JCTC Journal of Chemical Theory and Computation

Free Energies of Solvation with Surface, Volume, and Local Electrostatic Effects and Atomic Surface Tensions to Represent the First Solvation Shell

Junjun Liu,[†,‡] Casey P. Kelly,[§] Alan C. Goren,[‡,||] Aleksandr V. Marenich,[§] Christopher J. Cramer,*,[§] Donald G. Truhlar,*,[§] and Chang-Guo Zhan*,[‡]

Key Laboratory of Pesticide & Chemical Biology of Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, People's Republic of China, Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 725 Rose Street, Lexington, Kentucky 40536, Department of Chemistry and Supercomputing Institute, 207 Pleasant Street SE, University of Minnesota, Minneapolis, Minnesota 55455, and Division of Natural Sciences & Mathematics, Transylvania University, 300 North Broadway, Lexington, Kentucky 40508

Abstract: Building on the SVPE (surface and volume polarization for electrostatics) model for electrostatic contributions to the free energy of solvation with explicit consideration of both surface and volume polarization effects, on the SM*x* approach to including first-solvation-shell contributions, and on the linear relationship between the electric field and short-range electrostatic contributions found by Chipman, we have developed a new method for computing absolute aqueous solvation free energies by combining the SVPE method with semiempirical terms that account for effects beyond bulk electrostatics. The new method is called SMVLE, and the elements it contains are denoted by SVPE-CDSL, where SVPE denotes accounting for bulk electrostatic interactions between solute and solvent with both surface and volume contributions, CDS denotes the inclusion of solvent cavitation, changes in dispersion energy, and possible changes in local solvent structure by a semiempirical term utilizing geometry-dependent atomic surface tensions as implemented in SM*x* models, and L represents the local electrostatic effect derived from the outward-directed normal electric field on the cavity surface. The semiempirical CDS and L terms together represent the deviation of short-range contributions to the free energy of solvation from those accounted for by the SVPE term based on the bulk solvent dielectric constant. A solute training set containing a broad range of molecules used previously in the development of SM6 is used here for SMVLE model calibration. The aqueous solvation free energies predicted by the parametrized SMVLE model correlate exceedingly well with experimental values. The square of the correlation coefficient is 0.9949 and the slope is 1.0079. Comparison of the final SMVLE model against the earlier SM*x* solvation model shows that the parametrized SMVLE model not only yields good accuracy for neutrals but also significantly increases the accuracy for ions, making it the best implicit solvation model to date for aqueous solvation free energies of ions. The semiempirical terms associated with the outward-directed electric field account in a physical way for the improvement in the predictive accuracy for ions. The SMVLE method greatly decreases the need to include explicit water molecules for accurate modeling of solvation free energies of ions.

Introduction

Dielectric continuum solvation models[1,2] have been widely and successfully used for estimating solvation free energies. Such models are also called implicit solvation models because the solvent is not atomistically explicit but rather is implicit in the dielectric medium. In the self-consistent versions of such models, the solvent is considered to be a continuous dielectric medium that is polarized by the solute, leading to a reaction field that in turn polarizes the solute, which changes the solvent polarization, leading ultimately to a self-consistent reaction field (SCRF). A dielectric continuum solvation model accurately describes the long-range permanent-multipole-moment and inductive interactions between solute and bulk solvent; in the language conventionally used in the continuum solvation model literature, both permanent-multipole-moment and induction effects are labeled as electrostatic, and we will follow that convention in the rest of this article. SCRF methods require less computational effort than explicit-solvent approaches involving the same quality treatment of the solute, and this makes them appealing for the study of complex chemical, materials, and biochemical processes and for the rapid screening of many solutes in, for example, molecular docking studies. When the bulk-solvent model is augmented with additional terms to account for the deviation of short-range solute−solvent interactions from the bulk electrostatic model, useful accuracy can be obtained.[3]

One popular way to implement the SCRF approach is to describe the solvent polarization in terms of the electrostatic potential that it introduces inside the solute cavity under the assumption that all solute charge density resides inside the cavity; this is often called the polarized continuum model (PCM).[1,2] However, unconstrained quantum mechanical calculations of solute electronic structure always lead to a tail of the wave function penetrating outside the cavity, thereby causing an additional polarization effect called volume polarization.[4] It has been demonstrated[5] that neglecting charge penetration (also called outlying charge) leads to inconsistencies in the course of solving Poisson's equation. Such inconsistencies render many SCRF implementations sensitive to cavity size[6] and prone to overestimating solvent shifts of energy barriers in aqueous solution.[7,8] In previous studies,[5−7,9,10] a general model called surface and volume polarization for electrostatic interaction (SVPE), or the fully polarizable continuum model (FPCM),[11−18] was developed. This model, implemented for irregularly shaped solute cavities, fully accounts for both surface and volume polarization effects in solute−solvent electrostatic interactions. Therefore, the SVPE solvation model provides a theoretically well justified continuum methodology for studying long-range electrostatic interactions. It has also been useful in a practical sense, having been applied successfully to study

mechanisms for various chemical reactions and to make pK_a predictions.[7,8,19,20]

One must bear in mind that absolute solvation free energies result not only from long-range electrostatic interactions between solute and bulk solvent but also from significant short-range contributions, such as short-range and nonbulk electrostatics, as well as cavitation, exchange repulsion, dispersion, and disruption or formation of the nearby solvent structure. (Note that the short-range, nonbulk electrostatic effect may be considered to be a solvent structure effect.) These interactions are not treated satisfactorily within the framework of a pure dielectric continuum model. Previously, the difference between solvation free energies calculated by dielectric continuum solvation methods and experimental solvation free energies has been labeled in an SVPE context as the nondielectric or short-range contribution[21,22] and in other contexts[3] as a cavity-dispersion-solvent-structure (CDS) effect. Such short-range contributions are often[7,8] (but not always[23]) neglected in estimating energy barriers by implicit solvation methods, but for systems with strong hydrophobic effects or hydrogen bonding between solute and solvent molecules, the short-range contributions to the energy barriers may be very significant or even decisive. Furthermore, accounting for the short-range contributions is essential for calculating reliable absolute solvation free energies of neutral solutes.[3]

In a supermolecular approach, the nearby solvent molecules are represented explicitly as components of a cluster continuing the solute. It has been reported that, by employing a combined supermolecule-continuum approach,[15,18,24−27] involving both explicit and implicit solvent, the SVPE method can account for short-range contributions between solute and solvent and hence more accurately predict the absolute solvation free energies for a series of charged chemical species, including H^+, Li^+, OH^-, e^- (the hydrated electron), and F^-.[15,18,24,27] The supermolecule-continuum approach explicitly treats a portion of the solvent surrounding the solute at a high quantum mechanical level. The computational efficiency depends heavily on solute size and larger solutes require more solvent water molecules to be explicitly considered, thus in practice limiting the application to smaller solutes. Even more significantly, the proper application of the supermolecule-continuum method requires statistical mechanical averaging over the various possible sites and orientations of the explicit solvent molecules, and this become impractical for even a small number (for example, two or three) of explicit solvent molecules. Finally, one should note that the interaction of solvent molecules in the supersolute (cluster) with the continuum must be treated accurately, for example, through recourse to very high levels of electronic structure theory.

For these reasons, much effort has been made to augment the dielectric continuum model with short-range contributions.[1,2,28] For example, the SM*x* series of solvation models (with *x* being 1−6, 8, 8AD, or D)[29−46] augment and correct the bulk electrostatic portion, obtained by the generalized Born (GB) approximation[29,47] (for $x = 1−6$, 8, or 8AD) or the PCM approximation[2] (for $x = $ 5C or D), with a semiempirical term that accounts for short-range contribu-

* Corresponding authors: e-mail cramer@umn.edu (C.J.C.); truhlar@umn.edu (D.G.T.); zhan@uky.edu (C.-G.Z.).
† Central China Normal University.
‡ University of Kentucky.
§ University of Minnesota.
‖ Transylvania University.

tions. By employing a training data set containing a broad range of solutes, atomic radii used for defining the cavity in electrostatic calculations (such radii are called intrinsic Coulomb radii) were calibrated to calculate the bulk electrostatic interactions, and a set of atomic surface tension parameters was optimized to calculate the short-range contributions. It has been shown[3,43,44] that the accuracy of SMx models for predicting absolute aqueous solvation free energies is quite good, about ~0.5 kcal/mol for neutral solutes. Although SMx models, like all other solvation models, have larger absolute errors for predicting aqueous solvation free energies of ions, which are much larger than those of neutral solutes, SMx still outperforms other continuum models for ionic solvation free energies.[3,48] While the SMx models, by implicitly including local electrostatic effects as part of the semiempirical CDS terms, provide significant improvement over PCM models in predicting absolute solvation free energies,[43−45] it is worthwhile to consider more explicit ways to include local electrostatic effects.

The long-range electrostatic contribution, which is a function only of the solvent's bulk dielectric constant, is included in the bulk electrostatic term, but the bulk electrostatic term also includes a somewhat arbitrary approximation to the short-range electrostatic effect, because the solute−solvent boundary that surrounds the solute cavity is located within the region occupied by the first solvation shell, but this shell does not behave like a bulk dielectric. The deviation of short-range electrostatics from bulk electrostatics (this is called the nonbulk electrostatic contribution or the local electrostatic contribution) has previously been included in SMx models as a solvent-structure contribution to the CDS term. Here we incorporate a new function, denoted by L for local electrostatics, that treats the nonbulk electrostatic contribution explicitly.

The motivation for the new function is the observation that the short-range electrostatic contribution is linearly correlated with the maximum or minimum outward-directed normal electric field on the solute−solvent boundary surface.[21,22] Therefore we make the contribution beyond the bulk part calculated with SVPE dependent not only on the solvent-accessible atomic surface area of the solute, as in the CDS terms, but also on the outward-directed normal electric field on the cavity surface; the combination of the new local electrostatic terms (L) with the atomic surface tension terms (CDS) is denoted CDSL. Replacing the CDS terms by CDSL terms and replacing the generalized Born approximation (of SM1−SM8 or SM8AD)[3,29−39,41−45] or the PCM approximation (of SM5C and SMD)[40,46] by the SVPE treatment is the essence of the present attempt to make the solvation model more physical. As we have just explained, a long abbreviation for the new method is SVPE-CDSL. It will, however, be more convenient to simply call the new method SMVLE, which denotes solvation model with volume and local electrostatics, since the explicit accounting for volume polarization and local electrostatics are the new elements beyond those included in previous SMx solvation models.

The same training set that was used to calibrate parameters for SM6[43] is used here for optimizing the parameters of SMVLE. The absolute aqueous solvation free energies obtained by the parametrized SMVLE model are compared with experimentally measured aqueous solvation free energies to calculate the mean unsigned error (MUE), which measures the predictive accuracy of the solvation model. The accuracy of the SMVLE model is compared with SM6[43] and the recent SM8[44] and SMD[46] solvation models, and the role of the new kind of semiempirical term, called G_L, is discussed.

Methods

Description of the SMVLE Model. As explained in the Introduction, the free energy of solvation is a sum of three terms:

$$\Delta G_S^* = \Delta G_{SVPE} + G_{CDS} + G_L \tag{1}$$

Here ΔG_S^* is the fixed-concentration absolute solvation free energy,[49] ΔG_{SVPE} is the bulk electrostatic portion calculated by the SVPE method, G_{CDS} is the semiempirical term based on atomic surface tensions, and G_L is the semiempirical electric-field-dependent term, whose form is motivated by Chipman's work on ions[21,22] where two semiempirical terms were generated for anions and cations, respectively. If the standard-state solvation free energy, ΔG_S°, with a concentration corresponding to a solute partial pressure of 1 atm in the gas phase and a solute concentration of 1 M in the liquid phase, is desired instead of ΔG_S^*, then another term, $G_{conc}^\circ = 1.89$ kcal/mol, must be added to account for the change in concentration.[50] This value of G_{conc}° and all other free energies considered in this paper correspond to a temperature of 298 K. ΔG_{SVPE} can be expressed as

$$\Delta G_{SVPE} = \langle \Psi^{(1)} | H^0 + \tfrac{1}{2} V | \Psi^{(1)} \rangle - \langle \Psi^{(0)} | H^0 | \Psi^{(0)} \rangle \tag{2}$$

where Ψ is the solute wave function, H^0 is the solute Hamiltonian in vacuum, and V is the energy operator associated with the reaction field. The factor of $1/2$ in eq 2 stems from assuming a linear response of the surrounding medium to the solute's charge distribution so that half of the induced favorable solute−solvent interaction is canceled by the cost of reorganizing the solvent.[51] The superscripts (0) and (1) refer to the gas-phase isolated molecule and the liquid-phase solution, respectively.

The G_{CDS} term includes free energy changes associated with solvent cavitation, changes in dispersion energy, and possible changes in local solvent structure. It is calculated according to[29,43]

$$G_{CDS} = \sum_{\text{atoms } k} \sigma_k A_k \tag{3}$$

where A_k is the solvent-accessible surface area[52,53] of atom k, which depends on the solute geometry, atomic van der Waals radius, and solvent radius, and σ_k is the atomic surface tension of atom k. The physical basis for eq 3 is that deviations from bulk electrostatics, as well as cavitation,

1112 *J. Chem. Theory Comput., Vol. 6, No. 4, 2010*

Liu et al.

dispersion, and solvent-structural contributions, are all concentrated in the first solvation shell, and A_k is basically a continuous measure of the amount of solvent in the first solvation shell of atom k.[29,52,54] The atomic surface tensions are sensitive to local environment, and therefore they are computed according to

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{\text{atoms } k'} \sum_{m=1}^{M} \tilde{\sigma}_{Z_k Z_{k'}}^{(m)} T_{kk'}^{(m)} \quad (4)$$

where $\tilde{\sigma}_{Z_k}$ and $\tilde{\sigma}_{Z_k Z_{k'}}$ are the semiempirical surface tension coefficients for atom k and atom pair kk', and $T_{kk'}$ is a geometry-dependent switching function called a cutoff tanh (referred to as a COT).[34] In most cases the sum over m has only one term, and when $m = 1$, the superscript is omitted. For k, $k' = 7$, 6, we have $M = 3$.

The remaining term G_L is motivated by the work of Chipman, who found that more accurate solvation energies could be obtained for ions by adding the following terms to the bulk-electrostatic term:[21,22]

$$G_L \text{ (for anions)} \approx W_{\text{anion}}(E^{\min} - E_0^{\min}) \quad (5)$$

$$G_L \text{ (for cations)} \approx W_{\text{cation}}(E^{\max} - E_0^{\max}) \quad (6)$$

where G_L(for anions) and G_L(for cations) are "local" (or short-range) electrostatic contributions for anions and cations, respectively; $E^{\min}$ and $E^{\max}$ are the minimum and maximum outward-directed normal electric field on the cavity surface, respectively; and W_{anion}, $E_0^{\min}$, W_{cation}, and $E_0^{\max}$ are fitting parameters. One may interpret these terms as corrections for local electrostatics. Equations 5 and 6 reveal a linear relationship between the local electrostatic contribution and the minimum or the maximum outward-directed normal electric field on the cavity surface for anions and cations, respectively, and the existence of this relationship indicates that a local (short-range) electrostatic interaction between solute and solvent can physically be modeled in term of the outward-directed normal electric field. However, it is not straightforward to generalize the above linear relationships into a formula valid for all solutes, including neutrals and ions with either sign of the charge. For example, one cannot simply combine eqs 5 and 6 because that ignores the local electrostatic effects for neutrals and zwitterions, and therefore the solvation free energy would not vary smoothly along a reaction coordinate where charge is developed or neutralized. The quantities $E^{\min}$ and $E^{\max}$ are also not adequate to represent the local electrostatic effects for dianions or dications with separated charge centers where both the minimum/maximum and the second minimum/maximum normal electric field should be considered. Furthermore, $E^{\min}$ and $E^{\max}$ do not necessarily vary smoothly during geometry optimization. Therefore, a more complicated functional form that does not have these disadvantages is required. We obtain such a function by summing over terms involving the normal electric fields on each surface node and by taking advantage of the properties of the COT function. In particular, we postulate that

$$G_L = \sum_{i=1}^{I} \left\{ B_i \left[\sum_{m=1}^{M} T(A_i, x_i, -E_m) E_m^{\ i} w_m \right] + B_{i+1} \left[\sum_{m=1}^{M} T(A_i, x_i, E_m) E_m^{\ i} w_m \right] \right\} \quad (7)$$

$$T(A_i, x_i, E_m) = \frac{1 + \tanh \left[A_i (E_m - x_i) \right]}{2} \quad (8)$$

where E_m is the outward-directed normal electric field at node m on the cavity surface; w_m is the surface area of node m; M is the total number of surface nodes used ($M = 1202$ in the present study); I is an integer that represents the highest power of E_m; tanh is the hyperbolic tangent; and B_i, A_i, and x_i are semiempirical parameters.

In eq 7, the summation over all surface elements means that we consider the local electrostatic contribution not only for ions but also for neutral solutes and zwitterions, and all normal electric fields, instead of only one minimum or one maximum normal electric field for singly charged ions, are considered for all kinds of uncharged and charged systems. In addition, the high powers of E_m allow significant nonlinearity in the relationship. The functional form of eq 7 varies smoothly along a reaction coordinate.

Computational Details. The optimized geometries and the corresponding experimental aqueous solvation free energies were obtained from the data set used to calibrate the parameters in the development of SM6,[43] with two exceptions. One exception is that one neutral molecule (*O*-ethyl *O'*-4-bromo-2-chlorophenyl *S*-propyl phosphorothioate or profenofos) in the SM6 training set could not be treated by the SVPE program because of its irregular molecular shape. Thus the SM6 data set has 273 neutral solutes and 143 ions (416 data), and the SMVLE training set has 272 neutral solutes and 143 ions (415 data). All of these molecules and their experimental aqueous solvation free energies are provided as Supporting Information. The second exception concerns the data used for the solvation energies of the ions. Most experimental aqueous solvation free energies are calculated on the basis of thermodynamic cycles, in which the solvation free energies are based on the absolute aqueous solvation free energy of the proton, denoted by $\Delta G_S(H^+)$.[55] The parametrization of SM6 was based on the $\Delta G_S(H^+)$ value of Zhan and Dixon[18] of -264.3 kcal/mol, whereas the later SM8,[44] SM8AD,[45] and SMD[46] were based on the $\Delta G_{\text{sol}}(H^+)$ value of Tissandier et al.[56] of -265.9 kcal/mol. We used the value of Tissandier et al. for the present work.

All the solvation calculations with the SMVLE model were carried out at the HF/6-31+G* electronic structure method.

We previously[43] concluded that the partial charges in some ions are so large that they should be treated by a supermolecule-continuum approach. Therefore, we developed a procedure based on the criterion that if any atom of the ion has partial atomic charge greater than or equal to the partial atomic charge on oxygen in a water molecule, then the ion should be treated as a supermolecule consisting a cluster of a bare ion and one solvent molecule. The ionic data set started with 112 bare ions, and by this criterion, 31 of them should be clustered (so the unclustered instances of these 31 ions are called improperly unclustered). This gives three

sets of ions to consider: 81 properly unclustered (PU) ions, 31 improperly unclustered (IU) ions, and 31 monohydrated (MH) ions. The collection of 81 PU and 31 MH ions (total 112 ions) is called the selectively clustered (SC) set; the collection of 81 PU and 31 IU ions (total 112 ions) is called the unclustered (UC) set; and the collection of 81 PU, 31 IU, and 31 MH ions (total 143 ions) is called "all ions". We used all 143 ions set for parametrization, but we give statistics for various subsets for discussion purposes.

The bulk electrostatics were calculated by the SVPE method by using a local version[10] of Gaussian03.[57] The SVPE results depend only on the level and basis set of the quantum mechanical calculation and the isodensity contour value that defines the solute cavity. All the solvation calculations with the SMVLE model were carried out at the HF/6-31+G* level of theory. Previous studies have shown that contour values in the range of 0.0005−0.002 atomic unit lead to a satisfactory description of the electrostatic contributions to the solvation energies for many neutral[6,9] and ionic[21,22] solutes. For this reason, we chose 0.001 atomic unit as the contour value to determine the dielectric cavity. Cavity surface interactions were calculated from a set of 1202 Lebedev grid points and weights that are expected to yield precision of 0.1 kcal/mol or less for SVPE contributions to solvation free energies of all the solutes examined. Surface areas for the CDS term of eq 3 were calculated by the ASA algorithm[58] with the values of Bondi[59] for the atomic radii.

The molecules studied here are generally rigid, except for methyl rotors, whose conformation does not have a large effect on solvation free energies, and consequently no attempt was made to account for relaxation of geometry, change of conformation, or change in vibrational frequencies upon solvation. For solvation free energy calculations, we adopt the Ben-Naim convention[49] that the solute is transferred from a fixed position in the gas phase to a fixed position in solvent (this is called the fixed-concentration solvation free energy above). A value of 78.5 for the dielectric constant value of water is used in all our solvation calculations, which nominally corresponds to 298 K.

Calibration. After the bulk electrostatic interactions between solute and solvent were accurately determined for each molecule with the SVPE method, a set of target short-range contributions were obtained from the difference between bulk electrostatic interactions and experimental aqueous solvation free energies. All the CDSL parameters were then subjected to a fitting routine. First the atomic surface tension coefficients were optimized to minimize the root-mean-square error (RMSE) over the 272 neutral solutes. Then the semiempirical parameters in eq 7 were optimized against all 415 molecules (272 neutral solutes and 143 ions) with the atomic surface tension coefficients frozen. The optimization of A_i and x_i was carried out in steps:

(1) First I was set temporarily to 1, and values from 0.001 to 2 were tried for x_1; for each x_1, values from 0.1 to 2000 were tried for A_1. The values of x_1 and A_1 that produced the smallest value of the weighted root-mean-square deviation (WRMSD) for the 415 solvation energies in the multiple linear regression fitting process were selected for the next step. In computing the WRMSD in this step and all

Table 1. SMVLE Surface Tension Coefficients

k	σ (cal/Å²)	k, k'	σ (cal/Å²)
H	57.88	H, C	−75.22
C	114.49	C, C	−70.59
N	−30.82	H, O	110.62
O	−84.28	O, C	187.69
F	46.48	O, O	98.59
Cl	14.69	C, N	30.94
Br	12.56	N, C	−52.83
P	−31.35	N, C (2)[a]	−261.62
S	−4.13	N, C (3)[a]	97.52
		O, N	256.52
		O, P	79.30

[a] Number in parentheses is m when m is not 1.

Table 2. Parameters for Local Electrostatic Terms

	$i = 1$	$i = 2$	$i = 3$
A_i	1984.0	1528.0	1488.0
x_i	0.07	0.037	0.057
B_i	−2.679	−23.413	453.544
B_{i+3}	1.454	−5.64	−139.35

subsequent steps, the neutral solutes and the ions had relative weights of W:1 where W is an integer parameter of the optimization scheme.

(2) I was increased by 1. The values of A_i and x_i already obtained were fixed and the values of A_I and x_I were optimized in the same way as in step 1.

(3) Step 2 was repeated until an arbitrary maximum value of I was reached. In this way, a first-round set of values of A_i and x_i ($i = 1, 2, ..., I$) was obtained.

(4) Now with I fixed at its maximum value, each A_i and x_i was reoptimized with the remaining parameters fixed. For example, A_1 and x_1 were reoptimized with other A_i and x_i ($i = 2, 3, ..., I$) obtained from previous steps fixed; then the reoptimized values of A_1 and x_1 along with the values of other A_i and x_i ($i = 3, ..., I$) were fixed in the reoptimization of A_2 and x_2; and so forth.

(5) Step 4 was repeated until the values of A_i and x_i ($i = 1, 2, ..., I$) did not change.

(6) Steps 1−5 were repeated for several maximal values of I. The final value of I was chosen to be 3 as discussed in the next section.

Results and Discussion

We tested the SMVLE method by examining the errors obtained with various maximal values of I in the range from 1 to 5. We found that the mean errors for ions decreased when I was increased from 1 to 3; however, when I was increased further, the mean errors for the anions improved but those for cations increased. Therefore, we set $I = 3$.

Because solvation free energies of ions are about an order of magnitude larger than those for neutrals, the predictions would have larger errors for ions even if the relative errors were similar. Furthermore, the experimental data for ions usually have larger absolute uncertainties. Another consideration is that our training set contains more neutral data than ionic data. Thus it is a matter of subjective judgment what value to choose for the parametrization weight W to
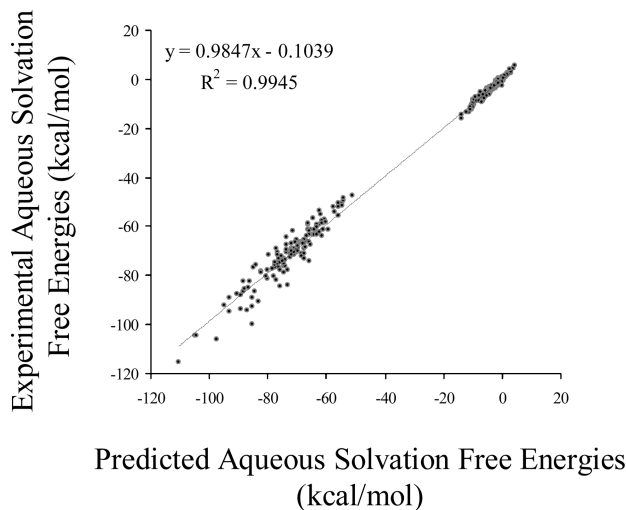
Figure 1. Correlation between the experimental and predicted aqueous solvation free energies.

balance the relative accuracies and desired accuracies for neutrals and for ions; we chose $W = 3$.

The calibrated parameters, namely, the surface tension parameters and the semiempirical G_L parameters, are given in Tables 1 and 2, respectively. The calculated absolute aqueous solvation free energies are obtained by the SMVLE method with these calibrated parameters. They are plotted in Figure 1 along with the experimental values. The predicted aqueous solvation free energies by the parametrized SMVLE are in excellent agreement with the experimental values. The square of the correlation coefficient between calculated and experimental aqueous solvation free energies is 0.9945. Moreover, the slope (0.9847) of the correlation equation is basically 1 and the corresponding constant (0.1039) is nearly 0, showing that the predicted solvation energies by the parametrized SMVLE method are in very good agreement with experimental values.

The mean unsigned errors for each solute class obtained from SMVLE were calculated to compare with those from SM6, SM8, SM8AD, and SMD. For the present study we recalculated the previously reported SM6 and SM8 errors using the reference solvation free energies for water-cluster data (data for ionic clusters and the water dimer) corrected by +2.38 kcal/mol to account for a recently discovered error due to a missing concentration correction term (see ref 46 for more detail) and to convert the SM6 error analysis to the scale based on the proton solvation energy of Tissandier et al.,[56] as discussed above. Note also that although SM8, SM8AD, and SMD were parametrized with 274 data for neutral solutes in water (the 272 data used here plus profenofos and tetramethylsilane), all mean errors given in the present article have been calculated for the 272 neutral data used here. The mean errors for the new SMVLE model and the previous SM6, SM8, SM8AD, and SMD models are shown in Table 3.

The error of the SMVLE model for neutrals is 0.55 kcal/mol and it is close to or better than that of the SMx methods, for which mean unsigned errors range between 0.47 (SM6/mPW1PW/6-31G*) and 1.31 (SMD/HF/6-31+G*) kcal/mol (Table 3). The SMVLE method not only retains good

accuracy for neutrals but also significantly improves the accuracy for ions. The SMVLE mean unsigned errors for the set of all 143 ions (3.25 kcal/mol) and for the set of 112 UC ions (3.07 kcal/mol) are smaller than the corresponding errors obtained with any other method tested in the present work (and we have shown previously[3,44] that the methods tested here are better than other available methods). The error of the SMVLE model for the set of 112 SC ions is 2.92 kcal/mol, which is close to or better that of the SMx methods, for which the MUE ranges between 2.80 (SM8AD/M06-2X/6-31G*) and 4.53 (SMD/mPW1PW/6-31G*) kcal/mol.

The local solvent environment is critical for ions, and it is difficult to simulate with implicit solvent. It was noticed in a previous study[43] that the overall error for aqueous ions decreased when one explicitly bound solvent molecule was introduced. The data for the two subsets of 31 ions (IU and MH) listed in Table 3 show that, in contrast to previous models, explicitly including one solvent water molecule with the ion just slightly increases the predictive accuracy for ions within the SMVLE method. The SMVLE error over 31 tested ions decreases from 4.45 kcal/mol (31 IU ions) to 3.88 kcal/mol (31 MH ions). The difference of ∼0.57 kcal/mol in SMVLE is much less than the ∼4 kcal/mol difference found on average with the non-SMVLE models tested in the present study (Table 3). The almost identical accuracies obtained with or without the addition of one explicit solvent molecule suggests that SMVLE is capable of successfully modeling the strong local electrostatic interactions between ionic solutes and solvent, and the addition of an explicit solvent molecule is unnecessary for SMVLE.

In an attempt to assess the reason for the success of SMVLE in modeling unclustered ions, we removed the electric field semiempirical term, G_L, in eq 1 and we recalculated the errors. The MUE for neutrals changes only about 4%, whereas the MUE for ions increases from ∼5 to ∼12 kcal/mol. This implies that the improvement in the predictive accuracy for ions can be attributed to the newly introduced G_L term. The values of the outward-directed normal electric fields are acting as indicators of the local solvent environment such that solute−solvent interactions stronger than would be anticipated from the bulk dielectric constant are associated with large values of the outward-directed normal electric field. As a special case, a hydrogen bond between solute and solvent may often be located by the direction of strongest outward-directed normal electric field, and the strength of the hydrogen bond might be represented by the magnitude of the strongest outward-directed normal electric field.[21,22]

SMVLE does not involve the optimization of intrinsic Coulomb radii for each atomic number, as in previous SMx models, or for atoms with various bonding types, as in some more empirical models. It is especially encouraging that SMVLE yields good results for ions despite not requiring this. In addition, this is of practical importance because it means SMVLE should be easier to extend to additional atomic numbers, if desired.

Although SMVLE provides significant improvement over all previous SMx models for ions, the improvement over SMD is particularly large and especially important. The

Computing Absolute Aqueous Solvation Free Energies

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1115**

Table 3. Mean Unsigned Errors in Aqueous Solvation Free Energies Calculated by SMVLE and Older SM*x* Methods[a]

ESM	mean unsigned errors (kcal/mol)										
	272 neutrals	81 PU ions	31 IU ions	31 MH ions	143 all ions	415 all solutes	60 SC anions	52 SC cations	112 SC ions	112 UC ions	384 proper solutes[b]
SMVLE											
HF/6-31+G*	0.55	2.55	4.45	3.88	3.25	1.48	3.17	2.63	2.92	3.07	1.24
SM6											
mPW1PW/6-31G*	0.47	2.55	8.57	4.34	4.24	1.77	3.45	2.58	3.05	4.22	1.22
mPW1PW/6-31+G*	0.55	2.90	8.52	4.29	4.42	1.89	3.00	3.61	3.28	4.46	1.35
SM8											
mPW1PW/6-31G*	0.56	2.51	8.41	4.16	4.15	1.80	3.44	2.42	2.97	4.14	1.26
mPW1PW/6-31+G*	0.63	2.59	8.46	4.03	4.17	1.85	2.96	3.03	2.99	4.21	1.32
M05-2X/6-31G*	0.59	2.54	8.41	4.18	4.17	1.83	3.49	2.42	2.99	4.16	1.29
SM8AD											
mPW1PW/6-31G*	0.60	2.94	6.25	2.64	3.59	1.63	2.86	2.85	2.86	3.86	1.26
M05-2X/6-31G*	0.52	3.06	6.25	2.68	3.67	1.61	2.95	2.96	2.95	3.94	1.23
M06-2X/6-31G*	0.61	2.88	6.05	2.59	3.50	1.61	2.78	2.82	2.80	3.76	1.25
SMD											
HF/6-31G*	0.91	3.28	8.10	3.63	4.40	2.12	3.86	2.82	3.38	4.61	1.63
HF/6-31+G*	1.31	3.59	8.93	3.60	4.75	2.50	4.53	2.50	3.59	5.07	1.98
mPW1PW/6-31G*	0.62	4.50	9.64	4.63	5.64	2.35	5.61	3.29	4.53	5.92	1.76
M05-2X/6-31G*	0.59	4.08	9.11	4.24	5.21	2.18	5.01	3.09	4.12	5.47	1.62
M06-2X/6-31G*	0.62	4.39	9.35	4.44	5.48	2.30	5.45	3.19	4.40	5.76	1.73

[a] ESM = electronic structure method, PU = properly unclustered, IU = improperly unclustered, MH = monohydrated, SC = selectively clustered, UC = unclustered. See text for detailed description of the subsets (Computational Details). [b] All solutes except IU ions.

reason it is especially important is that SMVLE and SMD do not require accurate partial atomic charges, which can sometimes be difficult to obtain for arbitrary levels of theory, extended basis sets, and complex systems. Thus SMVLE and SMD are more generally applicable. We also note that the present test of SMVLE includes diffuse functions (denoted by the "+" in 6-31+G*), and Table 3 shows that in previous methods the use of diffuse functions often decreases accuracy, which we interpreted as due to less stable partial atomic charges in SM6 and SM8 and to outlying charge in SMD. The good performance of SMVLE with a basis set containing diffuse functions is therefore particularly encouraging.

Concerning the computational complexity of the SMVLE method, the ratio of computing time spent for each SCRF cycle of the SMVLE calculation to that for each SCF cycle of the corresponding gas-phase calculation ranges from 1.2 to 1.6 when the number of basis functions used for the solute is larger than ~200.

Concluding Remarks

We have developed a new method, called SMVLE, for predicting absolute aqueous free energies of solvation by combining (1) the SVPE method, (2) semiempirical atomic surface tensions as used in the SM6 model, and (3) a new functional form, developed in the present study, that explicitly accounts for the local electrostatic effect. The SVPE term accounts for bulk electrostatics; the atomic surface tensions account for solvent cavitation, changes in dispersion energy, and possible changes in local solvent structure; and the final contribution accounts explicitly for nonbulk electrostatics in terms of the local electric field at the solute−solvent boundary. The parameters for SMVLE have been calibrated against a broad range of solutes, including 272 neutrals and 143 ions. The predicted

aqueous solvation free energies by the parametrized SMVLE method correlate very well with experiment and have a value of the square of the correlation coefficient equal to 0.9945 and a slope of 0.9847. Comparisons with previous SM*x* solvation models show that the SMVLE model not only has comparable accuracy for neutrals but also impressively increases the predictive accuracy for ions. The semiempirical terms (G_L) derived from the electric field are found to be primarily responsible for the increase in predictive accuracy for ions. The outward-directed normal electric fields that make the most important contributions account for strong interactions between the ionic solute and the nearby solvent, which makes the addition of explicit water molecules unnecessary. These encouraging results demonstrate that the parametrized SMVLE is accurate and effective in predicting absolute solvation free energies not only for neutral molecules but also for ions exhibiting strong solute−solvent interactions.

1116 *J. Chem. Theory Comput., Vol. 6, No. 4, 2010*

Liu et al.

Supporting Information Available: Three tables for all molecules involved in this study, including 272 neutral solutes, 112 unclustered ions, and 31 clustered monohydrated ions, and their experimental aqueous solvation free energies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

(2) (a) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094. (b) Rivail, J.-L.; Rinaldi, D. In *Computational Chemistry: Reviews of Current Trends*, Vol. 1; Leszczynski, J., Ed.; World Scientific: Singapore, 1996; pp 139−174. (c) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187–4225. (d) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093. (e) *Continuum Solvation Models in Chemical Physics*, Mennucci, B., Cammi, R., Eds.; Wiley: Chichester, U.K., 2007.

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

(4) (a) Chipman, D. M. *J. Chem. Phys.* **1997**, *106*, 10194–10206. (b) Chipman, D. M. *J. Chem. Phys.* **1996**, *104*, 3276–3289.

(5) Zhan, C.-G.; Bentley, J.; Chipman, D. M. *J. Chem. Phys.* **1998**, *108*, 177–192.

(6) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1999**, *110*, 1611–1622.

(7) Zhan, C.-G.; Landry, D. W.; Ornstein, R. L. *J. Phys. Chem. A* **2000**, *104*, 7672–7678.

(8) Chen, X.; Zhan, C.-G. *J. Phys. Chem. A* **2004**, *108*, 6407–6413.

(9) Zhan, C.-G.; Chipman, D. M. *J. Chem. Phys.* **1998**, *109*, 10543–10558.

(10) Vilkas, M. J.; Zhan, C.-G. *J. Chem. Phys.* **2008**, *129*, 194109.

(11) Zhan, C.-G.; Niu, S. Q.; Ornstein, R. L. *J. Chem. Soc., Perkin Trans. 2* **2001**, 23–29.

(12) Dixon, D. A.; Feller, D.; Zhan, C.-G.; Francisco, J. S. *J. Phys. Chem. A* **2002**, *106*, 3191–3196.

(13) Zhan, C.-G.; Dixon, D. A.; Sabri, M. I.; Kim, M. S.; Spencer, P. S. *J. Am. Chem. Soc.* **2002**, *124*, 2744–2752.

(14) Zheng, F.; Zhan, C.-G.; Ornstein, R. L. *J. Phys. Chem. B* **2002**, *106*, 717–722.

(15) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. B* **2003**, *107*, 4403–4417.

(16) Zhan, C.-G.; Spencer, P.; Dixon, D. A. *J. Phys. Chem. B* **2003**, *107*, 2853–2861.

(17) Dixon, D. A.; Feller, D.; Zhan, C.-G.; Francisco, J. S. *Int. J. Mass Spectrom.* **2003**, *227*, 421–438.

(18) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 11534–11540.

(19) Xiong, Y.; Zhan, C.-G. *J. Org. Chem.* **2004**, *69*, 8451–8458.

(20) (a) Chen, X.; Zhan, C.-G. *J. Phys. Chem. A* **2004**, *108*, 3789–3797. (b) Lu, H.; Chen, X.; Zhan, C.-G. *J. Phys. Chem. B* **2007**, *111*, 10599–10605.

(21) Chipman, D. M. *J. Chem. Phys.* **2003**, *118*, 9937–9942.

(22) Chipman, D. M.; Chen, F. W. *J. Chem. Phys.* **2006**, *124*, 144507.

(23) (a) Cramer, C. J.; Hawkins, G. D.; Truhlar, D. G. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1802–1804. (b) Storer, J. W.; Giesen, D. J.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G.; Liotard, D. A. *ACS Symp. Ser.* **1994**, *568*, 24–49. (c) Tuñón, I.; Ruiz-López, M. F.; Rinaldi, D.; Bertrán, J. *J. Comput. Chem.* **1996**, *17*, 148–155. (d) Chuang, Y.-Y.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quantum Chem.* **1998**, *70*, 887–896. (e) Cramer, C. J.; Truhlar, D. G. *Faraday Discuss. Chem. Soc.* **1998**, *110*, 477–479. (f) Chuang, Y.-Y.; Radhakrishnan, M. L.; Fast, P. L.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 4893–4909. (g) Chuang, Y.-Y.; Truhlar, D. G. *J. Am. Chem. Soc.* **1999**, *121*, 10157–10167. (h) Jaque, P.; Marenich, A.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. C* **2007**, *111*, 5783–5799. (i) Kim, Y.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2009**, *113*, 9109–9114. (j) Kim, Y.; Marenich, A. V.; Zheng, J.; Kim, K. H.; Kołodziejska-Huben, M.; Rostkowski, M.; Paneth, P.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 59–67.

(24) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2004**, *108*, 2020–2029.

(25) Xiong, Y.; Zhan, C.-G. *J. Phys. Chem. A* **2006**, *110*, 12644–12652.

(26) Zhan, C.-G.; Landry, D. W.; Ornstein, R. L. *J. Am. Chem. Soc.* **2000**, *122*, 2621–2627.

(27) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 9737–9744.

(28) (a) Rivail, J. L.; Terryn, B.; Rinaldi, D.; Ruiz-Lopez, M. F. *J. Mol. Struct. THEOCHEM* **1985**, *120*, 387–400. (b) Rinaldi, D.; Costa Cabral, B. J.; Rivail, J.-L. *Chem. Phys. Lett.* **1986**, *125*, 495–499.

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

(30) Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089–1097.

(31) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213–217.

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

(33) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(34) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257–3271.

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

(36) Zhu, T. H.; Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117–9133.

(37) Zhu, T. H.; Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111*, 5624–5624.

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

(39) Li, J. B.; Zhu, T. H.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178–2182.

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

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

(42) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2005**, *113*, 107–131 .

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

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

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

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

(47) (a) Hoijtink, G. J.; Deboer, E.; Vandermeij, P. H.; Weijland, W. P. *Recl. Trav. Chim. Pays-Bas Belg.* **1956**, *75*, 487–503. (b) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 393. (c) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1–11. (d) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164–170. (e) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129. (f) Cramer, C. J.; Truhlar, F. G. *Rev. Comput. Chem.* **1995**, *6*, 1–72. (g) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–153.

(48) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2009**, *42*, 493.

(49) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792–803.

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

(51) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertran, J., Eds.; Kluwer: Dordrecht, The Netherlands, 1996; pp 1−81.

(52) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754–2759.

(53) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.

(54) Nemethy, G.; Scheraga, H. A. *J. Chem. Phys.* **1962**, *36*, 3401–3417.

(55) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066–81.

(56) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.

(57) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

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

(59) Bondi, A. *J. Phys. Chem.* **1954**, *58*, 929–939.