JCTC Journal of Chemical Theory and Computation

Universal Solvation Model Based on the Generalized Born Approximation with Asymmetric Descreening

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

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

Abstract: We present a new self-consistent reaction field continuum solvation model based on the generalized Born (GB) approximation for the bulk electrostatic contribution to the free energy of solvation. The new model improves on the earlier SM8 model by using the asymmetric descreening algorithm of Grycuk to treat dielectric descreening effects rather than the Coulomb field approximation; it will be called Solvation Model 8 with asymmetric descreening (SM8AD). The SM8AD model is applicable 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. It does not require the user to assign molecular mechanics types to an atom or a group; all parameters are unique and continuous functions of geometry. This model employs a single set of parameters (solvent acidity-dependent intrinsic Coulomb radii for the treatment of bulk electrostatics and solvent description-dependent atomic surface tensions coefficients for the treatment of nonelectrostatic and short-range electrostatic effects). The SM8AD model was optimized over 26 combinations of theoretical levels including various basis sets (MIDI!, 6-31G*, 6-31+G*, 6-31+G**, 6-31G**, cc-pVDZ, DZVP, 6-31B*) and electronic structure methods (M05-2X, M05, M06-2X, M06, M06-HF, M06-L, mPW1PW, mPWPW, B3LYP, HF). It may be used with confidence with any level of electronic structure theory as long as self-consistently polarized Charge Model 4 or other self-consistently polarized charges compatible with CM4 charges are used, for example, CM4M charges can be used. With M05-2X/6-31G*, the SM8AD model achieves a mean unsigned error of 0.6 kcal/mol on average over 2 560 solvation free energies of tested aqueous and nonaqueous neutral solutes and a mean unsigned error of 3.9 kcal/mol on average over 332 solvation free energies of aqueous and nonaqueous ions.

1. Introduction

The electrostatic contribution[1] to the free energy of solvation results from the interaction of a solute with its reaction field, which is the electric field produced by the polarized charge density that the solute induces in the solvent. In self-consistent reaction field theory, the solute is polarized self-consistently by the reaction field, and it is the interaction of the mutually polarized solute and solvent subsystems that is called the electrostatic contribution.[2-4] The electrostatic contribution can be evaluated by solving the nonhomogeneous Poisson equation (NPE, by which we mean the Poisson equation for a nonhomogeneous medium in which the dielectric constant is unity inside the solute cavity but has a nonunit value outside it) in terms of the continuous charge density or by using alternative approaches, for instance, the generalized Born (GB) approximation,[5-11] which does not start with the NPE but instead employs a starting point based on representation of the solute as a collection of point charges, located at the nuclear positions. We have previously introduced a series of successively improved self-consistent reaction field solvation models based on the GB approxima-

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

tion (SM5.4,[12] SM5.42,[13−15] SM5.43,[16] SM6,[17] SM8[18]) or the nonhomogeneous Poisson equation (SM5C,[19] SMD[20]) for bulk electrostatics combined with empirical atomic surface tensions[21] that account for cavity formation, dispersion, and solvent structure effects and for shorter-range nonbulk electrostatic effects.

The GB models involve partial atomic charges, whose interaction with the solvent and with each other is dielectrically screened by the polarized solvent and descreened by other parts of the solute. Conventional GB models treat dielectric descreening effects in terms of the so-called Born radii of individual atoms in the solute molecule. The SM5.4, SM5.42, SM5.43, SM6, and SM8 models employ the Born radius based on the Coulomb field (CF) approximation of Still et al.[10] for the electric displacement induced by the partial atomic charge in a dielectric. In the case of the CF approximation, a charge-induced dipole interaction varies as r^{-4}, where r is the distance between the partial atomic charge and a volume element of the continuum solvent. Starting with the Kirkwood distributed monopole model[22] for biopolymer electrostatics, Grycuk has shown[23] that, when the individual partial atomic charges are asymmetrically situated in the molecule, i.e., located near the dielectric boundary rather than at the center of the molecular surface, one can apparently estimate the dielectric descreening more accurately by using a shorter-range function proportional to r^{-6} in evaluation of the polarization component of the free energy of solvation. The shorter-range bulk electrostatics can be approximated with a corrected formula for the Born radius suggested by Grycuk.[23] Tjong and Zhou have demonstrated[24,25] that, as measured against electrostatic energies calculated by solving the NPE for a set of 55 proteins in water as well as in low-dielectric media,[24,25] the GB method using the improved Born radius formula is more accurate than any of the GB/CF models tested in their study. We refer the reader to several other studies[26−37] dealing with Grycuk's method[23] or other approaches to improve the Coulomb field approximation of Still et al.[10]

The key element of the present article is the incorporation of the new descreening algorithm of Grycuk[23] into a self-consistent reaction field solvation model that should improve on conventional GB models with regard to predicting solvation free energies, liquid-phase molecular geometries, solute response properties, such as NMR chemical shifts in solution, and other molecular properties. The resulting model will be called Solvation Model 8 with Asymmetric Descreening (SM8AD) because the SM8AD model extends the earlier SM8 model based on the CF algorithm of Still et al.[10] to those solutes for which the conventional GB/CF approach might be particularly poor, for instance, in situations when one or more polar residues in the solute molecule lie near the dielectric boundary. Both SM8AD and SM8 use the universal cavity dispersion solvent structure (CDS) formalism[38] to account for the nonbulk electrostatic contributions to the solvation free energy, arising from interactions between the solute and the solvent molecules in the first solvation shell. The solute electronic relaxation and, therefore, the solvent-induced changes in the atomic charges result from all the solute−solvent bulk electrostatic interactions and —

in the case of clustered ions — from solute−cluster molecule interactions, but the CDS terms are added post-SCF and do not affect the solute charge distribution. 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 uncertainties in the precise definition of a solute cavity. The SM8AD model is a universal continuum 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 SM8AD model has been tested against the earlier SM8 and SMD models for 2 892 solvation data. In contrast to SM8AD and SM8, both of which employ GB approximation for bulk electrostatics and represent the solute molecule as a collection of partial atomic charges in a cavity, the SMD model[20] 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 nonhomogeneous Poisson equation by the Integral Equation Formalism Polarizable Continuum Model (IEFPCM) algorithm.[39−42]

2. Computational Details

The free energy of solvation is defined as the standard-state free energy of transfer from the gas phase to the condensed phase according to

$$\Delta G_S^o = \Delta E_E + \Delta E_N + G_P + G_{CDS} + \Delta G_{conc}^o \qquad (1)$$

where ΔE_E is the change in the solute's internal electronic (E) energy in moving from the gas phase to the liquid phase at the same geometry, ΔE_N is the change in the solute's internal energy due to changes in the equilibrium nuclear (N) positions in the solute that accompany the solvation process, G_P is the polarization free energy, and G_{CDS} is the component of the free energy that is nominally associated with cavitation, dispersion, and solvent structure. Because all calculations reported here are based on gas-phase geometries, the ΔE_N component is assumed to be zero in this article, although not in the model in general. 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}^o is zero.[43] It would be 1.89 kcal/mol if we instead used a gas-phase standard state of 1 atm.

Bulk Electrostatics Formalism. The electronic relaxation term ΔE_E and the polarization term G_P in eq 1 comprise the bulk electrostatic contribution ($\Delta G_{EP} = \Delta E_E + G_P$) to the solvation free energy. The bulk electrostatic contribution is calculated from a self-consistent molecular orbital calculation,[13] where the generalized Born approximation[5−11] is used to calculate the polarization component G_P according to

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right) \sum_{k,k'}^{atoms} q_k \gamma_{kk'} q_{k'} \qquad (2)$$

In the above equation, the summations go over atoms k in the solute, ε is the dielectric constant of the solvent, q_k is the partial atomic charge of atom k, and $\gamma_{kk'}$ is a Coulomb integral involving atoms k and k'. The Coulomb integrals $\gamma_{kk'}$ are calculated according to ref 10:

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

where $R_{kk'}$ is the distance between atoms k and k', and α_k is the effective Born radius of atom k, which is described below. In the above equation, d is an empirical constant that is set to the value of 3.7, which was found to be optimal in earlier work.[17] For atoms and monatomic ions, the GB result reduces to the original Born formula:[44]

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\frac{q^2}{\alpha} \qquad (4)$$

which is the exact classical result for the case where the solute is a conducting sphere of radius α, the charge q being located in the center of the sphere.

The effective Born radius of atom k in eq 3 can be expressed using the assumption that the electric displacement field induced by the charge q_k is a Coulomb field. In this case the charge-induced dipole interaction (G_{P_k}) varies as r^{-4}, according to the formula:

$$G_{P_k} = -\frac{q_k^2}{2}\left(1 - \frac{1}{\varepsilon}\right)\int_{\rho_k}^{\infty} \frac{dV}{4\pi r^4} \qquad (5)$$

where r is the distance between the partial atomic charge q_k and the volume element of the continuum solvent, ρ_k is the so-called Coulomb radius of atom k that defines the boundary between the solute cavity ($r < \rho_k$) and the bulk solvent ($r \geq \rho_k$). The Coulomb field approximation leads to the following formula for the Born radius:[45,46]

$$\alpha_k = \left(\frac{1}{R'} + \int_{\rho_{Z_k}}^{R'} \frac{A_k(r)}{4\pi r^4}dr\right)^{-1} \qquad (6)$$

In eq 6, R' is the radius of the sphere centered on atom k that completely engulfs all the other spheres centered on the other atoms of the solute, ρ_{Z_k} is the intrinsic Coulomb radius of atom k, which in the present study depends only on the atomic number Z_k, and $A_k(r)$ is the exposed (solvent accessible) area of a sphere of radius r that is centered on atom k. This area calculated here using the ASA algorithm[45] depends on the geometry of the solute and the radii of the spheres centered on all the other atoms in the solute.[47] If the analytical gradient of G_P with respect to the position R_a of an arbitrary atom a is desirable, the derivative of α_k must be taken according to[46]

$$\frac{\partial \alpha_k}{\partial R_a} = -\alpha_k^2 \int_{\rho_{Z_k}}^{R'} \frac{\partial A_k(r)}{\partial R_a} \frac{dr}{4\pi r^4} \qquad (7)$$

where the derivative of $A_k(r)$ is evaluated analytically using the ASA formulation.[45] The GB approximation that uses eq 6 for the Born radius will be called the generalized Born Coulomb field (GB/CF) approximation hereafter. The earlier SM8 model[18] uses the GB/CF approximation for bulk electrostatics.

Following Grycuk's formulation,[23] the performance of the GB approximation can be improved by replacement of eq 6 with an alternative functional form for the Born radius α_k, which is given as follows:

$$\alpha_k = \left(\frac{1}{R'^3} + \int_{\rho_{Z_k}}^{R'} \frac{3A_k(r)}{4\pi r^6}dr\right)^{-1/3} \qquad (8)$$

The GB approximation that uses eq 8 for the Born radius will be called the GB approximation with asymmetric descreening (GB/AD). The SM8AD model proposed in the present study uses the new GB/AD approximation for bulk electrostatics. The derivative of α_k with respect to the position R_a of an arbitrary atom a is given as follows:

$$\frac{\partial \alpha_k}{\partial R_a} = -\alpha_k^4 \int_{\rho_{Z_k}}^{R'} \frac{\partial A_k(r)}{\partial R_a} \frac{dr}{4\pi r^6} \qquad (9)$$

where the derivative of $A_k(r)$ is evaluated analytically using the ASA formulation.[45] The $\partial \alpha_k/\partial R_a$ derivative in eq 9 can be used in analytical computation of the $\partial G_P/\partial R_a$ gradient within the GB/AD formalism along with the other components of $\partial G_P/\partial R_a$ derived in earlier work.[46]

According to eq 2, the GB models are based on partial atomic charges, and therefore, their accuracy for a particular level of electronic structure theory depends on whether meaningful partial charges can be computed for that theoretical level. Like the earlier SM8 model, the new SM8AD model is designed to employ class IV charge models, in particular, Charge Model 4 (CM4)[17] and Charge Model 4M (CM4M).[48] 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,[49] Löwdin,[50] or redistributed Löwdin[51] 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.[18] 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.

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

$$G_{CDS} = \sum_k^{atoms} \sigma_k A_k(R, \{R_{Z_k} + r_s\}) + \sigma^{[M]} \sum_k^{atoms} A_k(R, \{R_{Z_k} + r_s\}) \qquad (10)$$

where σ_k is the atomic surface tension of atom k, A_k is the solvent accessible surface area (SASA)[52,53] of atom k, and $\sigma^{[M]}$ is the molecular surface tension. The SASA depends on the geometry R, the set $\{R_Z\}$ 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 according to the ASA algorithm.[45] The van der Waals radii used in eq 10 for the SASA calculation are not the same as the intrinsic Coulomb radii used in eqs 6−9 for solution of the bulk electrostatic problem. In fact, in eq 10 we use the values of R_Z fixed at Bondi's values[54] and the value of r_s fixed at the value[16] of 0.4 Å, whereas the intrinsic Coulomb radii used in eqs 6−9 have been optimized according to the algorithm that will be described later in the article.

The atomic surface tensions are given by

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{k'}^{\text{atoms}} \tilde{\sigma}_{Z_k Z_{k'}} T_k(\{Z_{k'}, R_{kk'}\}) \quad (11)$$

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

The atomic surface tensions in eq 11 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 as follows:

$$\tilde{\sigma}_i = \tilde{\sigma}_i^{[n]} n + \tilde{\sigma}_i^{[\alpha]} \alpha + \tilde{\sigma}_i^{[\beta]} \beta \quad (12)$$

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[55−58] hydrogen bond acidity parameter of the solvent (which Abraham denotes as $\Sigma\alpha_2$), β is Abraham's hydrogen bond basicity parameter of the solvent (which Abraham denotes as $\Sigma\beta_2$), and $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, and $\tilde{\sigma}_i^{[\beta]}$ are empirical parameters that depend on i.

The molecular surface tension in eq 10 is also a function of solvent descriptors, and it is given by

$$\sigma^{[M]} = \tilde{\sigma}^{[\gamma]}(\gamma/\gamma_\text{o}) + \tilde{\sigma}^{[\phi^2]}\phi^2 + \tilde{\sigma}^{[\psi^2]}\psi^2 + \tilde{\sigma}^{[\beta^2]}\beta^2 \quad (13)$$

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 = 1.43932 cal mol^{-1} Å^{-2}), and $\gamma_\text{o} = 1$ cal mol^{-1} Å^{-2}, ϕ^2 is the square of the fraction of non-hydrogenic atoms in the solvent molecule that are aromatic carbon atoms (carbon aromaticity), ψ^2 is the square of the fraction of non-hydrogenic atoms in the solvent molecule 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. According to eq 10, the molecular surface tension is multiplied by the total SASA of the given solute. The latter is equal to the sum of the SASAs of each of the individual atoms in the solute.

The SM8AD model may be applied to any medium for which the relevant macroscopic descriptors such as dielectric constant, refractive index, bulk surface tension, and acidity and basicity parameters are either known or may be estimated. Because water is so important as a solvent, there are advantages to using a less general and more specifically parametrized solvation model for this solvent. Therefore, water is treated as a special solvent that is given its own set of surface tension coefficients, so that eqs 12 and 13 are not used for water, and the molecular surface tension $\sigma^{[M]}$ in eq 10 is set to equal zero. Thus, in the case when the model is employed to compute solvation free energies in aqueous solvent, the parameters $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ used in eq 11 to obtain σ_k are simply numbers that do not depend on solvent descriptors.

SM8AD Training Set. The SM8AD training set is part of the Minnesota Solvation Database[59] and it contains 2 892 experimental solvation data for 233 ionic and 482 neutral 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 SM8AD primary testing set contains six data subsets:

(i) 274 aqueous free energies of solvation for 274 neutral compounds;[17,18,20]

(ii) 71 aqueous free energies of solvation for an additional 71 neutral compounds;[60,61]

(iii) 2 072 free energies of solvation in 90 nonaqueous solvents for 276 neutral solutes (232 of the 276 solutes are also included in the set of 274 aqueous solutes, and 44 solutes are additional);[18,20]

(iv) 143 transfer free energies between water and 15 organic solvents for an additional 93 neutral solutes;[18,20]

(v) 112 aqueous free energies of solvation for 112 selectively clustered singly charged ions (there are 81 unclustered ions and 31 clustered ions);[17,18,20,62]

(vi) 220 free energies of solvation in acetonitrile, dimethyl sulfoxide (DMSO), and methanol for 166 singly charged ions (45 ions out of 166 are also included in the set of 81 unclustered aqueous ions, and 121 ions are additional).[18,20,63]

Several technical points should be mentioned here. The SM8AD training set is similar to those used in the parametrization of the SM8[18] and SMD models,[20] except for 71 new solutes added to the previously used set of 274 aqueous neutrals.[18,20] The addition of the 71 compounds, many of which are compounds of complex functionality (for instance, agricultural pesticides), is essential in extending the applicability of the SM8AD model to the classes of compounds which are poorly represented in the training sets of many solvation models (for instance, compounds with oxidized sulfur and phosphorus functionalities). The subset of 71 aqueous solutes includes 13 compounds from the set of 17 compounds described in ref 60 and 58 compounds from the set of 63 compounds presented in the SAMPL1 challenge[61] organized by Openeye Software (February, 2008). The remaining four of the 17 solutes[60] (benzylbromide, benzylchloride, diethyl sulfide, and 1,4-dioxane) were already included in the set of 274 aqueous solutes. Two compounds out of the 63 SAMPL1 solutes,[61] in particular cup08042 and cup08062, were discarded from consideration here for the reason explained in ref 64, and the other three compounds

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2451**

in SAMPL1 but not in subset ii (dichlobenil, fenuron, methyl parathion) were already among the 274 aqueous solutes of subset i. The SM8AD training set has five fewer data than the SM8AD primary testing set because subset ii includes five sulfonylureas (bensulfuron methyl, chlorimuron ethyl, metsulfuron methyl, sulfometuron methyl, thifensulfuron), which were used for testing the SM8AD model but not used in the SM8AD parametrization because we considered the experimental solvation free energy targets assigned for these compounds to be suspect, see ref 64 for more detail. Thus, for the SM8AD parametrization, we use only 66 out of 71 additional aqueous solutes.

The 143 transfer free energies associated with transferring the solute from aqueous solution to an organic solvent were determined directly from the corresponding experimental partition coefficients. The 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.

The single-ion solvation free energies were evaluated in previous work[17,63] based on the corresponding thermochemical cycle that relates the solvation free energy of the cation BH^+ or the anion A^- to the gas-phase basicity of the base B or the gas-phase acidity of the acid AH using the reference solvation free energy of the proton.[18,20] For the 1:1 M standard-state free energies of solvation for the proton in acetonitrile, DMSO, methanol, and water, we use −260.2, −273.3, −263.5, and −265.9 kcal/mol, respectively.[62,63] 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 an unclustered form). The criteria for whether to cluster an ion are explained elsewhere.[17] In all cases, the clusters were represented by a single, lowest-energy conformation.

The estimated average uncertainty for experimental free energies of solvation and transfer free energies of neutral solutes in subsets of 274, 2072, and 143 data is 0.2 kcal/mol.[17,60] The uncertainty for free energies of solvation for a set of 71 aqueous data is about 1 kcal/mol on average in the range of 0.1−1.93 kcal/mol.[61] The estimated average uncertainty for solvation free energies of ionic solutes is 3 kcal/mol.[17]

Table 1 lists 92 solvents including water used in the SM8AD parametrization. The corresponding values of solvent descriptors such as dielectric constant, refractive index, bulk (macroscopic) surface tension, and acidity and basicity parameters were taken from the Minnesota Solvent Descriptor Database[65] (these values are given in the Supporting Information). Experimental values for the 2 892 solvation free energies are also given in the Supporting Information. All computed solvation free energies in this study are based on rigid, gas-phase geometries. The molecular geometries of all unclustered neutral and ionic solutes were optimized at the mPW1PW[66]/MIDI![67,68] level of electronic structure theory.[59] The molecular geometries of aqueous clustered ions were optimized at the B97−1[69]/MG3S[70] level of theory.[17]

Table 1. Solvents Used in the SM8AD Training Set[a]

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

[a] All solvents except methanol have data for free energies of solvation for neutral solutes. The asterisk denotes the solvents which have data for free energies of solvation for ionic solutes. The names of 15 solvents for which we used solvent−water transfer free energies of neutral solutes are italicized.

SM8AD Parametrization. As for the earlier SM8 model,[18] the SM8AD parametrization effort is focused on two types of parameters: (i) the intrinsic Coulomb radii used for construction of the cavities for the bulk electrostatic calculation; and (ii) the atomic surface tensions $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ in eq 11 and the parameters $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, $\tilde{\sigma}^{[\beta^2]}$ in eq 13 used for the nonbulk electrostatic calculation within the CDS formalism.

We recall here that there is no thermodynamically unique way to separate the electrostatic contribution to the free energy of solvation from the nonelectrostatic one because only their sum is a state function and physical observable.[71−73] Therefore, continuum solvation models differ from one another in the way in which the electrostatic and nonelectrostatic components are defined. It is also widely recognized that the electrostatic terms may depend strongly on the model radii.[12,74−77] Keeping in mind that the magnitudes of solvation free energies of ions are much larger than those of neutral solutes and are dominated by large electrostatic contributions, an optimization of intrinsic atomic Coulomb radii to provide accurate solvation free energies of ions is a reasonable way to determine these parameters. If one assumes that the solute cavity is charge independent, then by using the same radii one might also achieve a reasonable estimation of the bulk electrostatics for cases where electrostatic and nonelectrostatic terms are comparable. The nonelectrostatic terms can then be defined as the difference between the experimentally available and path-independent total free

2452 *J. Chem. Theory Comput., Vol. 5, No. 9, 2009*

Marenich et al.

Table 2. Intrinsic Coulomb Radii (Å) of Various Models and Bondi's van der Waals Radii (Å)

atom	Z	SM8AD(aq)[a]	SM8AD (DMSO)[a]	SM8(aq)[b]	SM8 (DMSO)[b]	SMD(aq)[c]	SMD (DMSO)[c]	Bondi[d]
H	1	1.02	0.80	1.02	0.80	1.20	1.20	1.20
C	6	1.75	1.75	1.57	1.57	1.85	1.85	1.70
N	7	1.94	1.94	1.61	1.61	1.89	1.89	1.55
O	8	1.52	2.29	1.52	2.18	1.52	2.29	1.52
F	9	1.68	1.68	1.47	2.63	1.73	1.73	1.47
Si	14	2.47	2.47	2.10	2.10	2.47	2.47	2.10
P	15	2.12	2.12	1.80	2.13	2.12	2.12	1.80
S	16	2.16	2.16	2.12	2.45	2.49	2.49	1.80
Cl	17	2.40	2.40	2.02	2.63	2.38	2.38	1.75
Br	35	2.62	2.62	2.60	2.85	3.06	3.06	1.85

[a] The SM8AD radii for H and O are defined as a function of Abraham's hydrogen bond acidity parameter (α) for a given solvent according to eqs 14 and 15. We adopted the SMD values[20] for the SM8AD radii for Si and P. [b] The intrinsic Coulomb radii used by the SM8 model for any solute in water and DMSO.[18] [c] The intrinsic Coulomb radii used by the SMD model for any solute in water and DMSO.[20] [d] Bondi's values of van der Waals radii.[54]

energy of solvation and the modeled electrostatic contribution. We use this approach in the present study.

As in earlier work,[18] we optimize the SM8AD radii in calculations only on ions, then fix these parameters and optimize the nonbulk electrostatic term on data for neutrals. The optimization of SM8AD radii was done by minimizing the sum of mean squared errors calculated over 332 data points corresponding to 220 ions in acetonitrile, dimethyl sulfoxide, methanol and 112 selectively clustered ions in water. The ΔG_{EP} values used in the SM8AD optimization of radii were calculated with a locally modified version of Gaussian 03[78] called the Minnesota Gaussian Solvation Module (MN-GSM),[79] the M05-2X density functional,[80] and the 6-31G* basis set.[81] For simplicity, we did not use basis sets with diffuse functions in the optimization of radii. Although diffuse functions are important for calculations on small gas-phase anions, especially for atoms and diatomic molecules, they are less important for large anions and in solution where diffuse charge clouds of gas-phase anions are contracted (although perhaps only a little) by solvation effects.

After a testing of various optimization schemes we have found that the hydrogen and oxygen radii strongly depend on the solvent's value of Abraham's hydrogen bond acidity parameter α (the parameter that Abraham calls $\Sigma\alpha_2$),[55−58] whereas the radii for other elements do not demonstrate such a dependency. To approximate the hydrogen radius we adopted the scheme previously elaborated for the SM8 model:[18]

$$\rho = \begin{cases} 1.02 & \alpha \geq 0.43 \\ 1.02 - 0.52\,(0.43 - \alpha) & \alpha < 0.43 \end{cases} \quad (14)$$

where α is the solvent's value of Abraham's hydrogen bond acidity parameter. For the oxygen radius we adopted the SMD dependence of the radius on the α parameter:[20]

$$\rho = \begin{cases} 1.52 & \alpha \geq 0.43 \\ 1.52 + 1.8\,(0.43 - \alpha) & \alpha < 0.43 \end{cases} \quad (15)$$

Using eqs 14 and 15 to define the hydrogen and oxygen radii, we optimized the radii for the remaining elements (C, N, F, S, Cl, Br) by taking them to be independent of the solvent. Since the training set of 332 ions used in the SM8AD optimization does not include any solute containing silicon

or phosphorus we have opted to fix the SM8AD radii for Si and P at their SMD values.[20] Table 2 lists the optimized values of SM8AD intrinsic Coulomb radii given for water ($\alpha = 0.82$) and DMSO ($\alpha = 0$) compared to the values of the radii used by our most recent solvation models SM8[18] and SMD[20] and to the values of Bondi.[54] This set includes the radii for H, C, N, O, F, Si, P, S, Cl, and Br. For any other atom, the SM8AD model can use the van der Waals radii of Bondi[54] and Mantina et al.[82] for those atoms for which they defined radii; in cases where the atomic radius is not given in those papers, a radius of 2.0 Å is used.

The cavity dispersion solvent structure term (eq 1) associated with nonbulk electrostatic effects is parametrized by means of the atomic surface tension coefficients $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, $\tilde{\sigma}_i^{[\beta]}$ (eq 12) and the molecular surface tension coefficients $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ (eq 13) for nonaqueous solvents and the atomic surface tension coefficients $\tilde{\sigma}_i$ (eq 11) for water. The optimization of the sigma parameters involves a minimization of the following error function:

$$\chi = \sum_{J=1}^{2\,555} \left[\Delta G_S^o(\text{expt.}, J) - \frac{1}{26} \sum_{j=1}^{26} \Delta G_{EP}(j, J) - G_{CDS}(J) \right]^2 \quad (16)$$

where the J-summation runs over all data points in the neutral training set, including 2 412 solvation free energies and 143 transfer free energies, the j-summation runs over all levels of electronic structure theory used in the parametrization, $\Delta G_S^o(\text{expt.}, J)$ is the experimental standard-state solvation or transfer free energy, $\Delta G_{EP}(j, J)$ is the bulk electrostatic energy computed for a given theoretical level (or bulk electrostatic contribution to a transfer free energy), and $\Delta G_{CDS}(J)$ is the nonbulk electrostatic energy defined by eqs 10−13 (or the corresponding contribution to a transfer free energy). The 2 412 data points do not include the five sulfonylureas in water mentioned before, which were used in the SM8AD testing but not in the SM8AD parametrization. The CDS term was parametrized by averaging over 26 combinations of charge models, electronic structure methods, and basis sets as listed in Table 3. Namely, we used two charge models (CM4[17] or CM4M[48]), nine density functionals (M05-2X,[80] M05,[80,83] M06-2X,[84,85] M06,[84,85] M06-HF,[85,86] M06-L,[85,87] mPW-PW,[66] mPW1PW,[66] and B3LYP[88−91]), the Hartree−Fock (HF) method, and nine basis sets for which the charge models

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2453**

Table 3. Twenty Six Combinations of Charge Models, Electronic Structure Levels, and Basis Sets Used in the SM8AD Parametrization

charge model	functional	basis set
CM4	M05-2X	MIDI!
CM4	M05-2X	MIDI!6D
CM4	M05-2X	6-31G*
CM4	M05-2X	6-31+G*
CM4	M05-2X	6-31+G**
CM4	M05-2X	6-31G**
CM4	M05-2X	cc-pVDZ
CM4	M05-2X	DZVP
CM4	M05-2X	6-31B*
CM4M	M06-2X	MIDI!
CM4M	M06-2X	MIDI!6D
CM4M	M06-2X	6-31G*
CM4M	M06-2X	6-31+G*
CM4M	M06-2X	6-31+G**
CM4M	M06-2X	6-31G**
CM4M	M06-2X	cc-pVDZ
CM4M	M06-2X	DZVP
CM4M	M06-2X	6-31B*
CM4	M05	6-31G*
CM4M	M06	6-31G*
CM4M	M06-HF	6-31G*
CM4M	M06-L	6-31G*
CM4	mPWPW	6-31G*
CM4	mPW1PW	6-31G*
CM4	B3LYP	6-31G*
CM4	HF	6-31G*

Table 4. Nine Additional Combinations of Charge Models, Electronic Structure Levels, and Basis Sets Used in the SM8AD Testing but not Used in the SM8AD Parametrization[a]

charge model	electronic structure level	basis set
CM4	M05-2X	6-31B**
CM4M	M06-2X	6-31B**
CM4	M05-2X	cc-pVTZ
LPA	M05-2X	6-31G*
RLPA	M05-2X	6-31+G**
CM2	AM1	n.a.
CM3	AM1	n.a.
CM2	PM3	n.a.
CM3	PM3	n.a.

[a] The n.a. denotes not applicable.

are available (MIDI![67,68] MIDI!6D,[67,68] 6-31G*,[81] 6-31+G*,[81] 6-31+G**,[81] 6-31G**,[81] cc-pVDZ,[92] DZVP,[93] or 6-31B*[94]). We used the CM4M charge model only with the M06 suite[85] of density functionals for which this model was designed,[48] and we use the CM4 model with any other density functional or with the Hartree−Fock method.

Table 4 lists nine additional levels of theory that were not used in parametrization, but that are tested in the present study. In particular, we tested the SM8AD model using partial atomic charges obtained from Löwdin population analysis (LPA)[50] or redistributed LPA (RLPA),[51] the 6-31B**[94] and cc-pVTZ[92] basis sets, and the semiempirical models Austin Model 1 (AM1)[95] and Parametrized Model 3 (PM3)[96] combined with Charge Model 2 (CM2)[97] and Charge Model 3 (CM3).[98]

In the present study we have employed essentially the same strategy for optimizing the sigma coefficients $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, $\tilde{\sigma}_i^{[\beta]}$, $\tilde{\sigma}_i^{[water]}$, $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$, as we did in our earlier

Table 5. Atomic Surface Tension Parameters (cal mol^{-1} Å^{-2}) for SM8AD that Depend on Atomic Numbers[a]

i	$\tilde{\sigma}_i^{[water]}$	$\tilde{\sigma}_i^{[n]}$	$\tilde{\sigma}_i^{[\alpha]}$	$\tilde{\sigma}_i^{[\beta]}$
H	32.74	25.35		
C	65.00	32.05	146.39	
H, C	−41.80	−69.24		
C, C	−50.56	−44.81	−89.04	
O	−79.20	−15.09		−41.39
H, O	−54.88	−84.48		−160.75
O, C	183.85		276.59	
O, O	76.58			
N		44.96	−100.55	
H, N	−111.09	−93.29		
C, N	41.77	−73.49	174.33	
N, C	−57.40		−82.30	
O, N	176.32			136.00
F	27.28			
Cl	−3.36	−23.09		
Br	−13.16	−31.68		
S	−20.49	−34.96		
O, P	151.00		283.85	
O, S	277.16			
S, P	76.43			
Si		−78.22		

[a] Any possible atomic number-dependent surface tension parameter that is not in this table is set equal to zero in SM8AD.

Table 6. Molecular Surface Tension Parameters (cal mol^{-1} Å^{-2}) for SM8AD that Do Not Depend on Atomic Numbers

$\tilde{\sigma}^{[\gamma]}$	0.19
$\tilde{\sigma}^{[\phi^2]}$	−2.71
$\tilde{\sigma}^{[\psi^2]}$	−8.25
$\tilde{\sigma}^{[\beta^2]}$	2.10

work,[18,20] with one exception. In the previous studies,[18,20] we optimized these parameters separately for atoms involving at most H, C, N, and O, atoms involving F, S, Cl, and Br, and atoms involving Si and P. In the present study, we have abandoned this scheme, and we optimized all the sigma coefficients simultaneously. The final set of SM8AD surface tension coefficients is listed in Tables 5 and 6. The SM8AD model uses 46 nonzero surface tension parameters compared to the 54 parameters that are used by SM8. Functional forms for atomic surface tensions used by SM8AD are given in the Supporting Information.

3. Results

Tables 7 and 8 show the mean signed errors (MSE) and the mean unsigned errors (MUE) in 2 892 solvation energies calculated by SM8AD, SM8, and SMD used in combination with five selected electronic structure methods (M05-2X, M06-2X, mPW1PW, B3LYP, HF) and four basis sets (MIDI!6D, 6-31G*, 6-31+G**, 6-31G**). The mean signed and unsigned errors as well as the root mean squared errors over all theoretical levels used in the present study are given in the Supporting Information. Tables 9−12 show in more detail the errors in SM8AD solvation energies calculated using M05-2X/6-31G*. Table 9 gives a breakdown of the errors in calculated aqueous solvation free energies for neutrals by solute class. In Tables 10 and 11, the errors are broken down by solute class for calculated solvation free energies of neutral solutes in nonaqueous solvents and for calculated transfer free energies, respectively. Table 12 gives

Table 7. Mean Signed Errors (kcal/mol) in the Free Energies of Solvation Calculated using SM8AD, SM8, and SMD[a]

salvation model	charge model	functional	basis set	2 560 neutral data					332 ionic data	
				aqueous data			nonaqueous data	transfer energies	aqueous data	nonaqueous data
				274	66[b]	5[c]	2 072	143	112	220
SM8AD	CM4	M05-2X	MIDI!6D	0.09	0.78	−2.00	0.04	−0.09	0.62	−1.69
SM8	CM4	M05-2X	MIDI!6D	0.28	0.97	−8.26	0.14	−0.14	1.55	−1.59
SMD		M05-2X	MIDI!6D	0.56	1.60	−4.35	0.34	−0.48	4.59	0.95
SM8AD	CM4	M05-2X	6-31G*	−0.02	0.40	−3.02	−0.02	0.04	−0.18	−2.36
SM8	CM4	M05-2X	6-31G*	0.20	0.69	−8.96	0.08	−0.08	0.71	−2.31
SMD		M05-2X	6-31G*	−0.06	0.40	−8.64	−0.06	−0.02	3.74	0.42
SM8AD	CM4	M05-2X	6-31+G**	0.00	0.03	−6.05	−0.01	−0.22	0.22	−2.62
SM8	CM4	M05-2X	6-31+G**	0.20	0.29	−12.10	0.07	−0.21	1.01	−2.61
SMD		M05-2X	6-31+G**	−0.73	−0.80	−10.75	−0.50	0.24	4.32	0.60
SM8AD	CM4	M05-2X	6-31G**	0.03	0.45	−2.19	−0.01	−0.05	−0.16	−2.30
SM8	CM4	M05-2X	6-31G**	0.24	0.70	−8.38	0.09	−0.13	0.66	−2.30
SMD		M05-2X	6-31G**	−0.12	0.24	−8.93	−0.09	0.00	3.79	0.38
SM8AD	CM4M	M06-2X	6-31G*	−0.33	−0.30	−5.50	−0.16	0.36	−0.32	−2.44
SM8	CM4M	M06-2X	6-31G*	−0.11	0.12	−10.77	−0.08	0.19	0.62	−2.37
SMD		M06-2X	6-31G*	0.22	0.87	−7.43	0.10	−0.21	4.08	0.78
SM8AD	CM4	mPW1PW	6-31G*	−0.30	−0.34	−5.43	−0.14	0.29	−0.23	−2.24
SM8	CM4	mPW1PW	6-31G*	−0.08	0.10	−10.73	−0.06	0.13	0.71	−2.18
SMD		mPW1PW	6-31G*	0.23	0.94	−6.99	0.10	−0.24	4.25	0.92
SM8AD	CM4	B3LYP	6-31G*	−0.12	−0.08	−3.68	−0.05	0.26	0.04	−1.94
SM8	CM4	B3LYP	6-31G*	0.12	0.34	−9.14	0.07	0.10	0.98	−1.85
SMD		B3LYP	6-31G*	0.59	1.38	−6.13	0.31	−0.38	4.72	1.42
SM8AD	CM4	HF	6-31G*	0.03	0.18	−3.49	−0.01	0.05	−0.57	−2.90
SM8	CM4	HF	6-31G*	0.30	0.51	−9.29	0.14	−0.05	0.36	−2.79
SMD		HF	6-31G*	−0.71	−1.54	−14.14	−0.33	0.71	2.67	−0.70

[a] Description of the SM8AD data set is given in Section 2. [b] Part of a subset of 71 aqueous data used in the SM8AD parametrization but not used in the SM8 and SMD parametrization. [c] Five sulfonylureas as part of a subset of 71 aqueous data not used in the parametrization of either model.

Table 8. Mean Unsigned Errors (kcal/mol) in the Free Energies of Solvation Calculated using SM8AD, SM8, and SMD[a]

salvation model	charge model	functional	basis set	2 560 neutral data					332 ionic data	
				aqueous data			nonaqueous data	transfer energies	aqueous data	nonaqueous data
				274	66[b]	5[c]	2 072	143	112	220
SM8AD	CM4	M05-2X	MIDI!6D	0.59	1.62	2.85	0.57	0.63	3.33	3.98
SM8	CM4	M05-2X	MIDI!6D	0.65	1.93	8.26	0.59	0.72	3.36	4.57
SMD		M05-2X	MIDI!6D	0.88	2.34	4.62	0.70	0.82	4.83	4.19
SM8AD	CM4	M05-2X	6-31G*	0.53	1.72	3.23	0.56	0.63	2.96	4.38
SM8	CM4	M05-2X	6-31G*	0.59	2.01	8.96	0.57	0.73	3.00	4.93
SMD		M05-2X	6-31G*	0.60	1.85	8.64	0.63	0.64	4.13	4.15
SM8AD	CM4	M05-2X	6-31+G**	0.67	2.41	6.05	0.61	0.80	2.50	4.90
SM8	CM4	M05-2X	6-31+G**	0.72	2.34	12.10	0.61	0.71	3.11	5.70
SMD		M05-2X	6-31+G**	0.96	1.97	10.75	0.79	0.68	4.64	4.08
SM8AD	CM4	M05-2X	6-31G**	0.53	1.74	2.71	0.56	0.59	2.90	4.38
SM8	CM4	M05-2X	6-31G**	0.59	2.01	8.38	0.57	0.68	2.94	4.95
SMD		M05-2X	6-31G**	0.62	1.84	8.93	0.64	0.64	4.15	4.13
SM8AD	CM4M	M06-2X	6-31G*	0.62	1.69	5.50	0.59	0.72	2.79	4.38
SM8	CM4M	M06-2X	6-31G*	0.57	1.91	10.77	0.58	0.76	2.92	4.94
SMD		M06-2X	6-31G*	0.62	1.94	7.43	0.63	0.69	4.41	4.13
SM8AD	CM4	mPW1PW	6-31G*	0.60	1.69	5.43	0.59	0.71	2.86	4.28
SM8	CM4	mPW1PW	6-31G*	0.56	1.96	10.73	0.58	0.77	2.96	4.84
SMD		mPW1PW	6-31G*	0.63	1.97	6.99	0.63	0.69	4.54	4.20
SM8AD	CM4	B3LYP	6-31G*	0.59	1.60	3.68	0.57	0.72	3.00	4.14
SM8	CM4	B3LYP	6-31G*	0.58	1.87	9.14	0.58	0.79	3.07	4.68
SMD		B3LYP	6-31G*	0.81	2.11	6.13	0.67	0.74	4.94	4.24
SM8AD	CM4	HF	6-31G*	0.57	1.59	3.52	0.56	0.65	3.01	4.82
SM8	CM4	HF	6-31G*	0.65	1.93	9.29	0.58	0.74	2.91	5.26
SMD		HF	6-31G*	0.92	2.29	14.14	0.73	0.97	3.39	4.50

[a] See footnote a of Table 7. [b] See footnote b of Table 7. [c] See footnote c of Table 7.

a breakdown of MSEs and MUEs in 332 solvation free energies for ions by solute class. The errors in the M05-2X/6-31G* solvation energies for neutrals broken down by solvent name are given in the Supporting Information.

The SM8AD and SM8 solvation energies based on DFT or the Hartree−Fock method were calculated with a locally modified version of Gaussian 03[78] called MN-GSM.[79] The SMD solvation energies were calculated with the GESOL

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2455**

Table 9. Mean Signed and Mean Unsigned Errors in Aqueous Solvation Free Energies Calculated using SM8AD and SM8 with CM4/M05-2X/6-31G* by Solute Class[a]

solute class	N	SM8AD MSE[b]	MUE[b]	SM8 MSE[b]	MUE[b]
		274 Data			
H_2, NH_3, H_2O, $(H_2O)_2$	4	−1.74	1.74	−2.06	2.06
unbranched alkanes	8	−0.81	0.81	−0.85	0.85
branched alkanes	5	−0.75	0.75	−0.77	0.77
cycloalkanes	5	−0.47	0.47	−0.70	0.70
alkenes	9	−0.37	0.41	−0.30	0.40
alkynes	5	0.20	0.24	0.43	0.43
arenes	8	0.06	0.23	−0.09	0.24
alcohols	12	0.52	0.53	0.59	0.59
phenols	4	0.46	0.52	0.60	0.60
ethers	12	0.13	0.43	0.54	0.61
aldehydes	6	0.19	0.29	0.16	0.27
ketones	12	−0.12	0.30	0.30	0.36
carboxylic acids	5	0.42	0.42	0.74	0.74
esters	13	−0.43	0.47	−0.03	0.16
peroxides	3	−0.16	0.28	0.14	0.14
bifunctional H, C, O compounds	5	0.49	0.59	0.79	0.79
aliphatic amines	15	0.13	0.63	0.17	0.61
anilines	7	0.37	0.37	0.54	0.54
aromatic N-heterocycles (1 N)	10	0.41	0.41	0.46	0.46
aromatic N-heterocycles (2 Ns)	3	−0.70	0.70	−0.11	0.77
nitriles	4	−0.43	0.43	0.89	0.89
hydrazines	3	0.47	0.95	0.19	0.88
bifunctional H, C, N compounds	3	0.32	0.34	0.44	0.61
amides	4	0.45	0.62	0.85	0.97
ureas	2	0.27	0.59	0.12	0.12
nitrohydrocarbons	7	−0.11	0.30	0.30	0.40
bifunctional H, C, N, O compounds	3	−0.35	0.35	0.18	0.18
fluoroalkanes	5	−0.31	0.42	−0.07	0.30
fluoroarene	1	−0.09	0.09	−0.05	0.05
chloroalkanes	13	−0.17	0.32	0.10	0.26
chloroalkenes	6	0.17	0.27	0.68	0.68
chloroarenes	8	−0.43	0.43	−0.28	0.28
bromoalkanes	9	−0.11	0.12	−0.21	0.21
bromoalkene	1	−0.08	0.08	0.04	0.04
bromoarenes	4	−0.23	0.23	−0.41	0.41
multihalogenhydrocarbons	12	−0.34	0.37	0.03	0.25
halogenated bifunctional compounds	9	0.30	0.95	1.33	1.43
thiols	4	0.73	0.73	0.63	0.63
sulfides	5	0.87	1.01	0.77	0.86
disulfides	2	0.44	0.44	0.12	0.12
sulfur heterocycle	1	0.50	0.50	0.30	0.30
halogenated sulfur compounds	2	−0.98	2.63	−0.13	1.78
phosphorus compounds	14	0.20	1.08	0.78	1.62
silicon compound	1	−0.49	0.49	−0.12	0.12
all data	274	−0.02	0.53	0.20	0.59
		66 Data			
compounds containing H, C, O	9	−0.70	1.06	0.30	0.70
compound containing H, C, N	1	1.76	1.76	0.84	0.84
compounds containing H, C, N, O	17	0.18	2.57	0.80	2.37
compound containing H, C, F	1	−0.34	0.34	0.39	0.39
compounds containing H, C, Cl	3	0.48	0.48	1.58	1.58
halogen compounds containing H, C, N, and/or O	15	1.24	1.85	2.15	2.43
sulfur compounds not containing P	9	0.04	1.33	−1.11	2.55
phosphorus compounds	11	0.74	1.57	0.09	1.89
all data	66	0.41	1.72	0.69	2.01
		5 Data (Sulfonylurea Subset)			
compounds containing H, C, N, O, S	4	−3.16	3.43	−9.22	9.22
compound containing H, C, N, O, S, Cl	1	−2.44	2.44	−7.93	7.93
all data	5	−3.02	3.23	−8.96	8.96

[a] N is the number of data in a given solute class. [b] MSE and MUE are in kcal/mol.

program[99] that employs the external option of Gaussian 03. All solvation calculations using AM1 and PM3 were carried out with a locally modified version of GAMESS[100] called GAMESSPLUS.[101]

4. Discussion of Model Performance

First, we discuss the performance of SM8AD for predicting the solvation free energies of neutral solutes in aqueous

Table 10. Mean Signed and Mean Unsigned Errors in Nonaqueous Solvation Free Energies Calculated using SM8AD and SM8 with CM4/M05-2X/6-31G* by Solute Class[a]

solute class	N	SM8AD MSE[c]	MUE[c]	SM8 MSE[c]	MUE[c]
H_2, NH_3, H_2O	29	−1.72	1.73	−1.70	1.72
unbranched alkanes	85	0.19	0.43	0.23	0.45
branched alkanes	7	0.16	0.33	0.33	0.41
cycloalkanes	13	−0.12	0.35	−0.32	0.45
alkenes	18	−0.22	0.50	0.11	0.41
alkynes	9	−0.13	0.60	0.42	0.51
arenes	134	−0.43	0.54	−0.29	0.46
alcohols	272	−0.04	0.47	0.01	0.38
phenols	109	0.24	0.45	0.49	0.62
ethers	87	0.42	0.66	0.29	0.69
aldehydes	32	−0.01	0.73	−0.10	0.60
ketones	195	−0.51	0.57	−0.51	0.58
carboxylic acids	120	0.32	0.55	0.48	0.64
esters, including lactones [b]	243	0.12	0.35	0.27	0.45
peroxides	17	0.45	1.29	−0.07	0.58
bifunctional H, C, O compounds	24	0.81	1.38	1.01	1.33
aliphatic amines	154	0.18	0.38	0.19	0.43
anilines	61	−0.28	0.39	0.10	0.38
aromatic N-heterocycles (1 N)	52	0.00	0.58	−0.06	0.62
aromatic N-heterocycles (2 Ns)	9	0.63	0.82	0.68	0.82
nitriles	20	−0.33	0.59	0.18	0.46
hydrazines	5	0.83	1.19	0.80	1.27
bifunctional H, C, N compounds	2	−0.91	0.94	−0.74	0.79
amides	26	0.19	0.65	0.45	0.71
ureas	7	0.75	0.75	0.85	0.96
lactams	4	0.91	0.99	0.75	0.90
nitrohydrocarbons	86	−0.05	0.68	0.03	0.51
bifunctional H, C, N, O compounds	3	0.27	1.00	0.26	0.75
fluoroalkanes	5	−0.71	0.71	−0.12	0.62
fluoroarenes	11	−0.43	0.81	0.15	0.59
chloroalkanes	26	−0.61	0.61	−0.43	0.44
chloroalkenes	15	0.59	0.64	0.76	0.76
chloroarenes	31	−0.23	0.38	−0.06	0.33
bromoalkanes	21	−0.46	0.46	−0.67	0.68
bromoalkenes	2	−0.38	0.38	−0.16	0.16
bromoarenes	16	−0.40	0.54	−0.44	0.52
multihalogenhydrocarbons	14	−0.15	0.31	−0.26	0.36
halogenated bifunctional compounds	37	0.42	0.87	0.85	0.99
thiols	10	0.53	0.55	0.15	0.25
sulfides	13	−0.22	0.82	−0.30	0.91
disulfides	4	0.57	0.57	0.23	0.43
sulfur heterocycles	4	0.69	0.69	0.67	0.67
sulfoxide	1	−0.19	0.19	−0.41	0.41
phosphorus compounds	37	0.15	1.43	0.94	1.67
silicon compounds	2	0.97	0.97	1.61	1.61
all data	2 072	−0.02	0.56	0.08	0.57

[a] N is the number of data in a given solute class. [b] Five lactones and 238 other esters. [c] MSE and MUE are in kcal/mol.

solution. For the 274 neutral data in water that constitute subset i defined in Section 2, the MSE in the SM8AD free energies of solvation calculated using 26 theoretical levels listed in Table 2 is between −0.37 (M06-HF/6-31G*) and 0.58 kcal/mol (M05-2X/6-31B*). The corresponding MUE varies from 0.53 (M05-2X/6-31G*, M05-2X/6-31G**) to 0.87 kcal/mol (M05-2X/6-31B*). The MUE in the SM8 solvation free energies for the same data set varies from 0.55 (M06/6-31G*) to 1.04 kcal/mol (M05-2X/6-31B*), and the corresponding MUE in the SMD solvation free energies is between 0.60 (M05-2X/6-31G*) and 1.15 kcal/mol (M05-2X/6-31B*). For the 66 neutral data in water that constitute the training part of ii defined in Section 2, the MSE averaged over 26 theoretical levels used in these calculations is 0.1, 0.4, and 0.5 kcal/mol for SM8AD, SM8, and SMD, respec-

Table 11. Mean Signed and Mean Unsigned Errors in Transfer Free Energies between Water and Organic Solvents Calculated using SM8AD and SM8 with CM4/M05-2X/6-31G* by Solute Class[a]

		SM8AD		SM8	
solute class	N	MSE[b]	MUE[b]	MSE[b]	MUE[b]
lactones	10	0.40	0.84	0.06	0.89
aromatic N-heterocycles	6	−0.14	0.53	0.02	0.37
bifunctional H, C, N compounds	2	0.94	0.94	0.82	0.82
amides	13	−0.57	0.61	−1.05	1.05
ureas	11	−0.07	0.35	0.02	0.27
lactams	4	−1.35	1.35	−1.71	1.71
thymines and uracils	12	0.37	0.63	0.31	0.67
bifunctional H, C, N, O compounds	5	−0.15	0.81	−0.14	0.35
halogenated bifunctional compounds	39	0.32	0.64	0.38	0.67
sulfur compounds	19	0.10	0.37	−0.06	0.49
phosphorus compounds	9	−0.77	0.91	−1.36	1.38
silicon compounds	13	0.20	0.61	0.16	0.82
all data	143	0.04	0.63	−0.08	0.73

[a] N is the number of data in a given solute class. [b] MSE and MUE are in kcal/mol.

Table 12. Mean Signed and Mean Unsigned Errors in Ionic Solvation Free Energies Calculated using SM8AD and SM8 with CM4/M05-2X/6-31G*[a]

		SM8AD		SM8	
solute class	N	MSE[e]	MUE[e]	MSE[e]	MUE[e]
Acetonitrile					
H, C, N, O cations[b]	36	2.6	5.3	4.0	6.5
S cations[c]	3	9.1	9.1	16.1	16.1
All cations	39	3.1	5.6	4.9	7.2
H, C, N, O anions[b]	19	−3.2	3.2	−4.5	4.5
F, Cl, Br, S anions[c]	11	−4.2	4.3	−3.4	3.4
All anions	30	−3.6	3.6	−4.1	4.1
All ions	69	0.2	4.7	1.0	5.9
DMSO					
H, C, N, O cations[b]	4	0.1	0.7	−1.5	1.8
All cations	4	0.1	0.7	−1.5	1.8
H, C, N, O anions[b]	52	−4.2	5.0	−8.0	8.3
F, Cl, Br, S anions[c]	15	−5.8	5.8	−3.7	4.1
All anions	67	−4.6	5.2	−7.0	7.3
All ions	71	−4.3	4.9	−6.7	7.0
Methanol					
H, C, N, O cations[b]	26	−3.8	4.1	−1.3	2.1
Cl, Br cations[c]	3	−3.8	3.8	−1.0	1.0
All cations	29	−3.8	4.1	−1.2	2.0
H, C, N, O anions[b]	36	−2.5	3.4	−1.0	2.4
F, Cl, Br anions[c]	15	−1.7	2.9	−1.8	2.5
All anions	51	−2.3	3.3	−1.2	2.4
All ions	80	−2.8	3.6	−1.2	2.3
Water [d]					
H, C, N, O cations[b]	48	−2.1	3.0	0.4	2.4
Cl, S cations[c]	4	−0.7	2.7	1.5	2.5
All cations	52	−2.0	3.0	0.5	2.4
H, C, N, O anions[b]	43	1.5	3.0	1.3	3.7
F, Cl, Br, S anions[c]	17	1.3	2.8	0.0	3.0
All anions	60	1.4	3.0	0.9	3.5
All ions	112	−0.1	3.0	0.7	3.0

[a] N is the number of data in a given solute class. [b] Ions containing no elements other than H, C, N or O. [c] Ions containing any of the listed elements in addition to H, C, N or O. [d] 112 selectively clustered ions from the SM6 model training set as defined in ref 17. [e] MSE and MUE are in kcal/mol.

tively; the corresponding MUE is 1.9, 2.1, and 2.1 kcal/mol. Thus, SM8AD slightly outperforms SM8 and SMD on these data sets.

For five sulfonylureas (namely, bensulfuron methyl, chlorimuron ethyl, metsulfuron methyl, sulfometuron methyl,

thifensulfuron methyl) in water, the MSE averaged over 26 theoretical levels is equal to −4.2, −10.1, and −8.0 kcal/mol, respectively for SM8AD, SM8, and SMD, and the corresponding MUE is 4.6, 10.1, and 8.1 kcal/mol. These errors are rather large. However, one should keep in mind that Guthrie[61] assigned these five compounds the largest possible uncertainty (1.9 kcal/mol), meaning that the corresponding solubility and vapor pressure could not be found in the open source literature or that the primary vapor pressure temperature data were not available.[61] As the compounds are highly nonvolatile, their vapor pressures were measured by assessing slow mass loss from solid samples heated to rather high temperatures and then by extrapolating back to room temperature.[102] In addition to uncertainties associated with the extrapolation procedure, there is the possibility of thermal degradation of the sulfonylureas.[102] In view of the uncertainty in the experimental targets for these compounds, we did not include them in the parametrization of any of the solvation models tested in our study. Nevertheless, the SM8AD model is able to predict the free energies of solvation for such difficult compounds more accurately than any other tested model apparently due to more robust parametrization of its nonbulk electrostatic term with the use of a more diverse training set in spite of the fact that no sulfonylureas were used in the SM8AD parametrization. In this regard, we recall that the SM8AD training set has been extended by addition of 66 aqueous solutes (see Section 2) many of which are compounds of complex functionality, including compounds with oxidized sulfur that are not present in the SM8 and SMD training sets of aqueous neutrals.

Now we discuss the performance of SM8AD as compared to that of SM8 and SMD for predicting the free energies of solvation for neutral solutes in nonaqueous solvents. For 2 072 neutral data in 90 organic solvents, the MSE in the SM8AD free energies of solvation calculated using 26 theoretical levels listed in Table 2 is between −0.20 (M06-HF/6-31G*) and 0.31 kcal/mol (M05-2X/6-31B*). The corresponding MUE varies from 0.56 to 0.61 kcal/mol, only slightly depending on density functional and basis set. The MUE in the SM8 solvation free energies for the same data set varies from 0.57 to 0.69 kcal/mol, and the corresponding MUE in the SMD solvation free energies is between 0.63 (M05-2X/6-31G*) and 0.84 kcal/mol (M05-2X/6-31B*). For 143 transfer energies, the MSE averaged over 26 theoretical levels used in these calculations is 0.03, −0.04, and −0.13 kcal/mol for SM8AD, SM8, and SMD, respectively; the corresponding MUE is 0.71, 0.74, and 0.74 kcal/mol.

For 112 aqueous ions, the MSE in the SM8AD free energies of solvation calculated using 26 theoretical levels is between −0.59 (M06-HF/6-31G*) and 0.62 kcal/mol (M05-2X/MIDI!6D), and the corresponding MUE varies from 2.50 (M05-2X/6-31+G**) to 3.37 kcal/mol (M05-2X/MIDI!). The MSE in the SM8 free energies of solvation for the same data set is between 0.12 (M05-2X/DZVP) and 1.55 kcal/mol (M05-2X/MIDI!6D), and the corresponding MUE ranges from 2.75 (M06-HF/6-31G*) to 3.36 kcal/mol (M05-2X/MIDI!). The MSE in the SMD free energies of solvation for the same data set is between 2.67 (HF/6-31G*) and 5.17

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2457**

kcal/mol (mPWPW/6-31G*), and the corresponding MUE ranges from 3.39 (HF/6-31G*) to 5.34 kcal/mol (mPWPW/6-31G*). The MUE averaged over 26 theoretical levels used in the calculations for 112 aqueous ions is 2.9, 3.1, and 4.6 kcal/mol for SM8AD, SM8, and SMD, respectively. Thus, SM8AD produces on average slightly more accurate solvation energies for 112 aqueous ions tested in the present study.

For 220 ionic data in acetonitrile, dimethyl sulfoxide, and methanol, the MSE in the SM8AD free energies of solvation calculated using 26 theoretical levels is between −2.93 (M05-2X/DZVP) and −1.69 kcal/mol (M05-2X/MIDI!6D). The MSE in the SM8 free energies of solvation for the same data set is between −3.09 (M05-2X/DZVP) and −1.59 kcal/mol (M05-2X/MIDI!6D). The MSE in the SMD free energies of solvation for the same data set is between −0.70 (HF/6-31G*) and 1.86 kcal/mol (mPWPW/6-31G*). The MUE averaged over 26 theoretical levels used in the calculations for 220 ionic data in the three nonaqueous solvents 4.5, 5.1, and 4.2 kcal/mol for SM8AD, SM8, and SMD, respectively.

In general, the SM8AD model demonstrates satisfactory performance with any of the 26 theoretical levels listed in Table 2 for which it has been parametrized. The quality of the model does not substantially depend on the choice of density functional, and it can be recommended for the use with any density functional or the Hartree−Fock method. The SM8AD performance worsens when the model is used with the basis sets DZVP and 6-31B* and with basis sets that contain diffuse functions (6-31+G*, 6-31+G**). For instance, the MUE in the SM8AD free energies of solvation calculated for 274 neutral solutes in water using M05-2X with the CM4 charge model is 0.67 (6-31+G**), 0.69 (6-31+G*), 0.77 (DZVP), and 0.87 kcal/mol (6-31B*). The MUE in the SM8AD free energies of solvation calculated for the same data set using M06-2X with the CM4M charge model is 0.68 (6-31+G**), 0.69 (6-31+G*), 0.73 (DZVP), and 0.73 kcal/mol (6-31B*). The best performance for SM8AD can be achieved if it is used with MIDI!, MIDI!6D, 6-31G*, 6-31G**, and cc-pVDZ basis sets. Tables 9−12 give a breakdown of the errors in solvation calculations using SM8AD/CM4/M05-2X/6-31G* by solute class. For 274 neutral data in water, the MSE ranges from −1.74 to 0.87 kcal/mol (Table 9). For 2 072 neutral data in 90 organic solvents, the MSE ranges from −1.72 to 0.97 kcal/mol (Table 10). For 143 transfer energies, the MSE ranges from −1.35 to 0.94 kcal/mol (Table 11). For 124 cationic data (Table 12), the smallest error is observed for cations in DMSO (MSE = 0.1, MUE = 0.7 kcal/mol), and the largest error is observed for cations in acetonitrile (MSE = 3.1, MUE = 5.6 kcal/mol). The errors for the sulfur-containing cations in acetonitrile are somewhat large, but this discrepancy is mainly attributed to the H_3S^+ cation which is systematically (by ∼16 kcal/mol) undersolvated by the SM8AD model with respect to the experimental solvation free energy ΔG_s^o(expt.) = −100.2 kcal/mol. However, the SM8AD error for H_3S^+ is substantially smaller than the corresponding SM8 error. For 208 anionic data (Table 12), the smallest error is observed for anions in water (MSE = 1.4, MUE = 3.0 kcal/mol), and the largest error is observed for anions in DMSO (MSE = −4.6, MUE = 5.2 kcal/mol).

Other theoretical levels for which the SM8AD model has been tested include those listed in Table 3 (see the Supporting Information), though we did not use these theoretical levels in the SM8AD parametrization. The SM8AD model applied with M05-2X/6-31B** and the CM4 charge model to 274 aqueous solutes substantially undersolvates these compounds (MSE = 0.89 kcal/mol, MUE = 1.14 kcal/mol). The MUE in the free energies of solvation computed with SM8/CM4/M05-2X/6-31B** and SMD/M05-2X/6-31B** for the same data set are even larger, reaching 1.27 and 1.23 kcal/mol, respectively. The use of M06-2X/6-31B** instead of M05-2X/6-31B** slightly reduces these errors, resulting in 0.99, 1.13, and 0.93 kcal/mol for SM8AD, SM8, and SMD, respectively, the CM4M charge model being used with SM8AD and SM8. We also tested SM8AD and SM8 with the cc-pVTZ basis set for which we do not have a charge model; we used partial atomic charges from Löwdin population analysis (LPA) in this case. The two GB models used with cc-pVTZ perform poorly in comparison with the SMD model,[20] which is a density-based solvation model that does not require partial atomic charges (the MUEs for 274 aqueous data are equal to 2.46, 2.59, and 0.68 kcal/mol, respectively, for SM8AD, SM8, and SMD). The use of partial atomic charges obtained from Löwdin population analysis results in poor performance for SM8AD and SM8, even in the case of smaller basis sets such as 6-31G*. Indeed, the MUE in the SM8AD/LPA/6-31G* free energies of solvation calculated for 274 aqueous data is overly large (2.15 kcal/mol). Adding diffuse functions to the basis results in even larger errors in the SM8AD/LPA calculations, though the use of RLPA[51] instead of LPA can reduce the error (MUE = 1.46 kcal/mol for 274 aqueous data with SM8AD/RLPA/6-31+G**). As it was previously indicated,[18] the use of GB solvation models with partial atomic charges from population analyses (class II charges) results in less accurate solvation energies than with CM4 charges and other comparably reliable class IV charges.

In addition, we tested the SM8AD model using the semiempirical electronic structure methods AM1 and PM3 combined with the CM2[97] and CM3[98] partial atomic charges (see the Supporting Information). The charge models CM2[97] and CM3[98] were specifically parametrized for the use with AM1 and PM3. For 274 aqueous data, the MSE in the SM8AD free energies of solvation calculated with AM1/CM2, AM1/CM3, PM3/CM2, and PM3/CM3 varies from −1.75 to −2.20 kcal/mol. For 66 aqueous data, the corresponding MSE varies from −5.2 to −4.2 kcal/mol. For five sulfonylureas in water, the MSE varies from −18.9 to −15.1 kcal/mol. Similar errors were obtained for SM8/AM1/CM2, SM8/AM1/CM3, SM8/PM3/CM2, SM8/PM3/CM3 as well as for SMD/AM1 and SMD/PM3. These errors are substantially larger than those obtained, for instance, with SM8AD/CM4/M05-2X/6-31G*. Unfortunately, the applicability of the SM8AD model parameters developed in the present study for the use with any density functional or the Hartree−Fock method cannot be extended to use with AM1 and PM3, and the SM8AD model would require a special parametrization of its Coulomb radii and atomic surface tension coefficients to be used with AM1 and PM3. The same is true for SM8

and SMD. The earlier continuum models SM5.42/AM1 and SM5.42/PM3, which were parametrized for the use with AM1 and PM3, demonstrate good performance on 274 aqueous neutral data (MUE = 0.61 kcal/mol) and on 2 072 nonaqueous neutral data (MUE = 0.53−0.54 kcal/mol), and somewhat poorer performance on 112 aqueous ions (MUE = 4.2−4.9 kcal/mol) and on 220 nonaqueous ions (MUE = 6.7−6.9 kcal/mol).

Overall, the SM8AD errors are typically smaller than those of SM8 and SMD in many cases when all the three models perform well. However, the difference in the SM8AD, SM8, and SMD errors is usually smaller than the estimated uncertainty of the corresponding experimental targets (the uncertainty ranges from 0.1 to 1.9 kcal/mol for neutrals,[17,60,61] and it is about 3 kcal/mol for ions[17]), therefore, one can assume that all the three models do equally well. This indicates the fact that the CDS formalism used in the parametrization of SM8AD, SM8, and SMD is able to account for most of the nonelectrostatic solvation effects as well as for the deviations of the electrostatics from the assumed bulk model (which is different in all the models) due to the inexactness of the solvent permittivity model including the assumed values for intrinsic Coulomb radii, the uncertainties in the treatment of solute charge outside the solute cavity,[2,4,40,103] as in the case of SMD, and the inexactness of the solute charge model as in the case of SM8 and SM8AD.

In Table 13, we compare the free energies of solvation for 71 aqueous solutes calculated using SM8AD with those calculated using the IEFPCM algorithm[39−42] as implemented in Gaussian 03[78] and the Poisson−Boltzmann (PB) self-consistent reaction field solver as implemented in Jaguar.[104−106] For both electrostatic and nonelectrostatic contributions, we accept the defaults of these programs. Thus, the Gaussian 03 calculations include not only electrostatics but also cavitation, dispersion, and repulsion.[75] The molecular cavities in the IEFPCM/Gaussian 03 calculations were constructed using the united atom Hartree−Fock (UAHF) scheme[107] for atomic radii that is a recommended method for predicting solvation free energies with PCM according to the Gaussian 03 manual.[78] We have also tested another united atom scheme called UA0[108] that is a simplified united atom implementation based on the universal force field radii, available for the full periodic table.[109] The PB/Jaguar calculations employ atomic radii that depend on typing certain functional groups in a solute molecule.[105]

The MUE in the PB/Jaguar solvation free energies of 71 aqueous solutes (Table 13) is 1.5 times larger than the corresponding SM8AD error, whereas the MUE in the IEFPCM/UAHF/Gaussian 03 calculations for the same data set is twice as large. The solvation energies calculated using IEFPCM/UA0/Gaussian 03 are much less accurate (MSE = 9.9, MUE = 9.9, RMSE = 11.1 kcal/mol) and many of them are in disagreement with those calculated using IEFPCM/ UAHF. Table 14 shows individual contributions to the free energies of solvation such as polarization (G_P), electronic relaxation (G_E), cavitation (G_C), dispersion (G_D), and repulsion (G_R) calculated by IEFPCM/Gaussian 03 using UA0, UAHF, and Bondi's radii for four compounds selected out

Table 13. Standard-State Free Energies of Solvation (kcal/mol) Calculated for a Subset of 71 Aqueous Data using SM8AD, IEFPCM/Gaussian 03, and PB/Jaguar with M05-2X/6-31G*[a]

solute name[b]	exp	SM8AD	IEFPCM/UAHF Gaussian 03	PB Jaguar
1,1-diacetoxyethane	−5.0	−7.6	−3.5	−6.3
1,1-diethoxyethane	−3.3	−2.7	−2.7	−4.1
1,2-diethoxyethane	−3.5	−3.3	−4.2	−4.5
1,2-dinitroxypropane	−5.0	−1.6	−4.7	−5.5
1,4,5,8-tetraminoanthraquinone	−8.9	−17.3	−11.6	−18.3
1-amino-4-anilinoanthraquinone	−7.4	−11.0	−4.9	−11.7
1-amino-anthraquinone	−8.0	−9.7	−5.5	−9.1
2-butyl nitrate	−1.8	−0.1	−0.7	−1.6
4-amino-4′-nitroazobenzene	−11.2	−12.1	−8.2	−10.0
alachlor	−8.2	−5.5	−1.7	−8.4
aldicarb	−9.8	−7.6	−5.7	−9.1
ametryn	−7.7	−9.5	−7.7	−12.3
azinphos methyl	−10.0	−10.0	−4.7	−12.4
benefin	−3.5	−0.7	2.9	−2.6
bensulfuron	−17.2	−24.7	−21.0	−35.1
bis(2-chloroethyl) ether	−4.2	−3.1	−4.1	−4.2
bromacil	−9.7	−10.4	−5.8	−11.3
butyl nitrate	−2.1	−0.1	−1.7	−2.0
captan	−9.0	−6.3	−4.7	−7.0
carbaryl	−9.5	−9.9	−6.9	−8.5
carbofuran	−9.6	−11.2	−7.1	−8.1
carbophenothion	−6.5	−3.4	1.4	−4.7
chlordane	−3.4	−3.2	5.5	−1.3
chlorfenvinphos	−7.1	−7.5	2.2	−6.6
chlorimuron ethyl	−14.0	−16.4	−9.4	−25.5
chloropicrin	−1.5	−0.2	2.6	1.0
chlorpyrifos	−5.0	−3.5	4.8	−3.6
dialifor	−5.7	−8.6	−0.1	−12.0
diazinon	−6.5	−7.1	0.4	−9.2
dicamba	−9.9	−8.5	−4.9	−6.7
diethyl propanedioate	−6.0	−6.4	−4.7	−6.0
dimethoxymethane	−2.9	−3.4	−3.8	−4.7
dinitramine	−5.7	−3.9	−0.5	−6.1
dinoseb	−6.2	−9.2	−5.8	−9.7
endosulfan alpha	−4.2	−6.0	0.8	−8.2
endrin	−5.5	−6.3	0.1	−3.6
ethion	−6.1	−4.0	2.9	−10.6
ethylene glycol diacetate	−6.3	−8.1	−4.3	−6.0
ethylene glycol mononitrate	−8.2	−5.4	−8.5	−8.7
glycerol triacetate	−8.8	−11.3	−5.9	−10.0
heptachlor	−2.6	−2.6	5.4	−0.5
imidazole	−9.8	−8.0	−10.1	−10.6
isobutyl nitrate	−1.9	0.2	−0.6	−1.9
isophorone	−5.2	−4.4	−2.9	−5.9
lindane	−5.4	−4.2	−5.0	−4.8
malathion	−8.2	−8.8	0.0	−6.2
m-bis(trifluoromethyl) benzene	1.1	0.7	1.5	−0.6
methomyl	−10.7	−10.2	−7.5	−9.5
metsulfuron methyl	−15.5	−18.9	−11.9	−25.9
N,N,4-trimethylbenzamide	−9.8	−7.4	−4.2	−8.3
N,N-dimethyl-*p*-methoxybenzamide	−11.0	−9.2	−6.0	−17.3
nitralin	−8.0	−6.0	−0.8	−16.1
nitroglycol	−5.7	−1.1	−5.5	−5.4
nitroxyacetone	−6.0	−3.4	−5.4	−6.3
parathion	−6.7	−5.3	−0.2	−9.6
pebulate	−3.6	−2.8	−0.8	−5.0
phenyl formate	−3.8	−3.9	−3.9	−3.3
phorate	−4.4	−2.9	−0.4	−6.5
pirimor	−9.4	−10.3	−3.5	−12.0
profluralin	−2.5	−0.9	4.4	−3.5
prometryn	−8.4	−8.4	−7.0	−11.9
propanil	−7.8	−9.1	−5.9	−8.9
pyrazon	−16.4	−13.2	−7.8	−12.6
simazine	−10.2	−12.0	−10.4	−14.8
sulfometuron methyl	−20.3	−19.8	−12.5	−29.3
terbacil	−11.1	−9.3	−6.1	−11.0
terbutryn	−6.7	−8.8	−7.2	−15.9
thifensulfuron	−16.2	−18.5	−13.0	−26.8
trichlorfon	−12.7	−9.6	−5.0	−13.7
trifluralin	−3.3	−0.5	4.2	−2.9
vernolate	−4.1	−3.5	−1.2	−6.4
	MSE	0.2	3.5	−1.8
	MUE	1.8	3.8	2.7
	RMSE	2.3	4.7	4.3

[a] IUPAC names and other known names for these compounds are given in the Supporting Information. MSE denotes mean signed error; MUE denotes mean unsigned error, and RMSE denotes root mean squared error. [b] Sorted in alphabetical order.

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2459**

Table 14. Various Contributions (kcal/mol) to the Free Energies of Solvation for Selected Molecules Calculated using IEFPCM/Gaussian 03 and Various Schemes for Coulomb Radii[a]

model	G_P	G_E	ΔG_{EP}	G_C	G_D	G_R	G_{CDR}	ΔG_s^o
				Lindane				
UA0	−12.73	2.04	−10.69	26.45	−23.46	2.34	5.34	−5.35
UAHF	−12.99	1.98	−11.01	27.25	−25.31	4.11	6.05	−4.97
Bondi	−13.64	2.28	−11.36	28.78	−27.74	4.56	5.61	−5.75
				Pyrazon				
UA0	−18.81	3.51	−15.31	27.59	−21.66	1.79	7.72	−7.59
UAHF	−14.23	2.49	−11.74	26.23	−26.52	4.18	3.89	−7.84
Bondi	−23.05	4.93	−18.12	28.49	−27.43	4.01	5.07	−13.05
				Pirimor				
UA0	−7.86	1.33	−6.53	39.27	−19.01	0.59	20.84	14.31
UAHF	−9.94	1.64	−8.30	32.07	−33.89	6.65	4.83	−3.47
Bondi	−16.97	3.56	−13.41	38.91	−31.65	4.33	11.59	−1.82
				Prometryn				
UA0	−12.88	1.79	−11.09	39.54	−21.97	1.10	18.67	7.58
UAHF	−13.64	1.72	−11.92	33.28	−35.13	6.82	4.96	−6.96
Bondi	−16.35	2.48	−13.87	40.44	−34.17	4.86	11.12	−2.74

[a] The table shows contributions to the standard-state free energies of solvation in water (ΔG_s^o) calculated by IEFPCM/Gaussian 03 using UA0, UAHF, and Bondi's radii for four compounds selected out of the 71 compounds presented in Table 13. The first two compounds (sorted by name in alphabetical order) have the smallest deviation in the UA0 and UAHF values of ΔG_s^o, whereas the remaining two compounds have the largest deviation. The electronic structure method used for these calculations is M05-2X/6-31G*.

of 71 aqueous solutes presented in Table 13. The first two compounds (lindane, pyrazon) have the smallest deviation in the free energies of solvation calculated using UA0 and UAHF, whereas the remaining two compounds (pirimor, prometryn) have the largest deviation. In the last two cases, the deviation between the UA0 and UAHF total solvation energies is dominated by the deviation between the corresponding nonelectrostatic terms, though the bulk electrostatic contributions are roughly the same. For instance, the G_D term for prometryn in water calculated using UA0 is 13 kcal/mol larger than G_D(UAHF), contributing to the unphysically positive value of the total free energy of solvation for this solute (7.58 kcal/mol with UA0 versus −6.96 kcal/mol with UAHF, Table 14). Thus, the electrostatic and nonelectrostatic terms are not separately meaningful in a quantitative sense, and 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.[110] We also note that all the compounds with overly positive total solvation energies obtained using UA0 have three or more methyl groups, indicating a possible problem with the CH_3 parameters used by the UA0 united atom model. The poor performance by IEFPCM models may be caused in part by the difficulty of separately estimating the cavity, dispersion, and repulsion contributions. Table 14 shows that adding these contributions leads to considerable cancellation. In contrast, SM8AD, SM8, and SMD directly model the sum (the CDS term). To make this more clear, Table 15 shows, for the same four molecules as in Table 14, the P, E, EP, and CDS contributions of SM8AD, SM8, and SMD.

To conclude this section, we will make an additional comparison of our solvation models (SM8AD, SM8, SMD) with the IEFPCM/Gaussian 03 model that uses the UAHF scheme for atomic and group radii by considering four additional solutes (acetic acid, benzaldehyde, ethanol, nicotinamide) solvated by three solvents, in particular benzene, methylene chloride, and water (Table 16). These solutes

Table 15. Various Contributions (kcal/mol) to the Free Energies of Solvation for Selected Molecules Calculated using SM8AD, SM8, and SMD[a]

model	G_P	G_E	ΔG_{EP}	G_{CDS}	ΔG_s^o
			Lindane		
SM8AD	−4.62	0.83	−3.80	−0.38	−4.18
SM8	−3.91	0.64	−3.28	0.26	−3.02
SMD	−11.65	1.95	−9.70	2.47	−7.23
			Pyrazon		
SM8AD	−17.14	5.28	−11.86	−1.34	−13.20
SM8	−14.65	3.96	−10.69	−0.63	−11.32
SMD	−21.43	4.64	−16.79	4.85	−11.94
			Pirimor		
SM8AD	−15.46	3.37	−12.09	1.79	−10.30
SM8	−14.08	2.80	−11.27	0.63	−10.64
SMD	−16.95	3.65	−13.30	6.46	−6.84
			Prometryn		
SM8AD	−9.06	1.67	−7.39	−0.96	−8.35
SM8	−9.94	1.74	−8.20	−1.03	−9.23
SMD	−13.95	1.84	−12.11	3.70	−8.41

[a] The table shows contributions to the standard-state free energies of solvation in water (ΔG_s^o) calculated by SM8AD, SM8, and SMD for compounds presented in Table 14. The electronic structure method used for these calculations is M05-2X/6-31G*.

represent major classes of chemical compounds with various functionalities, and the set of solvents is chosen to span a range of dielectric constants and solvent properties. Table 16 also contains gas-phase and liquid-phase dipole moments and compares the former to experiment.[111] SM8AD, SM8, and SMD agree well between themselves in predicting the solvation free energies for any of these solutes with the difference in ΔG_s^o lying between 0 and 1.5 kcal/mol, whereas the individual (bulk electrostatic and nonbulk electrostatic) components to the free energy of solvation vary more significantly. For instance, the difference in ΔG_{EP} as well as in ΔG_{CDS} calculated by SM8 and SMD for any solute in water is about 4 kcal/mol or larger. SM8AD, SM8, and SMD agree with IEFPCM/Gaussian 03 better for aqueous solutes than for nonaqueous ones. The IEFPCM/Gaussian 03 free energies of solvation for the nonaqueous solutes are usually

Table 16. Standard-State Free Energies of Solvation (kcal/mol) and Dipole Moments (debye) for Selected Solutes in Benzene, Methylene Chloride, and Water Calculated using SM8AD, SM8, SMD, and IEFPCM/Gaussian 03[a]

	gas	benzene				methylene chloride				water			
model	μ	ΔG_{EP}	G_{CDS}	ΔG_s°	μ	ΔG_{EP}	G_{CDS}	ΔG_s°	μ	ΔG_{EP}	G_{CDS}	ΔG_s°	μ
Acetic Acid													
SM8AD	1.59	−1.48	−2.62	−4.10	1.74	−3.03	−1.62	−4.65	1.86	−7.52	0.90	−6.62	2.04
SM8	1.59	−1.38	−2.82	−4.20	1.73	−2.73	−1.96	−4.70	1.84	−6.04	−0.32	−6.36	1.99
SMD	1.59	−2.03	−1.79	−3.82	1.74	−4.95	−0.33	−5.27	1.89	−10.03	3.83	−6.20	2.15
IEFPCM	1.59	−1.61	0.68	−0.93	1.70	−3.08	0.10	−2.98	1.78	−7.92	1.35	−6.57	1.92
exp	1.70 ± 0.03			−4.02								−6.70	
Benzaldehyde													
SM8AD	3.32	−1.71	−4.81	−6.52	3.62	−3.19	−4.00	−7.18	3.92	−5.58	1.25	−4.32	4.38
SM8	3.32	−1.68	−4.87	−6.55	3.63	−3.08	−4.22	−7.30	3.92	−4.92	0.58	−4.34	4.26
SMD	3.32	−2.61	−3.98	−6.59	3.69	−5.53	−2.73	−8.26	4.08	−8.79	4.28	−4.51	4.67
IEFPCM	3.32	−1.24	0.82	−0.42	3.62	−2.41	0.11	−2.30	3.89	−5.40	2.17	−3.23	4.26
exp												−4.02	
Ethanol													
SM8AD	1.71	−0.57	−2.82	−3.38	1.79	−1.32	−2.46	−3.78	1.87	−4.13	−0.74	−4.87	1.97
SM8	1.71	−0.72	−2.61	−3.33	1.80	−1.51	−2.27	−3.78	1.87	−3.60	−1.19	−4.79	1.95
SMD	1.71	−1.42	−1.61	−3.02	1.85	−3.54	−0.94	−4.48	2.03	−7.46	2.43	−5.04	2.30
IEFPCM	1.71	−1.12	−0.08	−1.20	1.82	−2.26	−0.75	−3.01	1.93	−5.88	0.34	−5.54	2.14
exp	1.69 ± 0.03			−3.42				−3.82				−5.01	
Nicotinamide													
SM8AD	2.07	−4.19	−3.94	−8.13	2.16	−7.41	−3.63	−11.03	2.25	−11.32	−2.25	−13.57	2.51
SM8	2.07	−4.86	−3.01	−7.86	2.17	−8.34	−2.80	−11.14	2.25	−10.70	−2.16	−12.86	2.38
SMD	2.07	−4.95	−2.26	−7.21	2.24	−10.07	−1.17	−11.24	2.46	−14.91	2.80	−12.11	2.84
IEFPCM	2.07	−2.58	0.62	−1.96	2.21	−5.04	−0.34	−5.38	2.36	−12.08	1.70	−10.38	2.61

[a] The electronic structure method used for these calculations is M05-2X/6-31G*. The IEFPCM/Gaussian 03 calculations use the UAHF scheme for atomic and group radii. The ΔG_{EP} term refers to the bulk electrostatic contribution to the free energy of solvation ΔG_s°. The G_{CDS} term refers to the cavity dispersion solvent structure component as defined by SM8AD, SM8, and SMD, and it corresponds to the cavity dispersion repulsion (G_{CDR}) component as defined by IEFPCM/Gaussian 03. The quantity of μ refers to the dipole moment, in all cases the calculated value is calculated from the electron density. Experimental values of the free energies of solvation and the gas-phase dipole moments are taken from refs 59 and 111, respectively.

much less negative than the corresponding SM8AD, SM8, SMD, and available experimental values. Note that, according to the Gaussian 03 output, IEFPCM/Gaussian 03 by default scales the UAHF radii used in these calculations differently for different solvents, using the scaling factor of 1.2 for water and 1.4 for benzene and methylene chloride.

5. Discussion of Model Physics

In this section we will examine some general advantages in using the improved electrostatic algorithm based on the generalized Born approximation with asymmetric descreening (see Section 2) as incorporated now in the new SM8AD solvation model instead of the traditional generalized Born Coulomb field approximation.

Grycuk has tested[23] the GB/CF approximation against the Kirkwood model[22] applied to biopolymer electrostatics for the case of a spherical biopolymer (in this case the Kirkwood model is equivalent to the model based on the Poisson equation for electrostatics). In a simple case when a single charge is placed inside a spherical cavity embedded in a dielectric continuum, the GB/CF model provides the exact polarization energy due to the charge only if the charge is located in the center of the spherical cavity.[23] If the charge is located near the dielectric boundary, the effective Born radius is overestimated by up to a factor of 2, and therefore, the resulting polarization energy can be underestimated by the same factor.[23] As suggested by Grycuk,[23] the use of the corrected Born radius (i.e., eq 8 instead of eq 6) allows one to effectively reduce the errors of conventional GB/CF

models related to the deficiency of the Coulomb field approximation.

The observed discrepancy between the GB/AD polarization energies and the polarization energies obtained by solving the NPE can be attributed to the possible inaccuracy of any of these models. Although the NPE electrostatics is considered as the standard in this discussion, methods that solve the NPE may have uncertainties in the bulk electrostatic part due to the portion of the solute charge that lies outside the cavity and the assumed way in which the permittivity changes at and near the solute−solvent boundary.[2,4,40,103] On the other hand, even the GB/AD model does not eliminate all the deficiencies of the GB approximation, including the oversimplified 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. With that preface, the rest of this section uses the working hypothesis that IEFPCM is an accurate standard for a given set of radii.

Table 17 compares the polarization energies calculated by using the GB/CF and GB/AD approximations and by solving the NPE for bulk electrostatics in a medium with $\varepsilon = 78.3$ for the sodium-doped fullerene cation Na$^+$@C$_{60}$ (a system with nearly spherical symmetry) and the n-butylammonium cation n-CH$_3$(CH$_2$)$_3$NH$_3^+$ (a nonspherical system). The NPE was solved by the IEFPCM algorithm[39−42] as implemented in Gaussian 03[78] with the user-defined intrinsic atomic Coulomb radii and with the default tessellation settings. The GB/CF and GB/AD approximations were evaluated using a

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2461**

Table 17. Polarization Energies (kcal/mol) Calculated using the GB/CF and GB/AD Approximations and the Nonhomogeneous Poisson Equation for Bulk Electrostatics[a]

charge model[b]	Coulomb radius				G_P		
	H	C	N	Na	GB/CF	GB/AD	NPE
			$Na^+@C_{60}$				
delocalized[c]		1.57		3.55	−33.5	−35.6	−37.0[d]
localized[c]		1.57		3.55	−33.7	−33.7	
delocalized[e]		2.04		3.55	−30.2	−32.2	−33.2
localized[e]		2.04		3.55	−30.5	−30.5	
			$n\text{-}CH_3(CH_2)_3NH_3^+$				
delocalized[c]	1.02	1.57	1.61		−71.0	−76.3	−81.9
localized[c]	1.02	1.57	1.61		−92.4	−100.2	
delocalized[e]	1.44	2.04	1.86		−58.9	−63.6	−62.7
localized[e]	1.44	2.04	1.86		−72.7	−79.2	

[a] The electronic structure method used for these calculations is M05-2X/6-31G*. The molecular geometry was optimized at the mPW1PW/3-21G level for $Na^+@C_{60}$ and the mPW1PW/MIDI! level for $n\text{-}CH_3(CH_2)_3NH_3^+$. [b] We used either the delocalized CM4 partial atomic charges on H, C, N, and Na obtained using the unpolarized (gas-phase) wave function or the localized partial atomic charges which were defined as follows: the Na atom in $Na^+@C_{60}$ has a charge of +1, and each of the three hydrogen atoms in the NH_3 group in $n\text{-}CH_3(CH_2)_3NH_3^+$ has a charge of +1/3, the charges on all other atoms in these molecules are assigned to zero. [c] The electrostatic cavity was defined by superpositions of the nuclear-centered spheres corresponding to the SM8 values of intrinsic atomic Coulomb radii for H, C, and N, and the value of the Na radius equal to the half of the largest C−C distance in $Na^+@C_{60}$ unless noted otherwise. The Na radius was chosen in the way to exclude the space inside of the C_{60} cavity from the dielectric continuum. [d] The $Na^+@C_{60}$ cavity in this case was approximated by a sphere centered on the Na atom with the radius equal to 5.12 Å that is the half of the largest C−C distance plus the SM8 radius for C. [e] The electrostatic cavity was defined by superpositions of the nuclear-centered spheres corresponding to 1.2 times the Bondi values of intrinsic atomic Coulomb radii for H, C, and N and the value of the Na radius equal to the half of the largest C−C distance in $Na^+@C_{60}$. The Na radius was chosen in a way to exclude the space inside of the C_{60} cavity from the dielectric continuum.

locally modified version of Gaussian 03, in particular MN-GSM.[79] In all three cases we used the M05-2X/6-31G* unpolarized gas-phase wave function to obtain the continuous charge density for NPE calculations and to evaluate the CM4 class IV partial atomic charges[17] based on the charge density for GB calculations. The dielectric boundary was built to precisely enclose a superposition of nuclear-centered spheres with intrinsic Coulomb radii ρ_Z, which depend only on the atomic numbers of the atoms (Z). We tested two sets of the ρ_Z values corresponding to the SM8 radii[18] and the van der Waals radii of Bondi[54] scaled by a factor of 1.2. The other details of these calculations are given in footnotes to Table 17.

Table 17 indicates large differences among the polarization energies calculated by GB/CF, GB/AD, and NPE for both $Na^+@C_{60}$ and $n\text{-}CH_3(CH_2)_3NH_3^+$. In the case of a nearly spherical molecule with the charge located near the center of the sphere such as in $Na^+@C_{60}$, one might have expected that the GB/CF and GB/AD methods should agree among themselves and should agree well with the NPE. However, this is not the case for $Na^+@C_{60}$ because there is some charge transfer between the Na^+ cation and the C_{60} shell, and transferred charge is located near the dielectric boundary

rather than in the center of the molecular cavity. The latter circumstance makes the GB/CF approximation agree less closely than GB/AD with IEFPCM. The G_P value for $Na^+@C_{60}$ varies from −33.5 (GB/CF) to −35.6 (GB/AD) kcal/mol (with the SM8 radii used to construct the $Na^+@C_{60}$ electrostatic cavity), whereas for the situation in which all the charge was localized on the Na atom, the value of G_P was found to be −33.7 kcal/mol. The former calculation, with realistic charges, is labeled "delocalized" in Table 17, and the latter is labeled "localized". In fact, the GB/AD approximation is expected to be more realistic than the traditional GB/CF approach for almost any real solute other than a monatomic system (in which case the two models converge according to eq 4) because there is almost always one or more charged groups or atoms in the solute molecule that are exposed to the solvent, i.e., located near the dielectric boundary.

One advantage of GB models is their lower computational cost compared to that of the cost of NPE solvers. The use of the GB/AD approximation instead of GB/CF allows one to reproduce the NPE results more closely and extend the applicability of generalized Born models to a wider class of solutes with no additional cost.

6. Summary

We have presented a new self-consistent reaction field universal continuum solvent model based on the GB/AD approximation introduced by Grycuk.[23] The new model is called Solvation Model 8 with Asymmetric Descreening (SM8AD). "Universal" denotes its applicability to solvation in water or any nonaqueous 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). Water is treated as a special solvent that is given its own set of model parameters. The SM8AD model is applicable to any charged or uncharged solute or supersolute. This model was parametrized over 26 combinations of electronic structure methods and basis sets with the use of CM4[17] and CM4M class IV partial atomic charges.[48] Namely, we used nine density functionals (M05-2X, M05, M06-2X, M06, M06-HF, M06-L, mPWPW, mPW1PW, and B3LYP), the Hartree−Fock method, and nine basis sets for which the charge models CM4 and CM4M are available (MIDI!, MIDI!6D, 6-31G*, 6-31+G*, 6-31+G**, 6-31G**, cc-pVDZ, DZVP, or 6-31B*).

The SM8AD model was tested against the earlier SM8 model based on the GB/CF approximation and the density-based continuum solvent model SMD over a set of 2 892 solvation data, including 345 free energies of solvation for neutral solutes in water, 2 072 free energies of solvation for neutral solutes in 90 nonaqueous solvents, 143 transfer free energies for neutral solutes between water and 15 organic solvents, and 332 free energies of solvation for ions in acetonitrile, dimethyl sulfoxide, methanol, and water. The number of solvation energy calculations performed in this testing totals to 75 192 for each of the three models. The mean unsigned error averaged over 26 theoretical levels for 2 560 solvation data for neutral solutes is 0.6, 0.7, and 0.8 kcal/mol for SM8AD, SM8, and SMD, respectively. The

mean unsigned error averaged over 26 theoretical levels for 332 free energies of solvation for ions is 4.0 (SM8AD), 4.4 (SM8), and 4.3 kcal/mol (SMD).

Supporting Information Available: Two thousand four hundred seventeen reference solvation free energies and 143 reference transfer energies for neutral solutes in the SM8AD 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 SM8AD errors in the solvation free energies for neutral compounds (part II); the functional forms of atomic surface tensions used by SM8AD (part III). This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

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

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

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

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

(6) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 393.

(7) Klopman, G. *Chem. Phys. Lett.* **1967**, *1*, 200.

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

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

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

(11) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.

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

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

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

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

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

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

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

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

(20) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378.

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

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

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

(24) Tjong, H.; Zhou, H.-X. *J. Phys. Chem. B* **2007**, *111*, 3055.

(25) Tjong, H.; Zhou, H.-X. *J. Chem. Phys.* **2007**, *126*, 195102.

(26) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297.

(27) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1348.

(28) Pokala, N.; Handel, T. M. *Protein Sci.* **2004**, *13*, 925.

(29) Wojciechowski, M.; Lesyng, B. *J. Phys. Chem. B* **2004**, *108*, 18368.

(30) Sigalov, G.; Scheffel, P.; Onufriev, A. *J. Chem. Phys.* **2005**, *122*, 094511.

(31) Tanizaki, S.; Feig, M. *J. Chem. Phys.* **2005**, *122*, 124706.

(32) Schnieders, M. J.; Ponder, J. W. *J. Chem. Theory Comput.* **2007**, *3*, 2083.

(33) Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A. *J. Chem. Phys.* **2007**, *127*, 185101.

(34) Labute, P. *J. Comput. Chem.* **2008**, *29*, 1693.

(35) Cai, W.; Xu, Z.; Baumketner, A. *J. Comput. Phys.* **2008**, *227*, 10162.

(36) Bardhan, J. P. *J. Chem. Phys.* **2008**, *129*, 144105.

(37) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673.

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

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

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

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

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

(43) Ben-Naim, A. *Solvation Thermodynamics*; Plenum: New York, 1987; p 4.

(44) Born, M. *Z. Phys.* **1920**, *1*, 45.

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

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

(47) Tuñón, I.; Ruiz-López, M. F.; Rinaldi, D.; Bertrán, J. *J. Comput. Chem.* **1996**, *17*, 148.

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

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

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

Generalized Born Approximation

J. Chem. Theory Comput., Vol. 5, No. 9, 2009 **2463**

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

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

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

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

(55) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans.* **1989**, *2*, 699.

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

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

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

(59) 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. http://comp.chem.umn.edu/mnsol (accessed Jun 20, 2009).

(60) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008**, *51*, 769.

(61) Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501.

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

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

(64) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 4538.

(65) 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. http://comp.chem.umn.edu/solvation/mnsddb.pdf (accessed Jun 20, 2009).

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

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

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

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

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

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

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

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

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

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

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

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

(78) 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.

(79) Marenich, A. V.; Olson, R. M.; 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 2009*; University of Minnesota: Minneapolis, MN, 2009.

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

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

(82) Mantina, M.; Chamberlin, A. C.; Valero, R.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2009**, *113*, 5806.

(83) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(84) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.

(85) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.

(86) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 13126.

(87) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(88) Becke, A. D. *Phys. Rev. A:At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(89) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter* **1988**, *37*, 785.

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

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

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

(93) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.

(94) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 1643.

(95) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(96) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

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

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

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

(100) 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.

(101) Marenich, A. V.; Higashi, M.; Olson, R. M.; Chamberlin, A. C.; Pu, J.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. *GAMESSPLUS - version 2009−2*, University of Minnesota, Minneapolis, 2009.

(102) Schmuckler, M. E.; Barefoot, A. C.; Kleier, D. A.; Cobranchi, D. P. *Pest Manage. Sci.* **2000**, *56*, 521.

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

(104) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875.

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

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

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

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

(109) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

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

(111) *CRC Handbook of Chemistry and Physics*; Lide, D. R., Ed.; Taylor and Francis: Boca Raton, FL, 2008−2009, Vol. 89. http://www.hbcpnetbase.com (accessed Jun 20, 2009).

CT900312Z