Generalized Born Solvation Model SM12

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

[†]Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, Minnesota 55455-0431, United States

S *Supporting Information*

ABSTRACT: We present a new self-consistent reaction-field implicit solvation model that employs the generalized Born approximation for the bulk electrostatic contribution to the free energy of solvation. The new solvation model (SM) is called SM12 (where "12" stands for 2012), and it is available with two sets of parameters, SM12CM5 and SM12ESP. The SM12CM5 parametrization is based on CM5 partial atomic charges, and the SM12ESP parametrization is based on charges derived from a quantum-mechanically calculated electrostatic potential (ESP) (in particular, we consider ChElPG and Merz−Kollman−Singh charges). The model was parametrized over 10 combinations of theoretical levels including the 6-31G(d) and MG3S basis sets and the B3LYP, mPW1PW, M06-L, M06, and M06-2X density functionals against 2979 reference experimental data. The reference data include 2503 solvation free energies and 144 transfer free energies of neutral solutes composed of H, C, N, O, F, Si, P, S, Cl, Br, and I in water and in 90 organic solvents as well as 332 solvation free energies of singly charged anions and cations in acetonitrile, dimethyl sulfoxide, methanol, and water. The advantages of the new solvation model over our previous generalized Born model (SM8) and all other previous generalized Born solvation models are (i) like the SMD model based on electron density distributions, it may be applied with a single set of parameters with arbitrary extended basis sets, whereas the SM8 model involves CM4 or CM4M charges that become unstable for extended basis sets, (ii) it is parametrized against a more diverse training sets than any previous solvation model, and (iii) it is defined for the entire periodic table.
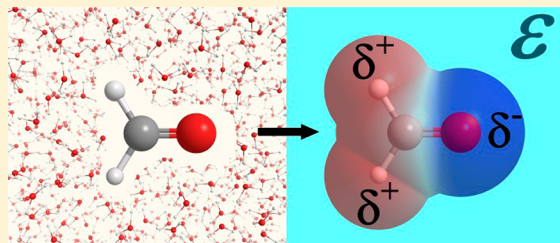
1. INTRODUCTION

Continuum solvation models (sometimes called implicit solvent models) are cost-efficient tools for studying solvent effects on molecular structure, spectra, energetics, and dynamics.[1−3] Implicit solvent models eliminate the difficulties of *explicit* solvation modeling associated with the statistical sampling of solvent configuration space. However, they do so at the expense of losing molecular-level detail. In particular, the description of the solvent in the first solvation shell is incomplete if one assumes that the solute−solvent interaction does not depend on the solvent's molecular structure or that such an interaction involves only long-range electrostatic polarization effects (bulk electrostatics) for which the dielectric response of the medium is uniform and linear at all positions outside the solute's molecular cavity. In general, a more accurate theoretical description of solvent effects should account not only for the bulk electrostatics but also for shorter-range polarization effects and nonelectrostatic effects such as cavitation, dispersion, and solvent structural effects (for example, hydrogen bonding effects, exchange repulsion, and the unique dielectric characteristics of the first solvation shell). On the other hand, there is no thermodynamically unique (path-independent) way to separate the bulk electrostatic component of the solvation free energy from other components, and only their sum is a state function.[4,5] Therefore, continuum solvation models differ from each other in the way in which the bulk-electrostatic term and the nonbulk-electrostatic terms are defined.

The electrostatic contribution to the free energy of solvation can be calculated self-consistently by solving the nonhomogeneous-dielectric Poisson equation (NPE) based on the continuous charge density or by using other approaches, for example, the generalized Born (GB) approximation[6−11] based on representation of the solute as a system of point charges located at the nuclear positions. In the past, we developed a number of successively improved self-consistent reaction-field solvation models that utilize the GB approximation (most recently, SM5.4,[12] SM5.42,[13−15] SM5.43,[16] SM6,[17] SM8,[18] and SM8AD[19]) or the NPE (SM5C,[20] SMD[21]) for bulk electrostatics combined with empirical atomic surface tension terms[22] (also called CDS terms) that account for cavity formation (C), dispersion (D), and solvent-structure (S) effects, where the S contribution includes the deviation of short-range electrostatic effects from the bulk electrostatics.

One advantage of GB models is their lower computational cost compared to that of the cost of NPE solvers (i.e., density-based models). However, since GB solvation models are based on partial atomic charges (in particular, our class IV charges[23] in all the GB models mentioned above), their accuracy for a particular level of electronic structure theory may depend on whether meaningful partial atomic charges can be computed for that level of theory, for example, with a particular basis set. When such meaningful charges are not available, it has been

necessary to use density-based solvation models, which are deemed to be less sensitive to the choice of basis set.[19,21]

Our previous GB solvation models utilize partial atomic charges obtained using the CMx charge models (x = 1−4 or 4M) (see refs 23−25 and references therein). These charge models were designed to map charges from Löwdin population analysis[26,27] or redistributed Löwdin population analysis[28] onto new sets of charges (for example, CM4[17] and CM4M[24] charges) using simple functional forms parametrized for several combinations of density functionals and basis sets to accurately reproduce the reference dipole moments for a large number of polar molecules. A major limitation of CM4, CM4M, and other models based on Löwdin charges is that they depend on the basis set and were optimized at most for 10 relatively small basis sets where 6-31+G(d,p) is the largest one.[24] An extension of the existing CMx models to larger basis sets (for example, to a basis of triple-ζ quality) is not advisable because the utility of Löwdin charges deteriorates rapidly upon increasing the number or diffuseness of the basis functions (as shown in ref 25), and the quality of Mulliken charges deteriorates even faster. We have recently presented a new charge model called Charge Model 5 (CM5).[25] CM5 partial atomic charges are derived from Hirshfeld population analysis[29] charges rather than from Löwdin or redistributed Löwdin charges. We have found that Hirshfeld charges are less sensitive to basis set size or the choice of basis as compared to those from other population analyses tested,[25] which made it possible to create a new model (CM5) that uses a single set of model parameters that depend neither on the electronic structure method nor on the basis set. In addition, the CM5 model can be used with any element of the periodic table, whereas, for example, the previous CM4 and CM4M models are only applicable to molecules containing H, Li, C, N, O, F, Si, P, S, Cl, and/or Br. In our recent work,[25] we have also found that partial atomic charges derived from quantum mechanical electrostatic potentials (ESP) yield dipole moments that are on average slightly more accurate than those obtained from CM5 charges in comparison with available reference data.[25] In particular, we considered ChElPG[30] and Merz−Kollman[31,32] (MK) electrostatic potential schemes. Overall, CM5 partial atomic charges are less computationally expensive than ESP charges, they are less dependent on molecular conformation, and they do not suffer from ill conditioning for buried atoms in larger molecules, so they avoid the three main deficiencies of ESP charges.

In the present work, we introduce a new generalized Born solvation model (called SM12 where "12" stands for 2012) that can be used with CM5 partial atomic charges, thereby overcoming the aforementioned deficiencies of our previous GB models (for example, SM8 and SM8AD as the most recent ones) based on class IV charges (in the case of SM8 and SM8AD, CM4 and CM4M charges) derived from Löwdin or redistributed Löwdin population analysis. In comparison with the SM8 training set,[18] the SM12 training set contains an additional 158 reference data, thereby making the new parametrization more robust. In addition, we have also parametrized the SM12 solvation model for use with ChElPG and MK partial atomic charges (i.e., ESP charges), though the ESP charges may exhibit some deficiencies as outlined in the previous paragraph. It is important for the CDS terms to be consistent with the chosen bulk-electrostatics model;[3] therefore, the SM12 model has two sets of CDS parameters, SM12CM5 and SM12ESP. This allows us to compare the

SM12 free energies of solvation computed using CM5 charges to those obtained using ChElPG or MK charges. The SM12 training set contains solutes composed only of H, C, N, O, F, Si, P, S, Cl, Br, and I (our earlier solvation models except for SM5.42[13] were not parametrized for iodine-containing solutes). We estimated the SM12 parameters for elements not present in the training set, thereby extending the SM12 model to the whole periodic system of the elements.

2. COMPUTATIONAL DETAILS

The theoretical free energy of solvation is computed according to

$$\Delta G_{\mathrm{S}}^{\otimes} = \Delta E_{\mathrm{E}} + G_{\mathrm{P}} + G_{\mathrm{CDS}} + \Delta G_{\mathrm{N}} + \Delta G_{\mathrm{conc}}^{\otimes} \tag{1}$$

where the symbol $\otimes$ denotes an arbitrary choice of standard states, ΔE_{E} is the change in the solute's internal electronic (E) energy in transferring from the gas phase to the liquid phase at the same geometry, G_{P} is the polarization free energy of the solute−solvent system when the solute is inserted, G_{CDS} is the component of the free energy that is nominally associated with cavitation, dispersion, and solvent structure, ΔG_{N} is the change in $\Delta G_{\mathrm{S}}^{\mathrm{o}}$ due to a change in nuclear coordinates, and $\Delta G_{\mathrm{conc}}^{\otimes}$ accounts for the difference in concentrations, if any, in the gas-phase standard state and the solution-phase one. The two most common choices of standard states are (i) 1 M ideal solution and 1 bar ideal gas, which are obtained by $\Delta G_{\mathrm{conc}}^{\otimes}$ = 1.9 kcal/mol, which yields $\Delta G_{\mathrm{S}}^{\mathrm{o}}$ and (ii) 1 M concentration in both solution and gas, which are obtained by $\Delta G_{\mathrm{conc}}^{\otimes}$ = 0 kcal/mol, which yields $\Delta G_{\mathrm{S}}^{*}$. In the present article we use choice (ii).

All calculations of $\Delta G_{\mathrm{S}}^{*}$ in the present article are based on rigid, gas-phase geometries; therefore, the ΔG_{N} component in eq 1 is taken to be zero here (although this does not restrict application of the model more generally).

Bulk Electrostatic Contribution. When geometries are not relaxed, the sum of the electronic relaxation term ΔE_{E} and the polarization term G_{P} in eq 1 gives the bulk electrostatic contribution ($\Delta G_{\mathrm{EP}} = \Delta E_{\mathrm{E}} + G_{\mathrm{P}}$) to the solvation free energy, and this is obtained from a self-consistent molecular orbital calculation (see, for example, ref 13) with the polarization component G_{P} defined using the GB approximation as follows

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

where the index k or k' runs over all nuclei in the solute molecule, ε is the bulk static dielectric constant of the solvent, q_k is the partial atomic charge of atom k, and $\gamma_{kk'}$ is a Coulomb integral. The SM12 model uses either CM5 charges or ESP charges (either ChElPG or MK).

The Coulomb integral is calculated in the present work according to the approximation of Still et al.[9] as

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

where $R_{kk'}$ is the internuclear distance, and α_k is the effective Born radius of atom k. The quantity d is an empirical constant that is set to the value of 3.7 which was found to be optimal in earlier work.[17] In the case of atoms and monatomic ions, the GB result reduces to the original Born formula[6]

$$G_{\mathrm{P}} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\frac{q^2}{\alpha} \tag{4}$$

which is the exact classical result for the case where the solute molecule is a conducting sphere of radius α and charge q. As in the case of all our earlier GB models except for SM8AD,[19] the effective Born radius of atom k in eq 3 is expressed using the Coulomb field approximation (i.e., we assume that the electric displacement field induced by the charge q_k is a Coulomb field) according to the following equation[33,34]

$$\alpha_k = \left(\frac{1}{R'} + \int_{\rho_{Z_k}}^{R'} \frac{A_k(r)}{4\pi r^4} dr \right)^{-1} \tag{5}$$

where R' is the radius of the sphere centered on atom k that fully engulfs all other spheres centered on the other atoms of the solute, ρ_{Z_k} is the intrinsic Coulomb radius of atom k which depends only on the atomic number Z_k, and $A_k(r)$ is the exposed or solvent-accessible surface area of a sphere of radius r that is centered on atom k. This area, calculated here using the analytic surface area (ASA) algorithm,[33] depends on the geometry of the solute and the radii of the spheres centered on all the atoms in the solute.[35] The choice of radius values to use as the SM12 intrinsic Coulomb atomic radii will be discussed in more detail below.

Cavity−Dispersion−Solvent-Structure Contribution. The G_{CDS} contribution to the free energy of solvation is defined in the same way as in our earlier work.[19] In particular, we use

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

where σ_k is the atomic surface tension of atom k, A_k is the solvent-accessible surface area (SASA) of atom k computed by the ASA algorithm,[33] and $\sigma^{[M]}$ is the molecular surface tension. The SASA depends on the geometry $\mathbf{R}$, the set $\{R_{Z_k}\}$ of atom-number-dependent van der Waals radii, and the solvent radius r_s added to each of R_{Z_k}. The model uses the values of R_{Z_k} equal to Bondi's values[36] for those elements for which such values are available or the values of van der Waals radii from ref 37 for any other element. We use the value of r_s equal to 0.4 Å.[16,17]

The quantities σ_k in eq 6 are defined by the following equation

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

where $\tilde{\sigma}_Z$ is a quantity that depends on the atomic number of atom k (we omit index k for simplicity), $\tilde{\sigma}_{ZZ'}$ is a quantity 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. Further details of eqs 6 and 7 and the COTs are described in the Supporting Information (SI).

The quantities $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ in eq 7 depend on a set of solvent descriptors as follows

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

where $\tilde{\sigma}_i$ is either $\tilde{\sigma}_Z$ or $\tilde{\sigma}_{ZZ'}$, n is the refractive index of the solvent at room temperature (which is conventionally taken as 293 K for this quantity), α is Abraham's[38−41] 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 the index i which is either Z or ZZ'.

The quantity $\sigma^{[M]}$ in eq 6 is also a function of solvent descriptors, and it is given by

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

where γ is the macroscopic surface tension of the solvent at the air/solvent interface at 298 K; we express surface tension in units of $cal \cdot mol^{-1} \cdot Å^{-2}$ (note that 1 dyn/cm $= 1.43932$ $cal \cdot mol^{-1} \cdot Å^{-2}$), and $\gamma_o = 1$ $cal \cdot mol^{-1} \cdot Å^{-2}$, ϕ^2 is the square of the fraction of non-hydrogen atoms in the solvent molecule that are aromatic carbon atoms (carbon aromaticity), ψ^2 is the square of the fraction of non-hydrogen 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 and of the atomic numbers of atoms in the solute. The quantity $\sigma^{[M]}$ in eq 6 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.

As in the case of SM8 and SM8AD,[18,19] the SM12 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. (Note that the acidity and basicity parameters that we use as solvent descriptors are taken from Abraham's values determined for solutes; this has been done in all of our universal solvation models because the solute tables are much more complete.) Since water is so important as a solvent, we use specific parameters for water. Thus water is treated as a special solvent that is given its own set of the $\tilde{\sigma}_i^{[n]}$ parameters (where i is Z or ZZ') used in eq 8 while the coefficients $\tilde{\sigma}_i^{[\alpha]}$ and $\tilde{\sigma}_i^{[n]}$ used in eq 8 and the coefficients $\tilde{\sigma}^{[\gamma]}$ and $\tilde{\sigma}^{[\beta^2]}$ used in eq 9 are the same as those for nonaqueous solvents. The coefficients $\tilde{\sigma}^{[\phi^2]}$ and $\tilde{\sigma}^{[\psi^2]}$ are not needed for water because the quantity $\sigma^{[M]}$ defined by eq 9 retains only the first and the fourth term. In our previous solvation models (for example, SM8), we set $\sigma^{[M]}$ to zero for water, but doing it with nonzero $\sigma^{[M]}$ for water makes it easier to compare water to other solvents and easier to understand the model.

The free energy of transfer from water (aq) to nonaqueous (naq) solvent is calculated as

$$\Delta G_{naq/aq} = -2.303RT \log P_{naq/aq} = \Delta G_{S,naq}^* - \Delta G_{S,aq}^* \tag{10}$$

where $P_{naq/aq}$ is the partition coefficient determined as the equilibrium concentration of a given solute in the nonaqueous solvent (for example, in n-octanol) divided by the concentration in water.

SM12 Training Set. The SM12 training set is part of the Minnesota Solvation Database[42] and contains 2979 experimental solvation data for 233 ionic and 525 neutral solutes composed of H, C, N, O, F, Si, P, S, Cl, Br, and I. 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 SM12 training set contains several data subsets:

(a) 274 aqueous solvation free energies for 274 unique neutral compounds used in our previous models;[17−19,21]

(b) 66 aqueous solvation free energies for 66 neutral compounds that include 13 unique compounds selected from refs 43 and 53 compounds as part of the SAMPL1[44] test set;

(c) 26 aqueous solvation free energies for 26 neutral compounds from SAMPL2;[45]

(d) 8 aqueous solvation free energies for 8 iodine-containing hydrocarbons introduced in earlier work;[13]

(e) 2129 solvation free energies in 90 nonaqueous (organic) solvents for 310 neutral solutes (232 of the 310 solutes are also included in subset [a], one solute is included in subset [b], 24 solutes are included in subset [c], 8 solutes are included in subset [d], and 45 solutes are additional);

(f) 144 transfer free energies between water and 14 organic solvents for 106 additional neutral solutes;

(g) 112 aqueous free energies of solvation for 112 selectively clustered singly charged ions (there are 81 unclustered ions and 31 clustered ions);[17−19,21,46]

(h) 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).[17−19,21,47]

The SM12 training set is similar to those used in the parametrization of our earlier models (for instance, the SM8 model[18]), but it also includes several improvements. The older training sets[18] were smaller; for example, the SM8 training set includes 274 aqueous solvation free energies (the same as subset [a]), 2072 nonaqueous solvation free energies, 143 transfer free energies, and 332 ionic data (the same as subsets [g] and [h]). In the present work, we have added 66 aqueous solvation free energies of subset (b) to the SM8 training set.[18] Note that these 66 data were also included in the training set of our latest SM8AD model[19] and were described in more detail in ref 19. Besides the 66 data, we have added the solvation free energy of 1,2,3,4,5,6-hexachlorocyclohexane (lindane) in n-octanol (−10.71 kcal/mol) obtained from the measured octanol−air partition coefficient,[48] 8 aqueous solvation free energies of subset (d) for 8 iodine-containing hydrocarbons (iodomethane, diiodomethane, iodoethane, 1-iodopropane, 2-iodopropane, 1-iodopentane, 1-iodobutane, and iodobenzene) from ref 13, and 20 nonaqueous solvation free energies for 9 iodine-containing hydrocarbons (the same as above, plus 3-iodopropene) in n-heptane, cyclohexane, n-octanol, n-hexadecane, decalin (cis/trans mixture), diethyl ether, chloroform, and carbon tetrachloride.[13] We have also added 18 water-to-octanol transfer free energies for 18 new compounds derived from experimental octanol−water log P values. In particular, we added the data for 2-, 3-, and 4-fluoroanisole,[49] 2-, 3-, and 4-chloroanisole,[49,50] pentachloroanisole,[49,50] 2-, 3-, and 4-bromoanisole,[49,50] 2,6-difluoropyridine,[49,51] 4-cyanopyridine,[49,51] 4-acetylpyridine,[49,51] methyl phenyl sulfoxide,[49,51] methyl phenyl sulfone,[49,51] cinnoline,[49,51] biphenyl,[49,52] and 9H-fluorene.[49,52] We have also added 26 aqueous solvation free energies for 26 SAMPL2 compounds,[45] and for five of them, namely, uracil, 5-fluorouracil, 5-(trifluoromethyl)uracil, 5-chlorouracil, and 5-bromouracil, we already had 17 reference transfer energies between water and n-butanol, n-heptane, n-octanol, chloroform, diethyl ether, or ethyl acetate in our older

(SM8 or SM8AD) training sets. These transfer energies were replaced with 17 solvation free energies in the corresponding solvents. For 19 out of the remaining 21 SAMPL2 compounds we added 19 solvation free energy data in n-octanol derived from experimental octanol−water log P values.[49]

Table 1 lists 92 solvents including water used in the SM12 parametrization. The corresponding values of solvent descrip-

Table 1. Solvents Used in the SM12 Training Set[a]

solvent		
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 (cis/trans mixture)	p-isopropyltoluene	undecane
decane	mesitylene	water*
decanol	methanol*	xylene (m/o/p 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 14 solvents for which we used solvent−water transfer free energies of neutral solutes are italicized.

tors such as dielectric constant, refractive index, bulk (macroscopic) surface tension, and acidity and basicity parameters were taken from the Minnesota Solvent Descriptor Database[53] (these values are given in the SI).

All computed solvation free energies in this study are based on rigid, gas-phase geometries. In earlier work,[17−19,21] the molecular geometries of all neutral and ionic solutes were optimized at the mPW1PW[54]/MIDI![55,56] level of theory. Here, we improve on the earlier work by using the M06-2X[57,58]/MG3S method to optimize the geometries. In particular, we use a large basis set (MG3S). The notation MG3S hereafter denotes the MG3S all-electron basis[59] on H, C, N, O, F, Si, P,

Table 2. SM12 CDS Parameters (cal·mol^{-1}·Å^{-2})^a

	SM12CM5				SM12ESP			
i	$\tilde{\sigma}_i^{[n]}$(water)	$\tilde{\sigma}_i^{[n]}$	$\tilde{\sigma}_i^{[\alpha]}$	$\tilde{\sigma}_i^{[\beta]}$	$\tilde{\sigma}_i^{[n]}$ (water)	$\tilde{\sigma}_i^{[n]}$	$\tilde{\sigma}_i^{[\alpha]}$	$\tilde{\sigma}_i^{[\beta]}$
H	15.20	38.03			17.50	37.37		
C	121.00	75.20	−175.54	107.91	111.65	73.93	−138.61	79.25
HC	−59.73	−90.54	40.83	−15.20	−57.71	−86.96	33.73	−9.42
CC	−91.66	−70.16	116.95	−71.80	−90.69	−71.73	96.82	−53.46
N	65.11	40.66	−125.46	117.85	92.24	64.05	−123.14	116.89
HN	−33.07	−60.95		−99.80	−36.20	−55.39		−55.86
CN	−172.49	−81.97	367.92	−205.14	−221.19	−101.18	444.91	−200.06
NC	−14.51	−6.25	−36.61		−14.75	−6.16	−50.97	
NC2	−222.94	−177.10			−161.16	−152.98		
O	−141.06	−42.86	101.40	−53.33	−196.96	−50.55	168.69	−33.57
HO	13.31	−52.98		−224.31	90.67	−25.14		−223.88
OC	61.90	18.00			149.14	42.80		
CO2	85.26	−35.75			33.95	−46.73		
ON	151.59	44.85	−84.35	69.18	244.55	65.57	−120.15	58.06
OO	23.54		78.28		69.59		−0.36	
F	6.45	−2.87			13.37	0.49		
Cl	−17.48	−27.89			−19.88	−28.62		
Br	−24.01	−35.36			−26.25	−35.59		
I	−27.72	−36.80			−28.85	−37.45		
Si	−13.86	−101.53			−14.45	−82.20		
P	−13.86				−14.45			
OP	73.15				94.32			
S	−48.09	−52.12			−47.99	−49.66		
OS	−117.83		478.06		−47.74		390.33	
SP	50.92	45.87			24.83	25.64		
X^b	−14.61	−14.00			−11.45	−12.00		

^{a}These are the CDS parameters that depend on atomic numbers. The index X refers to any element other than H, C, N, O, F, Si, P, S, Cl, Br, and I. Any possible atomic-number-dependent surface tension parameter that is not in this table is set equal to zero. There are also four CDS parameters that do not depend on atomic numbers. For the SM12CM5 set, they are equal to $\tilde{\sigma}^{[\gamma]} = 0.17$, $\tilde{\sigma}^{[\phi^2]} = -2.30$, $\tilde{\sigma}^{[\psi^2]} = -7.02$, and $\tilde{\sigma}^{[\beta^2]} = 6.98$ (in cal·mol^{-1}·Å^{-2}), and, for the SM12ESP set, they are equal to 0.18, −2.44, −6.78, and 5.00, respectively. ^{b}The row gives $\tilde{\sigma}_0^{[n]}$ for water and nonaqueous solvents.

S, and Cl, the ma-TZVP all-electron basis set on Br, and the ma-TZVP valence basis set on I along with the effective core potential ECP-28[60,61] (to replace 28 core electrons). The ma-TZVP basis sets on Br and I were derived from the corresponding def2-TZVP basis sets[60,62] according to a procedure described elsewhere.[63]

SM12 Parametrization. The SM12 model contains two types of parameters: (i) the intrinsic Coulomb radii used in eq 5 for constructing the cavities for the bulk electrostatic calculation; (ii) the atomic surface tensions $\tilde{\sigma}_Z$ and $\tilde{\sigma}_{ZZ'}$ used in eq 7 and the parameters $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ used in eq 9 for the CDS calculation.

While parametrizing the earlier SM8 model,[18] we optimized the values of the intrinsic Coulomb radii in calculations considering only ions (in particular, in water, acetonitrile, methanol, and dimethyl sulfoxide) because the magnitudes of solvation free energies of ions are many times larger than those of neutral solutes and are dominated by large electrostatic contributions.[18,19] In the present work, we set the SM12 intrinsic Coulomb radii on H, C, N, O, F, Si, P, S, Cl, and Br at their SM8 values optimized in earlier work.[18] The SM12 radius for iodine was set to the value of 2.6 Å optimized to reproduce (within 1 kcal/mol) the reference solvation free energy of I$^-$ in water (−63 kcal/mol)[13] and the aqueous solvation free energy of the H$_2$O·I$^-$ cluster (−60 kcal/mol, as estimated based on eq 15 of ref 21). For other elements, we recommend to use the

handbook values[37] of van der Waals radii for the whole periodic system, which were determined such that they are on the same scale as Bondi's values[36] for the most widely used radii. Note that the SM12 values of intrinsic Coulomb radii used in eq 5 do not depend on whether CM5 partial atomic charges or ESP charges are used in eq 2.

The cavity−dispersion−solvent-structure term (eq 1) that describes nonbulk-electrostatic effects is parametrized using only neutral solutes by optimizing $\tilde{\sigma}_i^{[n]}$, $\tilde{\sigma}_i^{[\alpha]}$, $\tilde{\sigma}_i^{[\beta]}$, $\tilde{\sigma}^{[\gamma]}$, $\tilde{\sigma}^{[\phi^2]}$, $\tilde{\sigma}^{[\psi^2]}$, and $\tilde{\sigma}^{[\beta^2]}$ (where the subscript i refers to Z or ZZ'). The optimization is done by minimization of the following error function

$$\chi = \sum_{K=1}^{N} \left[\Delta G_S^*(\text{ref.}, K) - \frac{1}{M} \sum_{k=1}^{M} \Delta G_{EP}(k, K) - G_{CDS}(K) \right]^2 \tag{11}$$

where the index K runs over all N data in the SM12 training set including aqueous and nonaqueous neutral solutes ($N = 2647$), the index k runs over all M levels of electronic structure theory used in the parametrization ($M = 10$), $\Delta G_S^*(\text{ref.}, K)$ is the experimental solvation free energy or transfer free energy, $\Delta G_{EP}(k, K)$ is the bulk electrostatic energy ($\Delta G_{EP} = \Delta E_E + G_P$) computed for a given theoretical level (or the bulk electrostatic

613

dx.doi.org/10.1021/ct300900e | J. Chem. Theory Comput. 2013, 9, 609−620

Table 3. Mean Unsigned Errors in Aqueous Solvation Free Energies of Neutral Compounds (in kcal/mol)[a]

| | | SM12CM5 | | | | SM12ESP | | | | | |
| | | B3LYP | | M06-2X | | MK | | ChElPG | | | |
solute class	N	MG3S	6-31G(d)	MG3S	6-31G(d)	B3LYP	M06-2X	B3LYP	M06-2X	SM8	SMD
274 Aqueous Data											
H_2, NH_3, H_2O, $(H_2O)_2$	4	1.63	2.02	1.68	2.05	1.66	1.70	1.59	1.62	1.92	1.74
hydrocarbons	40	0.70	0.76	0.73	0.78	1.00	1.23	0.66	0.64	0.62	0.25
alcohols and phenols	16	0.73	0.40	0.62	0.36	0.35	0.44	0.55	0.44	0.59	0.17
ethers	12	0.45	0.54	0.48	0.59	0.69	0.75	0.60	0.50	0.51	0.73
aldehydes and ketones	18	0.32	0.42	0.35	0.45	0.58	0.53	0.47	0.39	0.29	0.18
carboxylic acids	5	0.98	0.56	0.87	0.50	0.25	0.24	0.34	0.33	0.72	0.62
esters	13	0.14	0.44	0.19	0.52	0.72	0.75	0.60	0.60	0.25	0.24
peroxides	3	0.38	0.31	0.30	0.38	0.51	0.59	0.40	0.47	0.10	1.12
bifunctional HCO compounds	5	0.54	0.33	0.46	0.37	0.95	1.15	0.72	0.75	0.79	0.49
aliphatic amines	15	0.82	0.80	0.81	0.81	0.53	0.58	0.39	0.46	0.65	0.63
anilines	7	0.75	0.44	0.68	0.43	0.47	0.54	1.86	1.40	0.47	0.18
aromatic nitrogen heterocycles, including 1-methylthymine	13	0.82	0.56	0.75	0.54	0.66	0.70	1.05	0.90	0.34	0.56
nitriles	4	0.33	0.20	0.36	0.24	0.74	0.58	0.49	0.27	0.45	0.65
hydrazines	3	0.70	0.82	0.75	0.85	0.50	0.62	0.80	0.61	0.56	0.54
bifunctional HCN compounds	3	1.75	1.00	1.60	1.07	1.11	0.85	1.44	1.39	1.03	0.32
amides and ureas	6	0.78	0.95	0.79	1.00	0.65	0.61	1.08	0.98	0.67	0.93
nitrohydrocarbons	7	0.71	1.07	0.96	1.35	1.38	1.52	1.17	1.29	0.47	0.62
bifunctional HCNO compounds	3	0.36	0.44	0.41	0.48	0.63	0.86	0.40	0.61	0.27	0.03
compounds containing H, C, and F	6	0.38	0.49	0.47	0.58	0.78	0.89	0.58	0.67	0.53	0.47
compounds containing H, C, and Cl	27	0.26	0.40	0.29	0.46	0.81	1.00	0.49	0.53	0.40	0.75
compounds containing H, C, and Br	14	0.15	0.25	0.18	0.29	0.41	0.51	0.28	0.20	0.30	0.54
multihalogen (F, Cl, or Br) hydrocarbons	12	0.40	0.42	0.42	0.47	0.50	0.52	0.45	0.46	0.29	0.72
halogen (F, Cl, or Br) compounds containing H, C, N, and/or O	9	1.29	1.08	1.33	1.13	1.20	1.27	1.27	1.37	1.25	1.29
sulfur compounds not containing P	14	0.54	0.45	0.52	0.48	0.65	0.72	0.54	0.51	0.72	0.47
phosphorus compounds	14	0.81	0.74	0.71	0.90	1.13	1.15	1.03	1.00	1.16	1.55
silicon compound	1	0.26	0.28	0.27	0.28	0.98	0.97	0.74	0.68	0.16	0.65
all data	274	0.59	0.59	0.60	0.63	0.75	0.84	0.69	0.66	0.57	0.57
66 Aqueous Data (SAMPL1)											
compounds containing H, C, and O	9	0.50	0.75	0.58	0.85	0.78	0.86	0.68	0.69	0.71	0.66
compounds containing H, C, N, and/or O	18	1.37	1.54	1.41	1.57	1.51	1.60	1.28	1.30	2.27	2.24
halogen (F, Cl, or Br) compounds containing H, C, N, and/or O	19	1.38	1.15	1.29	1.10	1.51	1.59	1.43	1.48	1.82	1.61
sulfur compounds not containing P	9	1.35	1.24	1.27	1.21	1.24	1.24	1.25	1.23	2.69	2.06
phosphorus compounds	11	1.63	1.41	1.55	1.45	1.47	1.44	1.46	1.40	1.90	1.96
all data	66	1.29	1.26	1.27	1.27	1.37	1.42	1.27	1.28	1.92	1.77
26 Aqueous Data (SAMPL2)											
all data	26	1.47	1.11	1.36	1.06	1.05	1.04	1.36	1.25	1.58	2.29
8 Aqueous Data (Iodine-Containing Hydrocarbons)											
all data	8	0.35	0.35	0.45	0.41	0.59	0.62	0.32	0.27		

[a]The table shows mean unsigned errors (relative to reference data) in solvation free energies computed at the SM12/DFT/MG3S and SM12/DFT/6-31G(d) level of theory using CM5 charges and at the SM12/DFT/MG3S level using Merz−Kollman (MK) and ChElPG charges where DFT is B3LYP and M06-2X. The errors in the energies computed by SM8/CM4M/M06-2X/6-31G(d) and SMD/M05-2X/6-31G(d) are also reported. N is the number of data in a given solute class. Iodine-containing compounds are not included in the SM8 and SMD statistics because the models do not have parameters to treat iodine-containing compounds. The SM8 and SMD errors differ slightly from those published previously for the same subsets of data (see, for example, Table 8 in ref 19) because of different electronic structure methods used for optimizing the solute geometries (see Section 2 for more details).

contribution to a transfer free energy), and $G_{CDS}(K)$ is the nonbulk-electrostatic energy defined above (or the corresponding contribution to a transfer free energy).

Note that we optimized two independent sets of the atomic and molecular surface tensions: one is for the case when CM5 partial atomic charges are employed to calculate ΔG_{EP} (SM12CM5), and another is for the case when ESP partial atomic charges are used (SM12ESP). In the case of SM12CM5, the CDS term was parametrized by averaging over ten combinations of electronic structure methods and basis sets (i.e., $M = 10$ in eq 11). Namely, we used the CM5 charge model combined with the 6-31G(d)[64] and MG3S basis sets and the B3LYP,[65−68] mPW1PW,[54] M06-L,[69] M06,[57,58] and M06-2X[57,58] density functionals. In the 6-31G(d) calculations we use the 6-31G(d) basis sets defined by *Gaussian 09*[70] for all the tested elements except for iodine, for which we use the MIDI![55,56] basis.

Table 4. Mean Unsigned Errors in Nonaqueous Solvation Free Energies and Transfer Energies of Neutral Compounds (in kcal/mol)[a]

| | | SM12CM5 | | | | SM12ESP | | | | | |
| | | B3LYP | | M06-2X | | MK | | ChElPG | | | |
solute class	N	MG3S	6-31G(d)	MG3S	6-31G(d)	B3LYP	M06-2X	B3LYP	M06-2X	SM8	SMD
2129 Nonaqueous Data											
H$_2$, NH$_3$, H$_2$O	29	1.40	1.56	1.43	1.59	1.25	1.28	1.22	1.25	1.54	1.19
hydrocarbons	266	0.46	0.54	0.50	0.57	0.75	0.98	0.47	0.47	0.49	0.57
alcohols and phenols	381	0.45	0.40	0.43	0.39	0.38	0.44	0.52	0.45	0.45	0.51
ethers	87	0.45	0.43	0.43	0.42	0.46	0.52	0.51	0.46	0.66	0.58
aldehydes and ketones	227	0.42	0.46	0.44	0.47	0.48	0.50	0.42	0.40	0.73	0.67
carboxylic acids	120	0.48	0.47	0.47	0.47	0.49	0.50	0.50	0.50	0.61	0.57
esters, including lactones	243	0.40	0.40	0.40	0.41	0.47	0.45	0.50	0.48	0.41	0.50
peroxides	17	0.66	0.65	0.63	0.66	0.73	0.75	0.74	0.75	0.59	0.84
bifunctional HCO compounds	32	1.41	1.22	1.38	1.22	0.98	0.85	1.47	1.36	1.45	1.32
aliphatic amines	154	0.50	0.47	0.49	0.47	0.41	0.41	0.39	0.39	0.45	0.46
anilines	61	0.43	0.47	0.41	0.48	0.50	0.68	1.06	0.76	0.37	0.63
aromatic nitrogen heterocycles including 1-methylthymine and uracil	66	0.66	0.60	0.63	0.60	0.87	1.01	0.86	0.82	0.72	0.89
nitriles	20	0.41	0.49	0.41	0.47	0.85	0.79	0.67	0.58	0.55	0.60
hydrazines	5	1.20	1.30	1.22	1.32	1.00	1.01	1.06	1.05	1.03	1.29
bifunctional HCN compounds	2	0.21	0.95	0.21	1.15	1.96	2.17	0.66	0.70	1.39	1.55
amides, ureas, and lactams	37	0.85	0.95	0.86	0.99	0.92	0.96	1.11	1.05	0.59	0.73
nitrohydrocarbons	86	0.54	0.57	0.58	0.68	0.54	0.60	0.63	0.67	0.66	0.66
bifunctional HCNO compounds	7	0.78	0.73	0.83	0.81	1.02	0.97	1.02	0.96	0.96	2.12
compounds containing H, C, and F	17	0.64	0.68	0.67	0.70	0.99	1.17	0.74	0.81	0.64	0.84
compounds containing H, C, and Cl	74	0.50	0.54	0.51	0.56	0.73	0.87	0.54	0.58	0.51	0.81
compounds containing H, C, and Br	39	0.50	0.57	0.52	0.60	0.66	0.75	0.39	0.43	0.61	0.71
multihalogen (F, Cl, or Br) hydrocarbons	14	0.76	0.78	0.78	0.80	0.69	0.72	0.68	0.70	0.39	0.46
halogen (F, Cl, or Br) compounds containing H, C, N, and/or O	52	1.30	1.13	1.25	1.08	0.83	0.79	1.13	1.05	1.56	1.67
sulfur compounds not containing P	33	0.61	0.55	0.58	0.54	0.67	0.78	0.57	0.53	0.73	0.93
phosphorus compounds	37	1.23	1.21	1.21	1.22	1.45	1.46	1.43	1.40	1.49	1.71
silicon compounds	2	1.73	1.72	1.72	1.71	1.11	1.15	1.27	1.33	1.63	1.81
iodine hydrocarbons	21	0.39	0.40	0.42	0.42	0.50	0.60	0.38	0.35		
all data	2129	0.54	0.54	0.54	0.55	0.59	0.65	0.59	0.56	0.61	0.67
144 Transfer Energy Data											
hydrocarbons	2	0.29	0.24	0.27	0.23	0.18	0.12	0.34	0.30	0.10	0.79
lactones	10	0.83	0.76	0.82	0.78	0.75	0.73	0.84	0.80	0.87	0.74
aromatic nitrogen heterocycles	7	0.52	0.46	0.50	0.46	0.66	0.69	0.50	0.51	0.44	0.71
bifunctional HCN compounds	4	0.24	0.33	0.24	0.32	0.64	0.61	0.57	0.54	0.53	0.57
amides, ureas, and lactams	28	1.04	0.75	0.99	0.71	0.83	0.80	0.97	0.96	0.76	0.73
thymines or uracils	7	0.63	0.60	0.62	0.60	0.81	0.75	0.76	0.71	0.97	0.94
bifunctional HCNO compounds	5	0.69	0.75	0.69	0.74	0.62	0.58	0.68	0.63	0.54	0.46
halogen (F, Cl, or Br) compounds containing H, C, N, and/or O	38	0.61	0.65	0.62	0.66	0.71	0.71	0.69	0.69	0.73	0.67
sulfur compounds not containing P	21	0.43	0.42	0.43	0.42	0.38	0.38	0.39	0.38	0.44	0.48
phosphorus compounds	9	0.53	0.36	0.44	0.34	0.65	0.79	0.61	0.70	0.80	0.76
silicon compounds	13	0.65	0.65	0.65	0.65	0.49	0.50	0.49	0.49	0.83	0.54
all data	144	0.67	0.60	0.65	0.60	0.65	0.65	0.67	0.67	0.69	0.66

[a]See footnote a in Table 3

In the case of SM12ESP, the CDS term was parametrized by averaging over ten theoretical levels ($M = 10$) including two electrostatic potential schemes (ChElPG and MK) for charges, five density functionals listed above, and one basis set (MG3S). The value of $(\chi/N)^{1/2}$ after optimization equals 0.80 kcal/mol in the case of SM12CM5 and 0.83 kcal/mol in the case of SM12ESP. For comparison, if we assume that $G_{CDS} = 0$, the value of $(\chi/N)^{1/2}$ equals 3.56 and 3.30 kcal/mol, respectively, for SM12CM5 and SM12ESP.

The above procedure gives surface tension coefficients for H, C, N, O, F, Si, P, S, Cl, Br, and I. To estimate the SM12 parameters for the whole periodic table, we performed the following procedure. All the atomic and molecular surface tension coefficients were set to zero, except for $\tilde{\sigma}_Z^{[n]}$ where Z corresponds to H, C, N, O, F, Si, P, S, Cl, Br, or I. These parameters $\tilde{\sigma}_Z^{[n]}$ (one set for water, another for nonaqueous solvents) were then assumed to be independent of Z, i.e., $\tilde{\sigma}_Z^{[n]} = \tilde{\sigma}_0^{[n]}$. The values of $\tilde{\sigma}_0^{[n]}$ were optimized within this two-parameter model by minimizing the error function of eq 11,

Table 5. Mean Unsigned Errors in Solvation Free Energies of Ions (in kcal/mol)[a]

		SM12CM5				SM12ESP					
		B3LYP		M06-2X		MK		ChElPG			
solute class	N	MG3S	6-31G(d)	MG3S	6-31G(d)	B3LYP	M06-2X	B3LYP	M06-2X	SM8	SMD
112 Selectively Clustered Ions in Water											
anions	60	3.7	3.1	3.5	2.9	3.5	3.3	3.8	3.6	3.0	4.6
cations	52	3.5	3.0	3.5	2.9	3.3	3.3	3.5	3.5	2.5	2.9
all data	112	3.6	3.0	3.5	2.9	3.4	3.3	3.7	3.5	2.7	3.8
80 Ions in Methanol											
anions	51	2.3	2.7	2.5	3.0	2.8	2.9	2.8	2.8	3.0	2.1
cations	29	2.3	1.9	2.2	1.8	2.0	1.7	2.5	2.3	3.0	2.2
all data	80	2.3	2.4	2.4	2.6	2.5	2.5	2.7	2.6	2.6	2.2
69 Ions in Acetonitrile											
anions	30	4.0	4.4	4.2	4.7	3.3	3.6	2.8	3.0	4.5	3.9
cations	39	7.7	7.3	7.6	7.2	7.2	7.2	7.3	7.3	7.3	7.8
all data	69	6.1	6.0	6.1	6.1	5.5	5.6	5.3	5.4	6.1	6.1
71 Ions in Dimethyl Sulfoxide											
anions	67	6.3	6.9	6.5	7.0	6.8	7.1	5.9	6.0	7.7	4.7
cations	4	3.1	2.5	3.0	2.5	2.1	2.0	2.5	2.3	1.8	8.4
all data	71	6.1	6.6	6.3	6.8	6.5	6.8	5.7	5.8	7.4	4.9

[a]See footnote *a* in Table 3

yielding values of $(\chi/N)^{1/2}$ equal to 1.48 (SM12CM5) and 1.61 (SM12ESP) kcal/mol. We define these two values of $\tilde{\sigma}_0^{[n]}$ as those to be used for all elements *other* than H, C, N, O, F, Si, P, S, Cl, Br, and I, with $\tilde{\sigma}_{Z_k Z_{k'}} = 0$ (in eq 7) and $\tilde{\sigma}_{Z_k}^{[\alpha]} = \tilde{\sigma}_{Z_k}^{[\beta]} = 0$ (in eq 8) if either Z_k or $Z_{k'}$ is not 1, 6–9, 14–17, 35, or 53. Now our model is defined for the entire periodic table. The optimized values of the SM12 surface tension coefficients are listed in Table 2. For convenience of tabulation, atomic numbers are replaced with atomic symbols. Functional forms for atomic surface tensions used by SM12 are the same as in previous work[17,18] and given in the SI.

Further Details. The SM12 free energy of solvation is calculated using a locally modified version of *Gaussian 09* (Revision A.02)[70] called Minnesota Gaussian Solvation Module[71] or *MN-GSM*. The bulk electrostatic contribution ($\Delta G_{EP} = \Delta E_E + G_P$) to the solvation free energy is calculated using a self-consistent reaction field procedure as follows. First, we obtain the initial set of CM5 or ESP charges using the gas-phase electronic wave function of the solute molecule. Then, using the reaction field operator (see, for example, eq 32 in ref 72) based on the partial atomic charges fixed at the previously estimated values, we run a new reaction field calculation to obtain a new wave function that gives a new set of partial atomic charges (this time, in solution). Then, we repeat the reaction field calculation using a set of charges from the previous iteration until the ΔG_{EP} energy converges within 0.001 kcal/mol. Usually, it takes 4–5 iterations to converge. The ESP charges (i.e., MK or ChElPG) at each iteration are calculated using the keywords *pop=mk* and *pop=chelpg* in the *Gaussian 09*' input file, respectively for MK and ChElPG, whereas a calculation of CM5 charges requires a calculation of Hirshfeld charges[29] with the keyword *pop=hirshfeld*. The Hirshfeld charges are then converted into CM5 charges by running an external program called *CM5PAC*.[73] The CDS contribution is calculated independently of the solute's electronic wave function. CM5 charges will be available in an upcoming revision of *Gaussian 09*.

In the present study, we have also carried out SM8 solvation energy calculations based on the CM4M charge model[24] using

MN-GSM[71] and SMD calculation using the standard version[70] of *Gaussian 09*.

3. RESULTS

Table 3 shows mean unsigned errors ("theory" minus "experiment") for the 374 SM12 solvation free energies of neutral compounds in water by solute class calculated using B3LYP and M06-2X, in comparison with the corresponding SM8/M06-2X/6-31G(d) and SMD/M05-2X[74]/6-31G(d) errors. Table 4 shows the corresponding mean unsigned errors in the 2129 computed solvation free energies of neutral compounds in 90 nonaqueous solvents and in the 144 transfer free energies. Table 5 shows the mean unsigned errors in 332 ionic data. The corresponding mean signed errors by solute class over the whole training set are given in the SI.

4. DISCUSSION

We start with discussion of the performance of SM12 in predicting the solvation free energies of neutral solutes in aqueous solution. Table 3 shows that the SM12 mean unsigned errors over 274 neutral aqueous data used in the SM12 parametrization as well as in our previous works[18,19,21] vary from 0.6 (when CM5 charges are used) to 0.7 (with ChElPG charges) and 0.8 kcal/mol (with MK charges), that is 17–23% relative to the mean unsigned solvation energy (3.5 kcal/mol) obtained by averaging the 274 reference values. In the case of SM12CM5 on average, the errors do not depend significantly on the choice between the MG3S and 6-31G(d) basis set or between the B3LYP and M06-2X density functional (see Table 3) because atomic charges obtained using the CM5 model are only slightly sensitive on average to the choice of basis set or density functional.[25] The SM12 errors are overall similar to the errors in the 274 aqueous solvation free energies (0.6 kcal/mol) computed using our earlier solvation models, SM8 and SMD (Table 3).

Comparing the mean unsigned errors over individual solute classes, we note that the errors vary from 0.14 kcal/mol (for 13 esters with SM12CM5/B3LYP/MG3S) to 1.6–2.0 kcal/mol (for the class of small inorganic molecules containing H_2, NH_3,

Table 6. Solvation Free Energies in Water ($\Delta G^*_{S,w}$, in kcal/mol) and Octanol−Water log $P_{o/w}$ Values for Selected Solutes[a]

| | $\Delta G^*_{S,w}$ | | | | | log $P_{o/w}$ | | | | |
| | SM12 | | | | | SM12 | | | | |
solute name	CM5	CHG	SM8	SMD	exp	CM5	CHG	SM8	SMD	exp
acetylsalicylic acid	−9.5	−8.8	−9.6	−8.9	−9.9	1.9	1.7	1.0	0.6	1.19
5-bromouracil	−15.8	−16.7	−14.8	−13.3	−18.2	−0.3	−0.5	−0.4	−0.1	−0.21
butylparaben	−7.5	−7.7	−7.9	−8.6	−8.7	3.1	2.9	2.6	2.1	3.57
caffeine	−12.4	−11.8	−12.0	−11.4	−12.6	0.7	0.6	0.6	1.2	−0.07
5-chlorouracil	−15.9	−17.3	−14.8	−13.5	−17.7	−0.6	−0.8	−0.6	−0.3	−0.35
6-chlorouracil	−14.4	−15.3	−12.8	−10.8	−15.8					
cyanuric acid	−20.5	−21.6	−17.9	−13.9	−18.1	−1.7	−1.8	−1.4	−0.6	0.61
diflunisal	−11.6	−11.8	−12.2	−13.6	−9.4	3.3	3.3	1.9	1.2	4.44
ethylparaben	−7.9	−8.1	−8.4	−8.9	−9.2	2.2	2.1	1.6	1.1	2.47
5-fluorouracil	−15.7	−17.7	−15.2	−12.9	−16.9	−1.0	−1.2	−1.2	−0.7	−0.89
flurbiprofen	−7.2	−7.6	−7.5	−8.2	−8.4	4.1	4.1	3.5	3.0	4.16
hexachlorobenzene	−1.1	−1.9	−0.1	1.7	−2.3	4.6	4.4	4.6	4.9	5.73
ibuprofen	−5.4	−5.5	−5.2	−6.1	−7.0	3.6	3.8	3.4	3.3	3.97
5-iodoracil	−16.1	−16.9			−18.7	−0.2	−0.4			0.04
ketoprofen	−9.7	−9.0	−10.1	−11.1	−10.8	3.6	3.7	3.1	2.2	3.12
methylparaben	−8.3	−7.9	−8.6	−8.7	−9.5	1.6	1.4	0.9	0.5	1.96
naproxen	−9.3	−9.0	−8.7	−9.7	−10.2	2.8	2.8	2.6	2.0	3.18
4-nitroaniline	−9.0	−8.7	−9.2	−8.4	−9.5	0.7	0.5	0.6	0.4	1.39
octafluorocyclobutane	3.7	2.4	4.0	3.9	3.4	2.6	2.1	2.4	2.7	2.29
pentachloronitrobenzene	−3.0	−2.7	−0.9	2.3	−5.2	3.4	3.5	4.1	4.3	4.64
phthalimide	−12.0	−12.3	−10.7	−8.3	−9.6	0.2	0.2	0.6	0.4	1.15
propylparaben	−7.7	−8.3	−8.1	−8.7	−9.4	2.7	2.5	2.1	1.6	3.04
sulfolane	−7.1	−8.2	−13.2	−12.6	−8.6	−0.2	−0.6	−0.5	−1.0	−0.77
5-(trifluoromethyl)uracil	−14.8	−16.0	−14.3	−12.4	−15.5	−0.3	−0.5	−0.6	−0.1	0.04
trimethyl orthotrifluoroacetate	−2.1	−0.8	−1.0	−0.7	−0.8					
uracil	−16.3	−17.8	−15.2	−13.5	−16.6	−1.1	−1.4	−1.1	−0.8	−1.07
MSE	0.7	0.3	0.9	1.6		−0.3	−0.5	−0.6	−0.7	
MUE	1.4	1.3	1.6	2.3		0.6	0.6	0.7	0.9	

[a]These compounds are part of the SAMPL2 test set (ref 45). The reference values of aqueous solvation free energies are from ref 45, and the reference log P values are from ref 49. The notations MSE and MUE stand for mean signed and mean unsigned error, respectively. The reported theoretical numbers were calculated using SM12/M06-2X/MG3S based on CM5 charges and SM12/M06-2X/MG3S based on ChElPG (CHG) charges as well as using SM8/M06-2X/6-31G(d) and SMD/M05-2X/6-31G(d). Iodoracil is not included in the SM8 and SMD statistics because the models do not have parameters to treat iodine-containing compounds.

H_2O, and the water dimer in water). The latter value is dominated by the errors in the predicted solvation energies of H_2O and the water dimer. The aqueous solvation energy of H_2O predicted with any of the methods in Table 3 (including SM8 and SMD) is overestimated (i.e., overly negative) by 1−2 kcal/mol relative to the reference value,[42] −6.3 kcal/mol, and the aqueous solvation energy of $(H_2O)_2$ is also overestimated, by 3−4 kcal/mol relative to the reference value,[42] −11.3 kcal/mol. For example, the solvation free energy of water in water computed by SM12/M06-2X/MG3S using CM5 charges is equal to −8.14 kcal/mol where the bulk electrostatic contribution (ΔG_{EP}) equals −5.27 and the nonbulk-electrostatic (or CDS) contribution (G_{CDS}) equals −2.87 kcal/mol. At the same level of theory, the solvation free energy of the water dimer in water is equal to −14.65 kcal/mol with $\Delta G_{EP} = −9.79$ and $G_{CDS} = −4.86$ kcal/mol. The somewhat poor performance of our implicit solvent models on $(H_2O)_2$ and H_2O in water (with a relative error of 30% in each case) can be possibly improved by considering more than one explicit solvent (H_2O) molecule and by then doing a Boltzmann average of the free energies of the resulting solute−solvent clusters surrounded by continuum solvent. Alternatively, we could have given a higher weight to these data during the optimization.

The SM12 mean unsigned errors over 66 aqueous data (that come primarily from the SAMPL1 training set[44]) and over 26 (SAMPL2[45]) aqueous data vary from 1.1 to 1.5 kcal/mol or from 14 to 19% relative to the mean unsigned solvation energy averaged over the 92 reference values (7.8 kcal/mol). The corresponding SM8 and SMD errors are larger (between 1.6 and 2.3 kcal/mol) on these data because many of these solutes are compounds with complex functionalities (for instance, agricultural pesticides and/or compounds with oxidized sulfur and phosphorus functionalities) which were not adequately represented in the old training set[18,21] and, therefore, not included in the parametrization of SM8 or SMD. Table 3 shows that the SM12 errors on 8 aqueous iodine-containing solutes are relatively small (0.3−0.6 kcal/mol). Unlike SM12, the older (SM8 and SMD) models were not parametrized to treat iodine-containing solutes.

The SM12 mean unsigned errors for the 2129 solvation free energies of neutral solutes in 90 organic solvents (Table 4) vary from 0.5 to 0.7 kcal/mol, that is 9−13% of the mean unsigned solvation energy averaged over 2129 reference values (5.3 kcal/mol). The corresponding SM8 and SMD errors are similar (0.6−0.7 kcal/mol). The largest relative errors are found for 29 data in the already mentioned class of H_2, NH_3, and H_2O, for the 5 data that include hydrazine and methylhydrazine and for

the two data that include tetramethylsilane in hexadecane and octanol. The mean unsigned errors in the 144 transfer free energies for all of the models listed in Table 4 are in the range of 0.6–0.7 kcal/mol.

For 112 aqueous ions, the SM12 mean unsigned errors do not exceed 4 kcal/mol, and they are similar to the SM8 and SMD errors listed in Table 5. The SM12 mean signed errors listed in the SI range from −0.2 to 1.6 kcal/mol for anions and from 2.1 to 3.1 kcal/mol for cations. For 80 ions in methanol, the errors are no larger than for ions in water. The anions in methanol are oversolvated by 1–2 kcal/mol, whereas the cations are typically slightly undersolvated. Larger errors are observed for 69 ions in acetonitrile and 71 ions in dimethyl sulfoxide: 5–7 kcal/mol for SM12, 6–7 kcal/mol for SM8, and 5–6 kcal/mol for SMD. In general, the errors in the solvation free energies of ions are larger than those for neutrals because the absolute value of the solvation free energy of a charged solute is much larger on average. For example, the mean solvation free energy averaged over 332 reference values for ions equals −65 kcal/mol (all the 332 energies are negative).

Table 6 shows predicted solvation free energies and octanol–water log P values for 26 drug-like or biologically important compounds from the SAMPL2 data set.[45] These data were used in the SM12 parametrization (but they were not used in the SM8 or SMD parametrization). The SM12 model based on CM5 charges yields results that are very close to those obtained by SM12ESP using ChElPG charges. The mean unsigned errors in the aqueous solvation free energies vary from 1.3 or 1.4 kcal/mol for SM12 to 1.6 for SM8 and 2.3 kcal/mol for SMD. The computed log P values are underestimated on average by 0.3–0.7 log units against available experimental data,[49] though theory predicts a correct sign of log P in 20 out of 24 cases (experimental log P values are not available for two out of 26 compounds).

Table 7 shows predicted dipole moments in the gas phase and in water for several nucleic acid bases (namely, adenine,

Table 7. Dipole Moments (in debyes) for Selected Solutes[a]

	CM5		ChElPG		CM4M		density-based	
	gas	water	gas	water	gas	water	gas	water
adenine	2.2	2.9	2.4	3.3	2.2	3.4	2.4	3.5
cytosine	6.4	7.9	6.7	8.7	6.3	8.9	6.6	9.7
guanine	6.1	7.8	6.5	8.4	6.2	9.0	6.5	9.4
thymine	4.1	4.9	4.4	5.6	4.1	5.9	4.4	6.4
uracil	4.2	5.1	4.5	5.8	4.0	5.9	4.4	6.4

[a]See Section 4.

cytosine, guanine, thymine, and uracil) studied in earlier work.[75] In these calculations, we use the M06-2X/MG3S molecular geometries optimized in the gas phase. The gas-phase CM5, ChElPG, and CM4M dipole moments were calculated from the corresponding partial atomic charges obtained using the gas-phase electronic wave function. The liquid-phase CM5, ChElPG, and CM4M dipole moments were calculated using the corresponding charges obtained self-consistently within the GB approach (i.e., using the SM12 or SM8 solvation model). The gas-phase density-based dipole moment was obtained using the gas-phase electronic wave function of the corresponding solute molecule, and the liquid-phase density-based dipole moment was obtained self-consistently from an SMD calculation. The CM4M charges were obtained at the M06-2X/6-31G(d) level of theory,

whereas the other calculations were performed using M06-2X/MG3S. The density-based dipole moments of the studied molecules in aqueous solution increase by 45–47% relative to the gas phase, whereas the dipole moments derived from partial atomic charges increase by 20–32% (CM5), by 27–38% (ChElPG), and by 41–55% in the case of CM4M. One can say that the CM5 charges are less polarizable than the CM4M charges, though it is difficult to say how big these polarization effects should be because the molecular dipole moment in solution is not uniquely defined in terms of experimental observation.

The free energies of solvation have been calculated in the present work at fixed, gas-phase optimized geometries of solute molecules, ensuring that ΔG_N in eq 1 is equal to zero. The rationale for using rigid gas-phase geometries for parametrizing continuum solvation models has been discussed in previous works,[13] in which we found that when gas-phase geometries were replaced by solution-phase geometries, the mean deviation was only about 0.1 kcal/mol for neutral solutes and 0.5 kcal/mol for ions. Thus the approximation of using rigid, gas-phase geometries is usually acceptable.

The SM12 solvation model is defined for the entire periodic table by estimating the CDS parameters $\tilde{\sigma}_X^{[n]}$ (Table 2) where X refers to any element other than those already present in the SM12 training set (i.e., $X \neq$ H, C, N, O, F, Si, P, S, Cl, Br, and I). We tested the SM12 model for predicting the octanol–water log P value of phenylboronic acid, $PhB(OH)_2$. The log P value of $PhB(OH)_2$ computed at the SM12/M06-2X/MG3S//gas/M06-2X/MG3S level is equal to 1.27 and 1.40 with CM5 and ChElPG partial atomic charges, respectively. If we set $\tilde{\sigma}_X^{[n]}$ to zero (totally neglecting the CDS contribution from X, as would be done in the context of our earlier solvation models, for example, SM8[18]), the predicted log P value of phenylboronic acid would be 1.14 (CM5) and 1.24 (ChElPG). For comparison, the corresponding experimental value is equal to 1.58.[49]

5. SUMMARY

This article presents a new self-consistent reaction-field universal continuum solvent model called SM12 which is based on the generalized Born approximation using CM5 charges calculated from a quantum mechanical electron density or from ESP charges derived from quantum-mechanically calculated electrostatic potentials (for example, ChElPG and Merz–Kollman charges). The model has three key advantages over previous generalized Born models: (i) it can be used with any basis set, even extended ones, (ii) the model is parametrized over a more diverse training set, and (iii) it is defined for the entire periodic table. The model was parametrized using the B3LYP, mPW1PW, M06-L, M06, and M06-2X density functionals against 2979 reference experimental data including 2503 solvation free energies and 144 transfer free energies of neutral as well as 332 solvation free energies of singly charged anions and cations in water and organic solvents. The mean unsigned error averaged over the five density functionals used in combination with the MG3S and 6-31G(d) basis sets and CM5 charges equals 0.58 kcal/mol for 2647 neutral data and 4.3 kcal/mol for 332 ionic data. The corresponding errors over the five density functionals with the MG3S basis set and ChElPG charges equals 0.62 kcal/mol for neutrals and 4.2 kcal/mol for ions, and, if one uses Merz–Kollman charges, the corresponding errors equal 0.65 and 4.3

kcal/mol, respectively, for neutrals and for ions. The errors are quite small compared to competing methods.[3]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Reference solvation free energies and reference transfer energies, values of the solvent descriptors used in the parametrization, complementary tables with the errors in the computed solvation and transfer free energies over the training set, and the functional forms for the G_{CDS} contribution to the free energy of solvation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

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

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

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

(2) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3094.

(3) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760−768. Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2009**, *42*, 493−497.

(4) Smith, P. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *98*, 13735−13740.

(5) Boresch, S.; Archontis, G.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 25−33.

(6) Born, M. *Z. Phys.* **1920**, *1*, 45−48.

(7) Klopman, G. *Chem. Phys. Lett.* **1967**, *1*, 200−202.

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

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

(10) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129−152.

(11) Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817−4826.

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

(13) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117−9133; errata: **1999**, *111*, 5624; and **2000**, *113*, 3930.

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

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

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

(17) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133−1152.

(18) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011−2033.

(19) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2447−2464.

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

(21) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.

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

(23) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.

(24) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2046−2054.

(25) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.

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

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

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

(29) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129−138.

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

(31) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129−145.

(32) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.

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

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

(35) Tuñón, I.; Ruiz-López, M. F.; Rinaldi, D.; Bertrán, J. *J. Comput. Chem.* **1996**, *17*, 148−155.

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

(37) Mantina, M.; Valero, R.; Cramer, C. J.; Truhlar, D. G. In *CRC Handbook of Chemistry and Physics*, 91st ed. (**2010−2011**); Haynes, W. M., Ed.; CRC Press: Boca Raton, FL, 2010; pp 9-49−9-50.

(38) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Duce, P. P.; Morris, J. J.; Taylor, P. J. *J. Chem. Soc., Perkin Trans. 2* **1989**, 699−711.

(39) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73−83.

(40) Abraham, M. H. *J. Phys. Org. Chem.* **1993**, *6*, 660−684.

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

(42) 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 2012*; University of Minnesota: Minneapolis, MN, 2012. http://comp.chem.umn.edu/mnsol (accessed November 26, 2012).

(43) 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*, 769−779.

(44) Guthrie, J. P. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.

(45) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 259−279.

(46) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066−16081.

(47) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 408−422.

(48) Shoeib, M.; Harner, T. *Environ. Toxicol. Chem.* **2002**, *21*, 984−990.

(49) *Physical/Chemical Property Database (PhysProp)*; SRC Environmental Science Center: Syracuse, NY, 2012. http://www.syrres.com (accessed April 4, 2012).

(50) Zeng, X.-L.; Wang, H.-J.; Wang, Y. *Chemosphere* **2012**, *86*, 619−625.

(51) Toulmin, A.; Wood, J. M.; Kenny, P. W. *J. Med. Chem.* **2008**, *51*, 3720−3730.

(52) Jiang, X.; Xia, Z.; Deng, L.; Wei, W.; Chen, J.; Xu, J.; Li, H. *Chromatographia* **2012**, *75*, 347−352.

(53) 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 September 21, 2012).

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

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

(56) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192−196.

(57) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(58) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(59) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 1384−1388. The MG3S basis is available at the following website: http://comp.chem.umn.edu/basissets/basis.cgi (accessed September 21, 2012)

(60) Environmental Molecular Sciences Laboratory (EMSL) Basis Set Library at https://bse.pnl.gov/bse/portal (accessed September 24, 2012).

(61) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113−11123.

(62) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(63) Zheng, J.; Xu, X.; Truhlar, D. G. *Theor. Chem. Acc.* **2011**, *128*, 295−305.

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

(65) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3110.

(66) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

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

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

(69) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101/1−18.

(70) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*; Gaussian, Inc.: Wallingford, CT, 2009.

(71) Olson, R. M.; Marenich, A. V.; Chamberlin, A. C.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Hawkins, G. D.; Winget, P.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM, version 2012*; University of Minnesota: Minneapolis, MN, 2012.

(72) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Guido, C. A.; Mennucci, B.; Scalmani, G.; Frisch, M. *J. Chem. Sci.* **2011**, *2*, 2143−2161.

(73) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *CM5PAC*; University of Minnesota: Minneapolis, MN, 2011. http://comp.chem.umn.edu/cm5pac (accessed September 25, 2012)

(74) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.

(75) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74−80.