0.46	0.49	0.52
anilines	61	0.62	0.66	0.72	0.64	0.75	0.67
aromatic N-heterocycles (1 N)	52	0.82	0.71	0.67	0.66	0.91	0.69
aromatic N-heterocycles (2 Ns)	9	1.22	1.03	0.69	0.91	1.31	0.93
nitriles	20	0.38	0.59	0.96	0.93	0.50	1.02
hydrazines	5	1.19	1.37	1.74	1.24	1.21	1.40
bifunctional H, C, N compds	2	1.45	1.51	2.50	1.73	1.43	2.39
amides	26	1.47	0.73	0.83	0.72	1.20	0.75
ureas	7	2.09	1.07	0.59	0.96	1.85	0.74
lactams	4	1.94	0.92	0.91	0.71	1.39	0.95
nitrohydrocarbons	86	0.77	0.62	0.93	0.74	0.69	1.11
bifunctional H, C, N, O compds	3	1.35	1.28	1.19	1.21	1.40	1.26
fluoroalkanes	5	0.64	0.71	1.06	0.85	0.65	0.87
fluoroarenes	11	0.82	0.93	1.20	1.02	0.72	1.00
chloroalkanes	26	1.06	0.92	1.09	0.86	0.81	1.11
chloroalkenes	15	0.85	0.93	0.89	0.95	0.97	0.83
chloroarenes	31	0.50	0.51	0.61	0.50	0.50	0.56
bromoalkanes	21	0.52	0.76	0.97	0.86	0.61	0.81
bromoalkenes	2	0.44	0.83	1.17	0.94	0.50	0.94
bromoarenes	16	0.54	0.66	0.81	0.68	0.42	0.69
multihalogen hydrocarbons	14	0.48	0.47	0.49	0.47	0.46	0.50
halogenated bifunctional compds	37	1.11	0.92	0.92	0.97	1.14	0.88
thiols	10	0.55	0.46	0.64	0.40	0.36	0.45
sulfides	13	1.14	1.08	1.16	1.00	1.05	1.06
disulfides	4	1.41	1.41	1.41	1.39	1.40	1.39
sulfur heterocycles	4	0.53	0.57	0.46	0.60	0.94	0.47
sulfoxides	1	1.53	0.27	1.56	0.36	1.07	0.92
phosphorus compds	37	1.85	1.72	1.73	1.77	1.83	1.79
silicon compds	2	1.95	1.83	1.66	1.65	2.03	1.86
all neutral data	2072	0.70	0.64	0.79	0.67	0.67	0.73

[a] All the solvation free energies were obtained using the SMD model parameters and the IEF-PCM/Gaussian03 protocol for bulk electrostatics. *N* is the number of data in a given solute class. [b] 5 lactones and 238 other esters.

Fawcett[92] has presented alternative values of the absolute proton solvation energy in the three organic solvents, and using these values changes the reference values for both cations and anions. Table 11 presents the mean unsigned errors in the free energies of solvation for ions in acetonitrile, DMSO, and methanol using the ionic reference data evaluated by eqs 12 and 14 with use of the free solvation energies of the proton in the corresponding solvents reported by Fawcett:[92] −256.2, −270.5, and −263.8 kcal/mol for acetonitrile, DMSO, and methanol, respectively (the 1 M standard state correction, −1.89 kcal/mol, was applied to the originally reported values). The values for the free solvation energies of the proton used in the present parametrization and in Tables 9 and 10 are −260.2, −273.3, and −263.5 kcal/mol for acetonitrile, DMSO, and methanol, respectively. They were evaluated in previous work using the cluster pair approximation.[55,56,90] These values, compared to those reported by Fawcett,[92] are more negative for

the proton in acetonitrile and in DMSO and slightly more positive for the proton in methanol. The overall performances of all methods in Table 11 are even better when we use Fawcett's values of the free energy of solvation for the proton in the three organic solvents in spite of the fact that these values have not been used for the SMD parametrization. This observation is valid for the SM8 data as well (Table 11).

7. SMD Performance with Other Electrostatic Algorithms

So far we discussed the performance of the SMD model in solvation calculations using the SMD model parameters (atomic Coulomb radii and atomic surface tensions) along with the IEF-PCM algorithm[4−7] for bulk electrostatics as implemented in Gaussian03[62] (IEF-PCM/G03). Although the SMD model has been parametrized for the use with IEF-PCM/G03, we have also

TABLE 8: Mean Unsigned Errors (kcal/mol) in Transfer Free Energies between Water and Organic Solvents Calculated Using SMD, by Solute Class[a]

		M05-2X				B3LYP	HF
solute class	N	MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*
lactones	10	0.79	0.74	1.05	0.86	0.76	1.82
aromatic N-heterocycles	6	0.55	0.67	0.88	0.72	0.70	0.80
bifunctional H, C, N compds	2	0.69	0.84	0.98	0.87	0.71	0.92
amides	13	1.22	0.64	0.39	0.50	1.19	0.48
ureas	11	1.09	0.65	0.48	0.65	1.02	0.43
lactams	4	1.54	1.31	1.20	1.25	1.49	0.93
thymines and uracils	12	0.97	0.65	0.83	0.70	0.77	1.59
bifunctional H, C, N, O compds	5	0.38	0.62	0.79	0.63	0.44	1.07
halogenated bifunctional compds	39	0.90	0.56	0.63	0.55	0.68	1.01
sulfur compds	19	0.51	0.50	0.49	0.49	0.48	0.53
phosphorus compds	9	0.52	0.88	1.11	0.91	0.55	1.95
silicon compds	13	0.62	0.58	0.55	0.57	0.63	0.57
all neutral data	143	0.82	0.64	0.69	0.64	0.74	0.97

[a] All the solvation free energies were obtained using the SMD model parameters and the IEF-PCM/Gaussian03 protocol for bulk electrostatics. N is the number of data in a given solute class.

TABLE 9: Mean Signed Errors (kcal/mol) in Ionic Solvation Free Energies[a]

		M05-2X				B3LYP	HF	
solute class	N	MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*	SM8[b]
				Acetonitrile				
H, C, N, O cations[c]	36	7.4	6.9	6.3	6.9	7.1	6.3	4.0
S cations[d]	3	16.5	17.0	16.6	17.4	17.3	16.7	16.1
all cations	39	8.1	7.7	7.1	7.7	7.9	7.1	4.9
H, C, N, O anions[c]	19	−2.6	−3.1	−3.5	−3.2	−2.0	−3.9	−4.5
F, Cl, Br, S anions[d]	11	−2.4	−3.1	−2.8	−2.9	−2.3	−3.6	−3.4
all anions	30	−2.5	−3.1	−3.2	−3.1	−2.1	−3.8	−4.1
all ions	69	3.5	3.0	2.6	3.0	3.6	2.4	1.0
				DMSO				
H, C, N, O cations[c]	4	9.0	8.5	8.0	8.3	8.8	8.4	−1.5
all cations	4	9.0	8.5	8.0	8.3	8.8	8.4	−1.5
H, C, N, O anions[c]	52	−3.0	−3.2	−2.6	−2.8	−2.0	−4.0	−8.0
F, Cl, Br, S anions[d]	15	−2.2	−2.8	−2.2	−2.4	−2.0	−3.4	−3.7
all anions	67	−2.8	−3.1	−2.5	−2.7	−2.0	−3.8	−7.0
all ions	71	−2.1	−2.5	−1.9	−2.1	−1.4	−3.1	−6.7
				Methanol				
H, C, N, O cations[c]	26	0.2	−0.4	−1.2	−0.6	0.1	−1.0	−1.3
Cl, Br cations[d]	3	1.2	1.1	0.2	1.2	2.2	0.3	−1.0
all cations	29	0.3	−0.3	−1.1	−0.5	0.3	−0.9	−1.2
H, C, N, O anions[c]	36	1.8	0.9	1.8	1.7	2.8	−2.0	−1.4
F, Cl, Br anions[d]	15	2.2	1.4	2.5	2.1	2.9	−0.9	−1.8
all anions	51	1.9	1.1	2.0	1.8	2.8	−1.6	−1.5
all ions	80	1.3	0.6	0.9	1.0	1.9	−1.4	−1.4
				Water[e]				
H, C, N, O cations[c]	48	3.4	2.7	1.9	2.6	3.0	2.2	0.4
Cl, S cations[d]	4	4.6	4.6	3.9	4.8	5.3	4.0	1.5
all cations	52	3.5	2.9	2.1	2.8	3.2	2.3	0.5
H, C, N, O anions[c]	43	5.2	4.1	6.0	5.5	5.9	2.4	1.3
F, Cl, Br, S anions[d]	17	6.4	5.5	7.0	6.4	6.4	4.4	0.0
all anions	60	5.5	4.5	6.3	5.8	6.0	3.0	0.9
all ions	112	4.6	3.7	4.3	4.4	4.7	2.7	0.7

[a] All the solvation free energies were obtained using the SMD model parameters and the IEF-PCM/Gaussian03 protocol for bulk electrostatics. [b] The SM8 solvation free energies were obtained with a locally modified version of Gaussian03 called MN-GSM (ref 109) at the M05-2X/CM4 level of theory with the 6-31G* basis set. These SM8 calculations are based on the GB approximation for bulk electrostatics, the CM4 charge model (ref 55), and the SM8 intrinsic Coulomb radii and SM8 atomic surface tensions (ref 42). [c] Ions containing no elements other than H, C, N, or O. [d] Ions containing any of the listed elements in addition to H, C, N, or O. [e] 112 selectively clustered ions from the SM6 model training set as defined in ref 55.

tested this model for the use with other algorithms for solving the NPE for continuum solvation calculations in which the solute is represented by its electron density in real space. These tests include the IEF-PCM algorithm[4−7] as implemented in GAMESS,[63−65] the C-PCM[9−13] algorithm implemented in Gaussian03 and in GAMESS, and the COSMO algorithm[9,10] as implemented in NWChem.[66] Because all of the SMD calculations use the same surface tension parameters and radii and our own analytical surface area algorithm[61] for CDS terms, the differences in the resulting solvation energies are entirely due to the differences in the treatment of bulk electrostatics by different PCM algorithms.

New Continuum Solvation Model

J. Phys. Chem. B, Vol. 113, No. 18, 2009 **6389**

TABLE 10: Mean Unsigned Errors (kcal/mol) in Ionic Solvation Free Energies[a]

| solute class | N | M05-2X | | | | B3LYP | HF | SM8[b] |
		MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*	
				Acetonitrile				
H, C, N, O cations[c]	36	7.4	7.0	6.5	7.0	7.2	6.6	6.5
S cations[d]	3	16.5	17.0	16.6	17.4	17.3	16.7	16.1
all cations	39	8.1	7.7	7.3	7.8	8.0	7.3	7.2
H, C, N, O anions[c]	19	2.8	3.2	3.5	3.2	2.3	3.9	4.5
F, Cl, Br, S anions[d]	11	3.3	3.9	3.8	3.7	3.1	4.4	3.4
all anions	30	3.0	3.4	3.6	3.4	2.6	4.1	4.1
all ions	69	5.9	5.9	5.7	5.9	5.7	5.9	5.9
				DMSO				
H, C, N, O cations[c]	4	9.0	8.5	8.0	8.3	8.8	8.4	1.8
all cations	4	9.0	8.5	8.0	8.3	8.8	8.4	1.8
H, C, N, O anions[c]	52	4.7	4.8	4.3	4.5	4.3	5.3	8.3
F, Cl, Br, S anions[d]	15	3.1	3.5	3.2	3.2	3.0	4.0	4.1
all anions	67	4.3	4.5	4.0	4.2	4.0	5.0	7.3
all ions	71	4.6	4.8	4.3	4.4	4.3	5.2	7.0
				Methanol				
H, C, N, O cations[c]	26	2.0	2.2	2.4	2.2	2.4	2.4	2.1
Cl, Br cations[d]	3	1.2	1.1	0.4	1.2	2.2	0.3	1.0
all cations	29	1.9	2.1	2.2	2.1	2.4	2.2	2.0
H, C, N, O anions[c]	36	2.5	2.1	2.7	2.5	3.3	2.8	2.4
F, Cl, Br anions[d]	15	2.5	2.1	2.6	2.3	2.9	3.3	2.5
all anions	51	2.5	2.1	2.6	2.4	3.2	2.9	2.4
all ions	80	2.3	2.1	2.5	2.3	2.9	2.7	2.3
				Water[e]				
H, C, N, O cations[c]	48	3.5	3.0	2.5	2.9	3.2	2.7	2.4
Cl, S cations[d]	4	4.6	4.6	3.9	4.8	5.3	4.0	2.5
all cations	52	3.6	3.1	2.6	3.1	3.4	2.8	2.4
H, C, N, O anions[c]	43	5.6	4.8	6.2	5.7	6.2	3.5	3.7
F, Cl, Br, S anions[d]	17	6.7	5.7	7.0	6.5	6.6	4.8	3.0
all anions	60	5.9	5.0	6.4	5.9	6.3	3.9	3.5
all ions	112	4.8	4.1	4.6	4.6	4.9	3.4	3.0

[a-e] See footnotes *a–e* of Table 9.

TABLE 11: SMD Mean Unsigned Errors (kcal/mol) in Ionic Solvation Free Energies with Respect to Reference Data Obtained by Using Fawcett's Values of the Solvation Free Energies of H⁺ in Acetonitrile, DMSO, and Methanol

| solute class | N | M05-2X | | | | B3LYP | HF | SM8[a,b] |
		MIDI!6D	6-31G*	6-31+G**	cc-pVTZ	6-31G*	6-31G*	
				Acetonitrile				
H, C, N, O cations[c]	36	4.4	4.3	4.2	4.4	4.6	4.3	7.2
S cations[d]	3	12.5	13.0	12.6	13.4	13.3	12.7	12.1
all cations	39	5.1	4.9	4.9	5.1	5.3	4.9	7.6
H, C, N, O anions[c]	19	2.2	1.9	1.6	1.7	2.6	1.7	1.7
F, Cl, Br, S anions[d]	11	2.6	2.0	1.8	1.9	2.6	1.9	1.8
all anions	30	2.3	1.9	1.7	1.8	2.6	1.8	1.7
all ions	69	3.9	3.6	3.5	3.7	4.1	3.6	5.0
				DMSO				
H, C, N, O cations[c]	4	6.2	5.7	5.2	5.5	6.0	5.6	4.3
all cations	4	6.2	5.7	5.2	5.5	6.0	5.6	4.3
H, C, N, O anions[c]	52	4.2	4.1	3.5	3.9	4.2	4.3	6.1
F, Cl, Br, S anions[d]	15	2.6	2.7	2.3	2.5	2.8	2.8	2.7
all anions	67	3.8	3.8	3.2	3.6	3.9	4.0	5.4
all ions	71	3.9	3.9	3.3	3.7	4.0	4.1	5.3
				Methanol				
H, C, N, O cations[c]	26	2.0	2.1	2.3	2.1	2.3	2.3	2.0
Cl, Br cations[d]	3	1.5	1.4	0.5	1.5	2.5	0.6	0.7
all cations	29	2.0	2.0	2.1	2.0	2.4	2.1	1.9
H, C, N, O anions[c]	36	2.3	2.0	2.5	2.3	3.0	3.0	2.5
F, Cl, Br anions[d]	15	2.4	2.0	2.4	2.2	2.7	3.4	2.6
all anions	51	2.3	2.0	2.5	2.3	2.9	3.1	2.5
all ions	80	2.2	2.0	2.3	2.2	2.7	2.7	2.3

[a-d] See footnotes *a–d* of Table 9. The errors in the SMD and SM8 solvation free energies were calculated with respect to the reference data corrected by using the proton free energies of solvation given by Fawcett (ref 92).

TABLE 12: Deviations (kcal/mol) in Bulk-Electrostatics Solvation Free Energies Calculated Using Various Solvent Models and the SMD Intrinsic Coulomb Radii[a]

solute class	N	C-PCM Gaussian03	IEF-PCM GAMESS	C-PCM GAMESS	COSMO NWChem
		MSD			
all ionic data	332	−0.18	0.14	0.00	−0.94
neutral data ($\varepsilon \geq 32$)	482	−0.08	0.09	0.02	−0.25
neutral data ($\varepsilon < 32$)	2150	−0.49	0.00	−0.45	−0.56
all neutral data	2632	−0.42	0.02	−0.36	−0.50
		MUD			
all ionic data	332	0.18	0.16	0.12	1.61
neutral data ($\varepsilon \geq 32$)	482	0.08	0.10	0.06	0.27
neutral data ($\varepsilon < 32$)	2150	0.49	0.03	0.45	0.56
all neutral data	2632	0.42	0.05	0.38	0.50

[a] All the solvation free energies were obtained with the HF/6-31G* method. The mean signed deviations and the mean unsigned deviations (MSDs and MUDs) are calculated as deviations of the energies obtained using various bulk-electrostatics protocols from those obtained using the IEF-PCM/Gaussian03 protocol. The ionic data are calculated for solutes in acetonitrile, DMSO, methanol, and water, all of which have dielectric constants (ε) greater than or equal to 32. The neutral data are calculated for solutes in 90 nonaqueous solvents and in water.

C-PCM and COSMO solve the nonhomogeneous Poisson equation for an infinite dielectric constant but with scaled dielectric boundary conditions in order to approximate the result for a finite dielectric constant. For both methods tested in the present study we use the $(\varepsilon - 1)/\varepsilon$ dielectric screening factor for the conductor-like surface charge. This is a better approximation at high dielectric constant (e.g., 80) than at low dielectric constant (e.g., below 10). Even for a very low dielectric constant (e.g., 2), the difference between IEF-PCM and C-PCM is sometimes less than the errors in either due to the unrealistic treatment of dielectric response at the solute−solvent boundary (discontinuous change in dielectric constant from 1 to the bulk value at a fixed boundary whose location is somewhat arbitrary) and due to some solute charge lying outside the solute cavity; nevertheless the atomic surface tensions were parametrized using IEF-PCM, and so that model is preferred when available. For SMD calculations with the other electrostatic algorithms, we used the programs' default tessellation settings for calculations with Gaussian03 and GAMESS, but for the SMD/COSMO/NWChem calculations we employed finer grids (options *minbem=3, maxbem=4, ificos=1*) because the default NWChem tessellation parameters (options: *minbem=2, maxbem=3, ificos=0*) produced very large errors in solvation free energies. For all of the SMD electrostatic calculations in Gaussian03, GAMESS, and NWChem, we specified the solvent only by its dielectric constant with the solvent radius set to zero, and the program default nonelectrostatic contributions were turned off while the SMD nonelectrostatic contributions were added to the solvation free energies. We specified the SMD atomic Coulomb radii instead of the default radii used by these programs.

Table 12 shows the corresponding deviations in solvation electrostatic energies for ionic and neutral solutes in various media. The typical difference between IEF-PCM and C-PCM in Gaussian03 for solutes in solvents with dielectric constants greater than or equal to 32 is less than 0.2 kcal/mol for ions and less than 0.1 kcal/mol for neutrals. The typical difference between IEF-PCM and C-PCM for neutral solutes in solvents with dielectric constants less than 32 is less than 0.5 kcal/mol. Note that the differences in Table 12 between the results with essentially the same electrostatics algorithms as implemented in different programs are mostly due to the different tessellation algorithms used by these programs (cf. SMD/COSMO/NWChem and SMD/C-PCM/Gaussian03).

8. Comparison of SMD to Default Solvation Models of Popular Programs

In Table 13, we compare the errors obtained within the SMD model combined with the IEF-PCM, C-PCM, and COSMO algorithms implemented in Gaussian03, GAMESS, and NWChem to those obtained with the default solvation models in these programs (see the next section for a discussion of nondefault alternatives based on IEF-PCM). We used the solvation free energies calculated in previous work[42] employing the default parameters of these program packages for both nonelectrostatic and electrostatic terms, which is indicated by a "d" at the end of the acronym. Thus, in the present article, IEF-PCM-G03d denotes the default IEF-PCM of Gaussian03,[62] C-PCM-G06d denotes the default C-PCM of the 2006 version[65] of GAMESS, and COSMO-N06d denotes the default COSMO of version 4.7 of NWChem (released in 2006).[66] Thus, the Gaussian PCM calculations include not only electrostatics but also the default cavitation, dispersion, and repulsion terms, as described in the original references.[4−7] In contrast, the default in GAMESS and NWChem is to include only electrostatics. In all of these calculations we use the programs' default atomic radii and tessellation grids. All the tested programs by default use the same concentration (1 mol/L) in both the gaseous and solution phases.

For the IEF-PCM-G03d calculations we use the cavities provided by the model called UA0, which denotes the united atom topological model with the atomic radii of the UFF molecular mechanics model. There is more than one united atom model in Gaussian03. We have also tested the model called UAHF,[67] which denotes the united atom topological model with atomic radii optimized for the HF/6-31G(d) level of theory. The UAHF radii are recommended in the Gaussian03 manual for calculations of free energies of solvation with the *scfvac* keyword invoked (we used this keyword in calculations with both types of radii, UA0 and UAHF). The Gaussian03 calculations with the UAHF radii are further denoted as IEF-PCM/G03(UAHF).

To test all of the continuum models we used all the data for 332 ions in acetonitrile, DMSO, methanol, and water as described above. However, tests for neutrals were performed only for those 17 solvents including acetonitrile, DMSO, and water that are available for IEF-PCM in Gaussian03 (actually Gaussian03 supports 21 solvents, but we have neutral data in only 17 of them; see Table 1; for example, although methanol is also available in Gaussian03, we have no neutral data for this solvent). The calculations were performed at the B3LYP,[106−108] Hartree−Fock, M05-2X,[98] and mPW1PW[110] levels of theory using the 6-31G* and 6-31+G** basis sets.[99] Further details on the IEF-PCM-G03d, IEF-PCM/G03(UAHF), C-PCM-G06d, and COSMO-N06d calculations can be found in ref 42.

The SMD/IEF-PCM/G03 model which uses the SMD atomic Coulomb radii for the bulk-electrostatics calculation and the SMD atomic surface tension coefficients for the nonelectrostatic contributions to the free energies of solvation substantially

New Continuum Solvation Model

J. Phys. Chem. B, Vol. 113, No. 18, 2009 **6391**

TABLE 13: Errors (kcal/mol) in Solvation Free Energies Calculated Using Various Solvent Models[a]

method	aqueous neutrals[b]			organic neutrals[c]			ions[d]		
	MSE	MUE	RMSE	MSE	MUE	RMSE	MSE	MUE	RMSE
SMD/IEF-PCM/G03/M05-2X	−0.05	0.59	0.86	−0.06	0.65	0.89	1.49	4.13	5.52
SMD/IEF-PCM/G03/mPW1PW	0.24	0.62	0.91	0.09	0.65	0.89	2.00	4.30	5.43
SMD/IEF-PCM/G03/B3LYP	0.61	0.80	1.08	0.30	0.67	0.92	2.49	4.45	5.65
SMD/IEF-PCM/G03/HF	−0.69	0.91	1.24	−0.29	0.72	0.97	0.39	4.13	5.46
SMD/IEF-PCM/GAMESS/HF	−0.62	0.85	1.16	−0.29	0.72	0.97	0.52	4.13	5.49
SMD/C-PCM/G03/HF	−0.75	0.95	1.29	−0.83	0.97	1.24	0.22	4.15	5.46
SMD/C-PCM/GAMESS/HF	−0.67	0.89	1.20	−0.79	0.95	1.21	0.39	4.16	5.50
SMD/COSMO/NWChem/HF	−0.89	1.06	1.42	−0.88	1.01	1.28	−0.36	4.60	5.96
SM8/M05-2X/CM4	0.21	0.58	0.83	0.11	0.61	0.83	−1.34	4.28	6.22
SM8/mPW1PW/CM4	−0.06	0.55	0.77	−0.02	0.61	0.83	−1.23	4.21	5.63
SM8/B3LYP/CM4	0.14	0.57	0.79	0.11	0.62	0.85	−0.92	4.14	5.57
IEF-PCM-G03d/mPW1PW[e]	4.86	4.87	5.70	5.94	5.99	6.63	7.22	9.53	11.07
IEF-PCM/G03(UAHF)/mPW1PW[f]	0.61	1.18	1.95	3.88	3.94	4.53	5.41	7.96	11.89
IEF-PCM/G03(UAHF)/HF[g]	0.03	1.10	1.75	3.62	3.71	4.31	3.83	7.50	10.60
IEF-PCM/G03(UAHF)/HF/6-31+G**[h]	−0.51	1.36	1.96	3.43	3.55	4.16	3.64	7.34	10.46
C-PCM-G06d/B3LYP[i]	−0.65	1.57	1.85	2.62	2.78	3.15	4.24	8.14	9.26
COSMO-N06d/B3LYP[j]	−8.17	8.17	9.50	−2.12	2.76	3.68	−12.43	12.64	15.06

[a] All data in this table are for the 6-31G basis set unless noted otherwise. MSE, MUE, and RMSE refer to mean signed, mean unsigned, and root mean squared error, respectively. The SMD calculations were based on IEF-PCM, C-PCM, or COSMO protocols for bulk electrostatics as implemented in Gaussian03 (G03), GAMESS, or NWChem, the SMD intrinsic Coulomb radii, and the SMD atomic surface tensions. The SM8 calculations were based on the GB approximation for bulk electrostatics, the CM4 charge model (ref 55), and the SM8 intrinsic Coulomb radii and SM8 atomic surface tensions (ref 42); they were performed with a locally modified version of Gaussian03 called MN-GSM.[109] The mPW1PW density functional (also called MPW25, mPW0, and mPW1PW91) is described elsewhere (ref 110). [b] 274 neutral data. [c] 666 neutral data from the present work for SMD and from previous work (ref 42) for SM8 and the default settings for IEF-PCM, C-PCM, and COSMO (see ref 42 for detail). See Table 15 for the corresponding SMD and SM8 results on a larger data set. [d] 332 data for ions in acetonitrile, DMSO, methanol, and water. Note that the corresponding errors for the SM8 model and the default solvation models in Gaussian03, GAMESS, and NWChem previously reported in ref 42 have been recalculated in the present work using the reference solvation energies of aqueous clustered ions corrected by 2.38 kcal/mol (see eq 15 and the text following the equation). [e] Based on the default settings for the IEF-PCM model implemented in Gaussian03 at the mPW1PW level of theory. For these calculations we use the cavities provided by the model called UA0. [f] Same as previous row except for radii used to define the electrostatic cavities. For these calculations we use the cavities provided by the model called UAHF. [g] Same as previous row except based on a Hartree−Fock wave function for the solute (whereas the previous row is based on an mPW1PW solute wave function). In this row the UAHF radii are used with the kind of wave function for which they were optimized. [h] Same as previous row except for the 6-31+G** basis set used. [i] Based on the default settings for the C-PCM model implemented in GAMESS at the B3LYP level of theory. [j] Based on the default settings for the COSMO model implemented in NWChem at the B3LYP level of theory.

outperforms the default IEF-PCM-G03d model and the default algorithm for calculation of the nonelectrostatic terms.[4−7] The IEF-PCM-G03d/mPW1PW mean unsigned errors compared to the SMD/IEF-PCM/G03/mPW1PW errors are 8−9 times larger for tested neutrals and 2 times larger for tested ions (Table 13). Such large errors indicate that the default settings for the IEF-PCM algorithm in Gaussian03 cannot be recommended for the quantitative prediction of solvation free energies. Such large errors in the IEF-PCM-G03d calculations can be attributed to the use of the UA0 radii by the default PCM in Gaussian03. The performance of the default PCM in Gaussian03 can be improved by use of the UAHF radii. Nevertheless, the resulting errors in the solvation free energies of tested compounds remain unacceptably large. For instance, the IEF-PCM/G03(UAHF)/HF/6-31G* mean unsigned errors compared to the SMD/IEF-PCM/G03/HF/6-31G* errors are 1.2, 5.2, and 1.8 times larger, respectively, for aqueous neutrals, nonaqueous neutrals, and ions (Table 13). The same large errors are found in calculations using the default C-PCM in GAMESS and the default COSMO in NWChem. The C-PCM-G06d model produces mean unsigned errors that are 1.7−3.0 times larger than the typical SMD/C-PCM errors. The COSMO-N06d mean unsigned errors are even larger (up to 12.64 kcal/mol on average for ions): this model produces overly negative solvation energies across all classes of the tested solutes (Table 13). The very poor performance of the COSMO-N06d model is mainly due to the default tessellation grid, which is inadequately coarse.

The ionic data in Table 13 are averaged over cations and anions in all solvents. Table 14 breaks their mean unsigned errors down into categories. As compared to previous models, the SMD model is more accurate than SM8 for nonaqueous anions, but not for the other three categories. It is more accurate than the default models of GAMESS and NWChem for all four categories, and it is more accurate than the default model of Gaussian03 for all categories except nonaqueous anions, for which the default model of Gaussian03 uses different parameters than for cations. Very recently, Ginovska et al.[15] have presented a new model called CD-COSMO, and they have shown that one can greatly reduce the errors for solvation free energies of aqueous ions by using atomic radii that depend on the partial atomic charges.

For the SMD and SM8 models, we can consider all 90 organic solvents. Table 15 shows errors for this larger set; the trends and magnitudes are the same as in Table 13.

9. Alternative Protocols Using IEF-PCM Electrostatics

In order to clarify the differences between SMD and alternative continuum approaches to the computation of solvation free energies, it is instructive to consider a number of different models that have been described in the literature based on IEF-PCM electrostatics[4] (or, in some cases, the earlier but closely related PCM model[2]).

First, however, it is perhaps appropriate to note that some confusion can arise because the term "IEF-PCM" can mean different things in different contexts. Ideally, IEF-PCM should probably be restricted to mean the formal integral equation construction of the PCM approach to solving the NPE for an

TABLE 14: Mean Unsigned Errors (kcal/mol) in Solvation Free Energies of Ions[a]

method	aqueous		nonaqueous	
	52 cations	60 anions	72 cations	148 anions
SMD/IEF-PCM/G03/M05-2X	3.1	5.0	5.5	3.5
SMD/IEF-PCM/G03/mPW1PW	3.3	5.6	5.7	3.4
SMD/IEF-PCM/G03/B3LYP	3.4	6.3	5.8	3.4
SMD/IEF-PCM/G03/HF	2.8	3.9	5.3	4.1
SMD/IEF-PCM/GAMESS/HF	2.9	4.0	5.4	4.0
SMD/C-PCM/G03/HF	2.8	3.8	5.3	4.2
SMD/C-PCM/GAMESS/HF	2.9	3.9	5.4	4.1
SMD/COSMO/NWChem/HF	3.4	3.0	5.5	5.2
SM8/M05-2X/CM4	2.4	3.5	4.8	5.0
SM8/mPW1PW/CM4	2.4	3.4	4.7	4.9
SM8/B3LYP/CM4	2.4	3.7	4.8	4.6
IEF-PCM-G03d/mPW1PW[b]	10.5	12.8	14.0	5.7
IEF-PCM/G03(UAHF)/mPW1PW[c]	5.9	9.9	16.5	3.8
IEF-PCM/G03(UAHF)/HF[d]	5.4	8.2	16.0	3.8
IEF-PCM/G03(UAHF)/ HF/6-31+G**[e]	4.9	8.1	15.7	3.8
C-PCM-G06d/B3LYP[f]	7.3	8.1	12.0	6.6
COSMO-N06d/B3LYP[g]	11.2	7.9	6.7	18.0

[a] All data in this table are for the 6-31G* basis set unless noted otherwise. 332 data were used for ions in acetonitrile, DMSO, methanol, and water. Other details are the same as in footnote *a* of Table 13. Note that the IEF-PCM default model in Gaussian03 uses different parameters for cations and anions, whereas the other methods in this table involve a single set of parameters, independent of charge state and independent of partial atomic charges. [b] Details are the same as in footnote *e* of Table 13. [c] Details are the same as in footnote *f* of Table 13. [d] Details are the same as in footnote *g* of Table 13. [e] Details are the same as in footnote *h* of Table 13. [f] Details are the same as in footnote *i* of Table 13. [g] Details are the same as in footnote *j* of Table 13.

TABLE 15: Errors (kcal/mol) in Solvation Free Energies for the SMD and SM8 Models for the Full Set of 2072 Neutral Data in Nonaqueous Solvents[a]

method	MSE	MUE	RMSE
SMD/IEF-PCM/G03/M05-2X	−0.06	0.64	0.86
SMD/IEF-PCM/G03/mPW1PW	0.10	0.63	0.86
SMD/IEF-PCM/G03/B3LYP	0.31	0.67	0.92
SMD/IEF-PCM/G03/HF	−0.33	0.73	0.99
SMD/IEF-PCM/GAMESS/HF	−0.33	0.72	0.98
SMD/C-PCM/G03/HF	−0.79	0.95	1.21
SMD/C-PCM/GAMESS/HF	−0.75	0.92	1.18
SMD/COSMO/NWChem/HF	−0.84	0.99	1.26
SM8/M05-2X/CM4	0.08	0.57	0.77
SM8/mPW1PW/CM4	−0.05	0.57	0.78
SM8/B3LYP/CM4	0.07	0.58	0.79

[a] Details are the same as in footnote *a* of Table 13.

arbitrary cavity.[4] In order to put this formalism into practice, various computational and parametric choices must be made, especially with respect to construction of the molecular cavity. Several choices for cavity construction have been explored in the literature, as described in more detail momentarily, but these choices have sometimes not been specified in subsequent applications by nondevelopers. In addition to alternative cavity construction protocols, several different strategies for accounting for non-bulk-electrostatic solvation effects have also been considered. However, as noted above, implementations of PCM or IEF-PCM electrostatics in different electronic structure programs may include different default options for handling these components, so that researchers who report results by simply labeling them as IEF-PCM are not fully specifying the calculations.

These points having been made, we turn now to the question of modeling the nonbulk-electrostatic component of the solvation free energy. Floris et al.[111-113] were the first to consider addition of an exchange-repulsion term that was computed from classical pair potentials for a solute X in a solvent S according to

$$G_{disp-rep}(X) = \rho_S \sum_{l \in S} N_l \sum_{m \in X} \sum_j A_j \left[\frac{d_{lm}}{3r_{mj}^6} + c_{lm} \left(\sum_{k=1}^{3} \frac{k}{\gamma_{lm} r_{mj}^k} \right) \exp(-\gamma_{lm} r_{mj}) \right] \hat{\mathbf{r}}_{mj} \cdot \hat{\mathbf{n}}_j \quad (20)$$

where ρ_S is the density of the solvent, l runs over solvent atom types, N_l is the number of atoms of type l in S, m runs over solute atoms, j runs over surface tesserae (defined for a *different* cavity than that used for PCM electrostatics), A_j is the area of tessera j, r_{mj} is the distance from solute atom m to the center of tessera j, $\hat{\mathbf{r}}_{mj}$ is the unit vector in this direction, $\hat{\mathbf{n}}_j$ is the inward normal vector at the center of tessera j, and c_{lm}, d_{lm}, and g_{lm} are empirical pair potential parameters between atom types l and m. Floris and Tomasi used the pair potential parameters of Claverie and co-workers,[114-116] which were fit to reproduce free energies of vaporization and sublimation in condensed phases. This approach continues to be used by default in Gaussian03. Also used by default is a model that predicts free energies of cavitation based upon the scaled particle theory of Pierotti[117,118] as modified by Claverie.[119] In this approach

$$G_{cav}(X) = \sum_{m \in X} \frac{A_m}{4\pi R_m^2} \sum_{j=0}^{3} K_j(P, T; \rho_S, R_S)(R_m + R_S)^j \quad (21)$$

where m runs over solute atoms (or groups), A_m is the area of the sphere about atom m having radius R_m that is exposed to the solvent, R_S is a characteristic radius associated with the solvent, and the factors K_j depend upon pressure P, temperature T, and the number density ρ_S of the solvent.

As noted in section 8, the default "total" IEF-PCM model in Gaussian03, which combines IEF-PCM electrostatics using UA0 radii[120] with a CDS component computed from eqs 20 and 21, predicts free energies of solvation in rather poor agreement with experiment. In 1997, however, Barone, Cossi, and Tomasi[67] proposed a modification to the *electrostatic* portion of PCM calculations to improve the accuracy of predicted *aqueous* solvation free energies. In particular, they proposed a new cavity generation protocol (the original PCM electrostatic cavity was built from spheres having Bondi radii scaled by a factor of 1.2, but the differences associated with using scaled Bondi vs UA0 radii and PCM vs IEF-PCM electrostatics do not have a major impact on accuracy) called united-atom Hartree−Fock (UAHF). In the UAHF scheme, spheres are not placed on hydrogen atoms, but spheres on heavy atoms have radii that depend on the number of H atoms attached. Heavy-atom spheres are additionally dependent on the hybridization state of the atom and on the number of attached "active" atoms (atoms having that same hybridization state and atomic number as the atom under consideration and are further bonded only to C and H atoms). In addition, atomic radii are dependent on formal charge. Assigning hybridization and charge is accomplished heuristically. (Note that SMD parameters do not depend on charges, hybridization states, or classifications of other atoms as attached or unattached and thus there are no discontinuities when the model is applied along reaction paths.) Combining PCM(UAHF)/

HF/6-31G* electrostatics with CDS free energies from eqs 20 and 21, Barone, Cossi, and Tomasi reported a mean unsigned error of 0.2 kcal/mol for 43 neutral solutes in water and 1 kcal/mol for 27 ions in water.[67] Over larger aqueous data sets of 274 neutrals and 112 ions, however, we found IEF-PCM(UAHF)/HF/6-31G* to predict mean unsigned errors of 1.1 and 6.9 kcal/mol, respectively (Tables 13 and 14). We note that another cavity scheme, united-atom Kohn−Sham (UAKS), is listed in the Gaussian03 users manual as having been optimized for PCM/PBE/6-31G* calculations, and it is presumably similar in design to UAHF. The errors on the larger data set suggest that the original UAHF training set lacks broad coverage of organic functionality. The UAHF model also poses complications insofar as a given solute cavity does not change in a smooth and continuous fashion as a heavy atom gains/loses substituents, changes charge, or changes hybridization. Finally, the IEF-PCM(UAHF)/HF/6-31G* model with CDS contributions from eqs 20 and 21 shows 3-fold larger errors for neutrals in organic solvents (Table 13), suggesting that the UAHF cavity construction protocol for water is not general for organic media.

Pursuing a different approach, Luque, Orozco, and co-workers have employed IEF-PCM electrostatics as part of overall solvation models[68] that they refer to as MST models (named for Miertuš, Scrocco, and Tomasi, the authors of the original PCM paper[2]). The MST models were originally developed for predicting the solvation free energies of neutral solutes in water,[121−123] octanol,[124] chloroform,[125] and carbon tetrachloride[126] using PCM electrostatics. Small quantitative changes in parameter values were associated with extension to IEF-PCM electrostatics.[68] In the MST models, there are separate cavities for the electrostatic and nonelectrostatic portions of the calculation. In the electrostatic portion, Pauling radii are scaled by solvent-dependent factors ranging from 1.25 for water to 1.80 for carbon tetrachloride; spheres are placed on all atoms, but a distinction is made between polar and nonpolar H atoms (radii of 0.9 and 1.2 Å, respectively). In addition, in water, N, O, and S atoms take on different radii depending on the numbers of attached hydrogen atoms. In the nonelectrostatic portion, no scaling factor is applied to the cavity. Cavitation is computed using eq 21, but instead of using eq 20, the MST models define a remaining so-called van der Waals free energy computed as

$$G_{vdw}(X) = \sum_{m \in X} \xi_m A_m \qquad (22)$$

where A_m is the solvent-exposed surface area of group m and ξ_m is the surface tension parameter empirically optimized for group m. In this respect, the MST model is similar in parametrization strategy to the earlier SMx models. Groups are defined as sums of heavy atoms and attached hydrogen atoms and the surface tension parameters ξ have been optimized over training sets of 72, 63, 56, and 48 neutral solutes in water,[121−123] octanol,[124] chloroform,[125] and carbon tetrachloride,[126] respectively. In these four solvents, the MST models, which have been optimized for HF and B3LYP calculations with the 6-31G* basis set, predict neutral solvation free energies with errors similar in magnitude to those observed with SMD. In the case of ionic solvation free energies in water, Curutchet et al. examined the aqueous MST model and reported rms errors of 5.8 and 6.9 kcal/mol for 25 anions and 22 cations, respectively, provided that the scaling factor used in construction of the electrostatic cavity was reduced from 1.25 to 1.15.[127]

Several other combinations of IEF-PCM electrostatics with different approaches to compute nonelectrostatic solvation free energies have appeared. Amovilli and Mennucci[128,129] have described an approach in which the dispersion−repulsion term is dependent on the solute wave function, so that it is included in the SCRF process. For water, octanol, chloroform, and carbon tetrachloride, Curutchet et al.[70] compared this fully SCRF approach to results from the MST models (computed using eq 22) and various empirical pair potentials, including the one used by default in Gaussian03. They noted that the default pair potentials in Gaussian03 agreed reasonably with SCRF dispersion−repulsion energies for water as a solvent, but large, nonsystematic rms deviations of 5−10 kcal/mol over 22 neutral solutes were observed in the organic solvents. The use of other classical pair potentials taken from commonly available force fields did not improve this situation. MST van der Waals energies, on the other hand, had rms deviations compared to QM-SCRF of about 3.5 kcal/mol in organic solvents.

Considering exclusively water as solvent, other groups have explored variations on the PCM scheme. In particular, Shimizu et al.[69] combined PCM electrostatics (using unscaled Bondi radii for all atoms) with an alternative formulation of the cavitation and dispersion-repulsion free energies

$$G_{cav}(X) = c_0 + c_1 SES_X \qquad (23)$$

$$G_{disp}(X) = \rho_S \sum_{m \in X} \sum_j A_j \left(\frac{d_{Z_m}}{3 r_{mj}^6} \right) \hat{\mathbf{r}}_{mj} \cdot \hat{\mathbf{n}}_j \qquad (24)$$

In eq 23, SES is the solvent-excluding surface area of solute X and the coefficients c are parameters to be optimized (the form of this equation is related to one first proposed by Tuñón et al.[130]). In eq 24, the various terms have the same meaning as in eq 20 except that the solute−solvent pair-potential parameters have been replaced with solute atomic-number specific parameters d_{Z_m} to be optimized for atoms H, C, N, O, F, S, Cl, Br, and I. The terms corresponding to repulsion in eq 20 do not appear in eq 24; their effect is absorbed into the empirical parameters in the optimization process. Shimizu et al.[69] named their model PCM2 and reported aqueous solvation free energies for 215 neutral molecules with a mean absolute deviation from experiment of 0.7 kcal/mol at the HF/6-31G* level. They obtained the same error with the aqueous PCM(UAHF) model. Ions were not considered.

A different aqueous variation has been reported by Basilevsky et al.,[72] who combined PCM(UA0) electrostatics with a binomial cell cavitation contribution[131]

$$G_{cav}(X) = \left[\theta_S \ln \left(\frac{\theta_S}{\theta_S - 1} \right) \right] \rho_S R T V_X \qquad (25)$$

where θ has a value between 1.2 and 1.3, ρ_S is the number density of the solvent, R is the universal gas constant, T is temperature, and V_X is the molecular volume of the solute nonelectrostatic cavity, in this case constructed from thermal radii determined from classical simulations. For dispersion−repulsion, Basilevsky et al.[72,131] adopted an approach similar to that in eq 24, except that the d parameters run over atomic types, rather than being specific to atomic numbers, and in addition to a 6th-power term, 8h- and 12th-power terms were included. The 23 atomic types were taken from the MMFF force field.[132] This model provided a rmsd error compared to experiment of about 0.6 kcal/mol for 278 neutral solutes; ions were not considered.[72,131]

In a more limited study, de O. Mendes et al.[71] considered the effects of different electrostatic and nonelectrostatic cavity formulations on solvation free energies for homologous series of alkanes, alcohols, and carbonyl compounds, and concluded that spheres on united-atom functional groups, e.g., C=O, could provide quantitatively useful predictions when combining PCM electrostatics with eqs 20 and 21 for nonelectrostatic effects. Finally, Böes et al. have considered the solvation free energies of various anions in acetonitrile,[133] dimethylformamide,[133] nitrobenzene, [134] formamide,[73] and methanol[73] as solvents. To model these species, they adopted an MST-like formalism (vide supra), fitting ξ parameters for H, B, C, N, O, F, S, Cl, Br, and I for each of the various solvents. In their most recent study,[73] Böes et al. reported mean unsigned errors of 1.3 and 2.0 kcal/mol for 13 anions in formamide and 16 anions in methanol. Such small errors, however, must be regarded as somewhat suspect given 10 fitted parameters for as few as 13 data (thus, for instance, since Br^- is the only anion incorporating Br the ξ parameter ensures an error of zero).

Having summarized all of the various approaches to combine IEF-PCM or PCM electrostatics with nonelectrostatic protocols to predict free energies of solvation, we should of course mention that there are other quantum mechanical continuum solvation models that have the same goal that do *not* follow the PCM or IEF-PCM procedures. We refer interested readers to recent reviews.[29,30] A number of statistical mechanical and fragment-based models for computing solvation free energies also exist, but these are outside the scope of our present discussion.

Given the various results summarized above, it should be clear that it is possible to combine PCM and IEF-PCM electrostatics with different protocols for the computation of the non-bulk-electrostatic free energy of solvation so as to arrive at a complete model that provides quantitatively useful free energies of solvation. Such models, however, inevitably rely on some parametrization against experimental data to achieve reasonable accuracy, e.g., the ξ parameters in MST models or the σ values of SM*x* models. When large sets of solvation data are available, as is the situation for chloroform, for example, solvent-specific solvation models can be created, like chloroform-specific MST[125] and SM5.4.[135] The fundamental difference between SMD and other PCM-based models, however, is the *universality* of the former in its *construction*. The other models are universal in their *philosophy*: that is, were sufficient data to be available, accurate parameters could presumably be determined for any solvent; unfortunately, data in solvents other than the few mentioned above in this section tend to be sparse. To address this challenge, SMD relies on parametric *functions* (as opposed to constants) that take as input physical properties associated with a given solvent (or more generalized condensed phase). This approach permits the construction of an experimental training set that combines data from 90 different solvents, guaranteeing broad coverage of solute and solvent functionality, and significantly reduces the potential for solvent-specific parameters to be underdetermined. Upon parametrization, the resulting model is immediately applicable to any condensed phase for which the relevant solvent physical properties are known or can reasonably be estimated. Thus, while SMD is no more (nor less) accurate than other well parametrized quantum mechanical solvation models based on PCM or IEF-PCM electrostatics, it is the only such model that is universally defined for any solvent and moreover available for immediate application in a number of widely distributed free and commercial electronic structure packages. It is also more accurate than the default nonuniversal models of several popular programs.

10. Conclusions

Using the training set of 2821 solvation free energies including 274 data for neutral solutes in water, 2072 data for neutral solutes in 90 nonaqueous solvents, 143 neutral transfer free energies between water and 15 organic solvents, and 332 data for ions in acetonitrile, dimethyl sulfoxide, methanol, and water, a new universal implicit solvent model called SMD has been developed for predicting solvation free energies of neutral and ionic solutes in water and in nonaqueous solvents. For nonaqueous solvents, SMD uses a small set of solvent descriptors that characterize the properties of the solvent.

The SMD model is based on a self-consistent reaction field treatment of bulk electrostatics that involves an integration of the nonhomogeneous Poisson equation and on a cavity-dispersion-solvent-structure protocol for the nonelectrostatic contribution to the free energy of solvation. The SMD model employs a single set of parameters optimized over six electronic structure methods, namely, M05-2X/MIDI!6D, M05-2X/6-31G*, M05-2X/6-31+G**, M05-2X/cc-pVTZ, B3LYP/6-31G*, and HF/6-31G*. The model parameters are intrinsic Coulomb radii for the bulk electrostatic calculation and atomic surface tension coefficients for the CDS contribution. With the 6-31G* basis set, the SMD model achieves mean unsigned errors of 0.6−1.0 kcal/mol in the solvation free energies of tested neutrals and mean unsigned errors of 4 kcal/mol on average for ions with either Gaussian03 or GAMESS, whereas the default solvation methods of these programs have mean unsigned errors of 1−6 kcal/mol for these neutrals and 7−10 kcal/mol for ions.

Unlike the earlier SM8 model based on the generalized Born approximation for bulk electrostatics, the density-based solvation model SMD does not utilize partial atomic charges and, therefore, the applicability of the SMD model does not depend on the availability of reasonable charges for a given level of theory. However, the SMD model is slightly less accurate than the SM8 model. For instance, for the mPW1PW1/6-31G* method with which the SM8 model was parametrized, the SMD model achieves mean unsigned errors that are slightly larger than those obtained with SM8/mPW1PW1/6-31G*: 0.62 kcal/mol vs 0.55 for neutrals in water (Table 13), 0.63 kcal/mol vs 0.57 for neutrals in organic solvents (Table 15), and 4.30 kcal/mol vs 4.21 for ions (Tables 13, 14). In many cases the errors of SMD relative to SM8 are not systematic.

The SMD model has been implemented in the GESOL program.[136] To run an SMD calculation, GESOL employs the External option of Gaussian03.[62] This requires a Gaussian03 executable but does not require Gaussian03 source code. The SMD/GESOL model is recommended for liquid-phase single-point energy calculations, geometry optimizations, and Hessian evaluations with any of the methods for which it has been parametrized. With a variable accuracy, it may also be used with any other ground-electronic-state electronic structure method implemented in Gaussian03 for which the IEF-PCM or C-PCM models are available.

Supporting Information Available: Two thousand three hundred forty-six reference solvation free energies and 143 reference transfer energies for neutral solutes in the SMD training set; reference free energies for 112 selectively clustered ions in water; and 220 unclustered ions in acetonitrile, DMSO, and methanol (part I); complementary tables with the SMD errors in the solvation free energies for neutral compounds (part II); and the functional forms of atomic surface tensions used by SMD (part III). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

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

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

(3) Miertuš, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(4) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

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

(6) Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506.

(7) Tomasi, J.; Mennucci, B.; Cancès, E. *J. Mol. Struct. (Theochem)* **1999**, *464*, 211.

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

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

(10) Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253.

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

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

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

(14) Gregerson, L. N.; Baldridge, K. K. *Helv. Chim. Acta* **2003**, *86*, 4112.

(15) Ginovska, B.; Camaioni, D. M.; Dupuis, M.; Schwerdtfeger, C. A.; Gil, Q. *J. Phys. Chem. A* **2008**, *112*, 10604.

(16) Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744.

(17) Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558.

(18) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351.

(19) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(20) Rivail, J.-L.; Rinaldi, D. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 1996; Series 1, Vol. 1, p 139.

(21) Hoijtink, G. J.; de Boer, E.; van der Meij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas Belg.* **1956**, *75*, 487.

(22) Peradejordi, F. *Cahiers Phys.* **1963**, *17*, 393.

(23) Jano, I. *C. R. Acad. Sci.* **1965**, *261*, 103.

(24) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Wiley: London, 1981; Vol. 2, p 25.

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

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

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

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

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

(30) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

(31) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 877.

(32) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325.

(33) Tschampel, S. M.; Kennerty, M. R.; Woods, R. J. *J. Chem. Theory Comput.* **2007**, *3*, 1721.

(34) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817.

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

(36) Marenich, A. V.; Olson, R. M.; Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2055.

(37) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066.

(38) Pethica, B. A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6253.

(39) Smith, P. E.; Van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735.

(40) Boresch, S.; Archontis, G.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25.

(41) 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, The Netherlands, 2006; p 112.

(42) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011.

(43) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760. Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* [Online early access]. DOI: 10.1021/ar900004j. Published Online: Feb 17, 2009.

(44) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552.

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

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

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

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

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

(50) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117. Errata: **1999**, *111*, 5624; **2000**, *113*, 3930.

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

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

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

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

(55) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(56) Kelly, C. P. Ph.D. Thesis, University of Minnesota, Minneapolis, 2007.

(57) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046.

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

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

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

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

(62) 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. *Gaussian03, Revisions C.01, C.02, D.02, and E.01*; Gaussian, Inc.: Wallingford, CT, 2004.

(63) 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.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. *J. Comput. Chem.* **1993**, *14*, 1347.

(64) 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.

(65) 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).

(66) 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.; Dyall, 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.

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

(68) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Orozco, M.; Luque, F. J. *J. Mol. Struct. (Theochem)* **2005**, *727*, 29.

(69) Shimizu, K.; Freitas, A. A.; Farah, J. P. S.; Dias, L. G. *J. Phys. Chem. A* **2005**, *109*, 11322.

(70) Curutchet, C.; Orozco, M.; Luque, F. J.; Mennucci, B.; Tomasi, J. *J. Comput. Chem.* **2006**, *27*, 1769.

(71) de O. Mendes, C. L.; da Silva, C. O.; da Silva, E. C *J. Phys. Chem. A* **2006**, *110*, 4034.

(72) Basilevsky, M. V.; Leontyev, I. V.; Luschekina, S. V.; Kondakova, O. A.; Sulimov, V. B. *J. Comput. Chem.* **2006**, *27*, 552.

(73) Böes, E. S.; Bernardi, E.; Stassen, H.; Gonçalves, P. F. B. *Chem. Phys.* **2008**, *344*, 101.

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

(75) Wangsness, R. K. *Electromagnetic Fields*; Wiley: New York, 1979; p 179.

(76) 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.

(77) Wrobel, L. C.; Aliabadi, M. H. *The Boundary Element Method*; Wiley: Hoboken, NJ, 2002.

(78) Pascual-Ahuir, J. L.; Silla, E. *J. Comput. Chem.* **1990**, *11*, 1047.

(79) Pomelli, C. S.; Tomasi, J.; Cammi, R. *J. Comput. Chem.* **2001**, *22*, 1262.

(80) Frediani, L.; Cammi, R.; Pomelli, C. S.; Tomasi, J.; Ruud, K. *J. Comput. Chem.* **2004**, *25*, 375.

(81) Scalmani, G.; Rega, N.; Cossi, M.; Barone, V. *J. Comput. Methods Sci. Eng.* **2002**, *2*, 469.

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

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

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

(85) 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*.

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

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

(88) Abraham, M. H. In *Quantitative Treatment of Solute/Solvent Interactions; Theoretical and Computational Chemistry Series*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; Vol. 1, p 83.

(89) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database version 2009*; University of Minnesota: Minneapolis, MN, 2009. Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G.; *Minnesota Solvent Descriptor Database version 1999*; University of Minnesota: Minneapolis, MN, 1999.

(90) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408.

(91) 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.

(92) Fawcett, W. R. *Langmuir* **2008**, *24*, 9868.

(93) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A., III. *J. Phys. Chem. B* **2008**, *112*, 9709.

(94) Latimer, W. M.; Pitzer, K. S.; Slansky, C. M. *J. Chem. Phys.* **1939**, *7*, 108.

(95) Hyun, J.-K.; Ichiye, T. *J. Phys. Chem. B* **1997**, *101*, 3596.

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

(97) Babu, C. S.; Lim, C. *Chem. Phys. Lett.* **1999**, *310*, 225.

(98) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.

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

(100) Jaguar 6.5, Release 112; Schrödinger, Inc.: Portland, OR, 2005.

(101) Klamt, A.; Eckert, F. *Fluid Phase Equilib.* **2000**, *172*, 43.

(102) Stefanovich, E. V.; Truong, T. N. *Chem. Phys. Lett.* **1995**, *244*, 65.

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

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

(105) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

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

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

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

(109) 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 2008*; University of Minnesota: Minneapolis, MN, 2008.

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

(111) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616.

(112) Floris, F. M.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784.

(113) Floris, F. M.; Tani, A.; Tomasi, J. *Chem. Phys.* **1993**, *169*, 11.

(114) Caillet, J.; Claverie, P.; Pullman, A. *Acta Crystallogr.* **1978**, *B34*, 3266.

(115) Claverie, P. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; Reidel: Dordrecht, 1982; Vol. 3, p 151.

(116) Vigné-Maeder, F.; Claverie, P. *J. Am. Chem. Soc.* **1987**, *109*, 24.

(117) Pierotti, R. A. *J. Phys. Chem.* **1963**, *67*, 1840.

(118) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.

(119) Claverie, P. In *Intermolecular Interactions: From Diatomics to Biomolecules*; Pullman, B., Ed.; Wiley: Chichester, U.K., 1978; p 69.

(120) Barone, V.; Improta, R.; Rega, N. *Theor. Chem. Acc.* **2004**, *111*, 237.

(121) Bachs, M.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 446.

(122) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847.

(123) Orozco, M.; Bachs, M.; Luque, F. J. *J. Comput. Chem.* **1995**, *16*, 563.

(124) Curutchet, C.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2001**, *22*, 1180.

(125) Luque, F. J.; Zhang, Y.; Alemán, C.; Bachs, M.; Gao, J.; Orozco, M. *J. Phys. Chem.* **1996**, *100*, 4269.

(126) Luque, F. J.; Bachs, M.; Alemán, C.; Orozco, M. *J. Comput. Chem.* **1996**, *17*, 806.

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

(128) Amovilli, C. *Chem. Phys. Lett.* **1994**, *229*, 244.

(129) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051.

(130) Tuñón, I.; Silla, E.; Pascual-Ahuir, J. L. *Chem. Phys. Lett.* **1993**, *203*, 289.

(131) Alexandrovsky, V. V.; Basilevsky, M. V.; Leontyev, I. V.; Mazo, M. A.; Sulimov, V. B. *J. Phys. Chem. B* **2004**, *108*, 15830.

(132) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490.

(133) Böes, E. S.; Livotto, P. R.; Stassen, H. *Chem. Phys.* **2006**, *331*, 142.

(134) Böes, E. S.; de Andrade, J.; Stassen, H.; Gonçalves, P. F. B. *Chem. Phys. Lett.* **2007**, *436*, 362.

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

(136) Marenich, A. V.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *GESOL-version 2008*; University of Minnesota: Minneapolis, MN, 2008.