Universal Quantum Mechanical Model for Solvation Free Energies Based on Gas-Phase Geometries

Gregory D. Hawkins, Christopher J. Cramer,* and Donald G. Truhlar*

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

Received: October 10, 1997; In Final Form: January 5, 1998

We present a new solvation model for predicting free energies of transfer of organic solutes from the gas phase to aqueous and organic solvents. The model is based on class II charges, gas-phase geometries, a generalized Born approximation to the polarization free energy, and SM5-type atomic surface tensions. The initial parametrization of the new model was developed to utilize the MNDO/d Hamiltonian, and we also present parameters for the MNDO, AM1, and PM3 Hamiltonians. These parametrizations are based on reasonably accurate gas-phase geometries for 43 ions and 260 neutral solute molecules composed of H, C, N, O, F, S, Cl, Br, and I and containing a wide variety of functional groups. For aqueous solutions, the parametrization is based on data for 248 of the neutrals and all of the ions. For organic solvents, it is based on 1836 experimental data points for 227 of the neutral solutes in 90 organic solvents. The parametrization based on the MNDO/d Hamiltonian is called SM5.2R/MNDO/d, and it yields a mean unsigned error of 3.8 kcal/mol for the free energy of hydration of ions and a mean unsigned error of 0.38 kcal/mol for the free energy of solvation of neutral solutes. Gas-phase geometries for all solute molecules were calculated at the Hartree−Fock level with a heteroatom-polarized valence-double-ζ basis set (HF/MIDI!), and we confirmed that the average errors increase only about 0.1 kcal/mol if we use the MNDO/d geometries.

1. Introduction

We have recently developed a universal solvation model called SM5.4, and we parametrized it for free energies of solvation of organic solutes in water and essentially any organic solvent.[1−5] This model builds on, and improves upon, our earlier solvation models SM1−SM3 for aqueous solutions[6−9] and SM4 for alkane solvents.[10,11] As in all previous SM*x* models ($x = 1−4$), the solute is described by molecular orbital theory with orbitals optimized self-consistently in the presence of a solvent electrostatic reaction field, geometries optimized in solution, and first-solvation-shell effects included by empirical atomic surface tensions. We developed a new strategy[1,3] in which atomic surface tensions depend on interatomic distances of the solute and four macroscopic characteristics (optical refractivity, acidity, basicity, and macroscopic surface tensile strength) of the solvent; this strategy is designed for use in a suite of models called the SM5 suite. The solute charge distribution and the reaction field[12] are calculated (without truncating a multipole expansion) from class IV charges,[13] and such charges are denoted by the 4 in SM5.4. We presented two parametrizations, denoted SM5.4/AM1 and SM5.4/PM3, which are for use with the AM1[14−16] and PM3[17] Hamiltonians, respectively. (Note that in earlier work, /AM1 and /PM3 were sometimes abbreviated as /A and /P, respectively.)

Although we expect the SM5.4/AM1 and SM5.4/PM3 parametrizations of the SM5 model to be very useful, there are several reasons why additional parametrizations can be valuable (just as more than one standard basis set and more than one way to include electron correlation find everyday use in gas-phase ab initio molecular orbital theory). In particular, a parametrization scheme based on class II charges (charge classes are explained in a previous paper[13]) and gas-phase geometries will be especially useful for two reasons.

(i) Class IV charges are so far available only for H, C, N, O, F, S, Cl, Br, and I, and they are available only for certain kinds of electronic structure calculations; e.g., they are available for AM1 and PM3 but not for MNDO and MNDO/d. In contrast, class II charges are calculable for any element from any kind of wave function. Furthermore, methods based on class II charges are less expensive and easier to incorporate in electronic structure packages.

(ii) Sometimes, especially for transition states, solutes with low-barrier torsions, multiple low-energy conformations, or five-membered rings, weakly bound complexes, and cases where one or more solvent molecules are treated explicitly, the semiempirical Hamiltonians such as AM1 and PM3 do not yield accurate geometries. In such cases one would prefer to estimate the solvation energy based on independently obtained gas-phase geometries rather than to reoptimize the geometry of the molecule with the Hamiltonian for which solvation parameters are available. Note that reparametrizing the present solvation model with explicit geometry optimization would not get around this difficulty. But combining a solvation calculation based on semiempirical Hamiltonians with a higher-level gas-phase geometry optimization, while not as satisfactory as simply calculating everything at a high level, will be an attractive option for many practical applications to complex systems.

Furthermore, there is a third reason for the present effort, namely, that AM1 and PM3 are not appropriate for all types of systems, and it will be useful to have solvation models based on other semiempirical Hamiltonians, on ab initio Hartree−Fock theory, and on density functional theory. The present paper, for example, will include parameters for use with MNDO/d[18−20] and MNDO[21] as well as AM1 and PM3. The inclusion

of MNDO/d will be especially significant because it includes d orbitals for P, S, Cl, etc., and d orbitals are required for a proper treatment of hypervalent systems and any credible treatment of transition metals. Although we do not consider transition metals in the present paper, we do obtain MNDO/d parameters for the atoms that occur in common ligands of organometallic complexes, and this is a required first step for eventually having a solvation model applicable to organometallics.

The new model presented here, which fills special needs i and ii, is called SM5.2R, and the parametrization for MNDO/d, which is especially important for the third reason mentioned above, is denoted SM5.2R/MNDO/d, where the 2 denotes class II charges and R denotes the use of rigid gas-phase geometries. We also present similar parametrizations, called SM5.2R/MNDO, SM5.2R/AM1, and SM5.2R/PM3 based on MNDO, AM1, and PM3, respectively. These also fill special needs i and ii, but without d orbitals, and they should be useful to the large community with experience or programs based on these widely used Hamiltonians, especially because of special need ii.

Since the SM5.2R model is designed to be used with any accurate gas-phase geometry, our nomenclature may be extended to include reference to the method used to obtain the gas-phase geometry upon which the SM5.2R calculation will be based. We adopt the standard "//" notation. For example, SM5.2R/MNDO/d//MNDO/d indicates that an SM5.2R/MNDO/d calculation is performed on a gas-phase geometry obtained using the MNDO/d Hamiltonian. The // may be omitted when the gas-phase geometry is clear from the context or specified elsewhere.

In the process of developing the SM5.2R/MNDO/d parameters, we made several improvements and simplifications in the SM5 parametrization process, and we anticipate that these will be useful for parametrizing atomic surface tensions for ab initio Hamiltonians,[22] for density functional theory,[23-25] for INDO/S,[26] and for developing an SM5.0R[27] organic model. Therefore, we will describe the new parametrization method carefully and completely so that in later work we can refer to this paper for such details and not repeat them.

Section 2 presents the theory of the SM5.2R model, and section 3 contains details about the parametrization. Results and discussion are located in section 4, while concluding remarks can be found in section 5.

2. Theory

All SMx models[1-11,28,29] partition the standard-state free energy of solvation, ΔG_S°, into two terms:

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

Equation 1 assumes the same concentration of solute molecules in the standard state in solution as in the standard state in the gas phase; if one uses a different standard state in the gas phase, one must add another term to account for this. The standard state used for the SMx models, assuming an ideal gas and ideal solution, is 1 mol/L in both the gas phase and solution. The first term in eq 1, ΔG_{ENP}, includes the changes in the electronic internal energy caused by inserting a solute into the solvent as well as the free energy change due to electric polarization of the solvent. SMx models that optimize the geometry of the solute self-consistently in the reaction field created by the solvent refer to this first term as ΔG_{ENP} because it includes the change in the nuclear internal energy caused by inserting a solute into the solvent in addition to the change in

the electronic internal energy and the free energy change due to electric polarization of the solvent. The present model assumes the changes in the nuclear internal energy upon solvation (denoted N in ΔG_{ENP}) will be small, and it neglects these changes and drops the N from the notation. The second term, G_{CDS}, contains the solvation free energy effects (such as cavitation, dispersion, and solvent-structure rearrangement) that arise within the first solvation shell and are not already included in ΔG_{EP}. The CDS term also accounts for deviations of the true electrostatics from the estimate included in ΔG_{EP}. (The true value may differ from the estimate for several reasons in addition to the neglected changes in the nuclear internal energy, including imperfect atomic charges, breakdown of the assumption of nuclear-centered atomic charges and discontinuous dielectric response at the solute/solvent boundary, uncertain atomic radii, deviation of the first-solvation-shell behavior from bulk dielectric behavior, and inexact calculation of the electrostatic free energy from the solute charge distribution.) In addition, the G_{CDS} term can make up for any systematic deficiencies in ΔG_{EP} that can be modeled in terms of the exposed surface area of the solute.

The ΔG_{EP} term can further be separated into an electronic solute term and a polarization term:

$$\Delta G_{EP} = \Delta E_E + G_P \qquad (2)$$

where ΔE_E represents the change in the electronic energy of the solute when inserted from the gas phase into solution and G_P is the free energy of electric polarization.

By use of Still's formulation[30] of the generalized Born approximation,[12,31-35] the polarization free energy can be written as

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k q_{k'}\left(R_{kk'}^2 + \alpha_k\alpha_{k'}\exp\left[-\frac{R_{kk'}^2}{d_{kk'}\alpha_k\alpha_{k'}}\right]\right)^{-1/2} \qquad (3)$$

where ϵ is the bulk dielectric constant of the solvent, q_k is the point-charge on atom k, $R_{kk'}$ is the interatomic distance between atoms k and k', and α_k is an effective radius that is dependent on a set of parametrized coulomb radii, ρ_k. The $d_{kk'}$ parameters are constants.

G_{CDS} is modeled as

$$G_{CDS} = \left(\sum_k \sigma_k^A A_k(\{r_k; r_S^{CD}\})\right) + \sigma^{CS}\sum_k A_k(\{r_k; r_S^{CS}\}) \qquad (4)$$

where σ_k^A is a partial atomic surface tension for atom k, and $A_k(\{r_k; r_S^{CX}\})$, with $X = D$ or S, is the solvent-accessible surface area of atom k that is dependent upon the set of parametrized CDS solute radii, $\{r_k\}$, and on one or another effective solvent radii, r_S^{CD} and r_S^{CS}. The first term in eq 4 models cavity-dispersion (CD) interactions, and the second term models cavity-structure (CS) effects. (Note that CD and CS effects were not treated separately for water in SM5.4/AM1 and SM5.4/PM3, and we will continue this approach by subsuming σ^{CS} into the σ_k^A terms for aqueous solvent in the present work. This is a special case of eq 4 in which $r_S^{CD} = r_S^{CS}$.)

In most previous SMx models,[1-11,29] the effective solvent radii used in eq 4 were nonzero,[36-38] but in recent work[27] we have found that an effective solvent radius of zero better correlates with the available experimental data. Using a zero effective solvent radius converts the solvent-accessible surface area to the van der Waals surface area; both of these areas may be considered special cases of the "exposed" area. We adopt

the zero-solvent-radius choice for the present work. Similarly, the development of previous models for organic solvents[3–5,10,11] utilized a second effective solvent radius for the second term of eq 4. We found empirically that using a second radius does not significantly improve the present model and thus have set both r_S^{CD} and r_S^{CS} to zero.

In SM5 models, the geometry dependence of the atomic surface tensions is based on a switching function called a COT[1] (cutoff tanh), which can be written as

$$T(R_{kk'}|\bar{R}, \Delta R) =$$
$$\begin{cases} \exp\left[-\left(\dfrac{\Delta R}{\Delta R - R_{kk'} + \bar{R}}\right)\right] & R_{kk'} \leq \bar{R} + \Delta R \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $R_{kk'}$ is the distance between two atoms k and k', and $\bar{R}$ is the midpoint of the switch, while $2\Delta R$ is the range over which the function switches. This function has the property that it vanishes identically for all $R_{kk'}$ greater than $\bar{R} + \Delta R$, but it has an infinite number of continuous derivatives for all $R_{kk'}$.

Restricting our attention at first to the solvent water and excluding solutes containing phosphorus, the atomic surface tensions for the SM5.2R models are written as

$$\sigma_k^A|_{k=H} = \tilde{\sigma}_H + \sum_{k'=C,O,N,S} \{T(R_{Hk'}|\bar{R}_{Hk'}, W)[\tilde{\sigma}_{Hk'} +$$
$$\tilde{\sigma}_{HO}^{(2)} \sum_{\substack{k''=H \\ k''\neq k \\ k'=O}} T(R_{k'k''}|\bar{R}_{k'k''}, W) + \tilde{\sigma}_{HN}^{(2)} \sum_{\substack{k''=N \\ k'=N \\ k'\neq k''}} T(R_{k'k''}|\bar{R}_{k'k''}, W)]\} \quad (6)$$

$$\sigma_k^A|_{k=C} = \tilde{\sigma}_C + \tilde{\sigma}_{CC} \sum_{\substack{k'=C \\ k'\neq k}} T(R_{kk'}|\bar{R}_{CC}, W) +$$
$$\tilde{\sigma}_{CC}^{(2)} \sum_{\substack{k'=C \\ k'\neq k}} T(R_{kk'}|\bar{R}_{CC}^{(2)}, W_{CC}) + \tilde{\sigma}_{CN}[\sum_{k'=N} T(R_{kk'}|\bar{R}_{CN}, W)]^2 \quad (7)$$

$$\sigma_k^A|_{k=O} = \tilde{\sigma}_O + \tilde{\sigma}_{OC} \sum_{k'=C} T(R_{kk'}|\bar{R}_{OC}^{(2)}, W_{OC}) +$$
$$\tilde{\sigma}_{OO} \sum_{\substack{k'=O \\ k'\neq k}} T(-T(R_{kk'}|\bar{R}_{OO}, W)|R_{TT}, W_{TT}) +$$
$$\tilde{\sigma}_{ON} \sum_{k'=N} T(R_{kk'}|\bar{R}_{ON}, W) \quad (8)$$

$$\sigma_k^A|_{k=N} =$$
$$\tilde{\sigma}_N + \tilde{\sigma}_{NC}\{\sum_{k'=C} T(R_{kk'}|\bar{R}_{CN}, W)[\sum_{\substack{k''\neq k \\ k''\neq k'}} T(R_{k'k''}|\bar{R}_{Ck''}, W)]^2\}^{1.3} +$$
$$\tilde{\sigma}_{NC}^{(2)} \sum_{k'=C} [T(R_{kk'}|\bar{R}_{CN}, W) \sum_{k''=O} T(R_{k'k''}|\bar{R}_{Ck''}, W)] +$$
$$\tilde{\sigma}_{NC}^{(3)} \sum_{k'=C} T(R_{kk'}|\bar{R}_{NC}^{(2)}, W_{NC}) \quad (9)$$

$$\sigma_k^A|_{k=F} = \tilde{\sigma}_F \quad (10)$$

$$\sigma_k^A|_{k=S} = \tilde{\sigma}_S + \tilde{\sigma}_{SS} \sum_{\substack{k'=S \\ k'\neq k}} T(R_{kk'}|\bar{R}_{SS}, W) \quad (11)$$

$$\sigma_k^A|_{k=Cl} = \tilde{\sigma}_{Cl} \quad (12)$$

$$\sigma_k^A|_{k=Br} = \tilde{\sigma}_{Br} \quad (13)$$

and

$$\sigma_k^A|_{k=I} = \tilde{\sigma}_I \quad (14)$$

where σ_k^A is the total CD surface tension for atom k, $\tilde{\sigma}$ is a surface tension coefficient, and the combination of COTs multiplied by a given surface tension coefficient is the surface tension functional associated with that coefficient.

With the exception of the surface tension functionals associated with the surface tension coefficients $\tilde{\sigma}_{CN}$, $\tilde{\sigma}_{NC}^{(2)}$, $\tilde{\sigma}_{NC}^{(3)}$, $\tilde{\sigma}_{HO}^{(2)}$, and $\tilde{\sigma}_{HN}^{(2)}$, all the surface tension functionals in the SM5.2R model were first developed as part of the SM5.4 aqueous solvation model.[1] The surface tension functionals associated with the surface tension coefficients $\tilde{\sigma}_{CN}$ and $\tilde{\sigma}_{NC}^{(2)}$ were developed in conjunction with the SM5.0R model.[27] The surface tension functionals associated with $\tilde{\sigma}_{NC}^{(3)}$, $\tilde{\sigma}_{HO}^{(2)}$, and $\tilde{\sigma}_{HN}^{(2)}$ were developed as part of the current work.

In preliminary studies for the SM5.2R model, we attempted to identify potential systematic errors that could possibly be addressed by an exposed-surface-area-dependent term within the nonelectrostatic portion of our model. The $\tilde{\sigma}_{NC}^{(3)}$ surface tension coefficient and related functional were developed to identify the nitrile functionality within solute compounds. Although we ultimately set this surface tension coefficient to zero in the SM5.2R models, we include the related functional because we believe it may be useful in future implementations of the SM5 models. We also identified two solute compounds from our training set, water and hydrazine, which would have relatively large errors in our final model if we used only the surface tension functionals that were developed for previous[1,27] SM5 models. Our predictions for the aqueous solvation free energy of water would be about 2 kcal oversolvated, and hydrazine would be about 2 kcal undersolvated. In addition to the aqueous data points, the training set contains free energy data for water in 18 organic solvents and for hydrazine in 4 organic solvents. In preliminary models, the errors for water and hydrazine in organic solvents typically mirrored the free energy of solvation errors in aqueous solution. To eliminate these systematic errors, we developed the surface tension functionals associated with the $\sigma_{HO}^{(2)}$ and $\sigma_{HN}^{(2)}$ surface tension coefficients in eq 6 that distinguish these two solutes from other types of solutes. The $\sigma_{HO}^{(2)}$ surface tension coefficient and related functional affect only three solutes in our training set, namely, water, H_3O^+, and $CH_3OH_2^+$ (21 data points, including organic solvents), and they successfully remove the systematic oversolvation for water in the SM5.2R model. The $\sigma_{HN}^{(2)}$ surface tension coefficient and related functional affect only hydrazine in our training set for a total of five data points, and they remove the systematic undersolvation for hydrazine in the SM5.2R model. (For a discussion of how the $\sigma_{HN}^{(2)}$ surface tension coefficient and related functional affect organic hydrazones, see the Results and Discussion section.)

For water, the surface tension coefficients in eqs 6–14 are constants, but for organic solvents, the surface tension coefficients depend on three properties of the solvents, namely, the index of refraction[39] and Abraham's $\sum \alpha_2^H$ and $\sum \beta_2^H$ descriptors.[40–42] In particular,

$$\tilde{\sigma}_i = \hat{\sigma}_i^{(n)} n + \hat{\sigma}_i^{(\alpha)} \alpha + \hat{\sigma}_i^{(\beta)} \beta \quad (15)$$

where i is any of the subscripts on the surface tension coefficients in eqs 6−14, n is the index of refraction, and α and β are Abraham's $\sum\alpha_2^H$ and $\sum\beta_2^H$ descriptors, respectively.

In the original SM5 models, σ^{CS} is nonzero for organic solvents[3−5] and it depends on the index of refraction and the macroscopic surface tension[39] of the solvent. Since the present models use the same zero solvent radius for both terms of eq 4, the σ^{CS} dependence on the index of refraction correlates strongly with the index of refraction dependence of the atomic surface tensions, σ_k, and this dependence was dropped out of σ^{CS}.

Preliminary work for the SM5.2R model indicated that, as in earlier models,[3−5] the general treatment for the organic solvents would have small systematic errors in treating aromatic solvents as well as solvents that contain one or more electronegative halogen atoms (F, Cl, and Br). Initial testing suggested that our general treatment for the organic solvents could be significantly improved by incorporating additional solvent descriptor dependencies into σ^{CS}. After considerable experimentation, we chose three new solvent descriptors for σ^{CS} in addition to the macroscopic surface tension (denoted γ) used in the original SM5.4 model.[4,5] As a consequence, for the SM5.2R model, σ^{CS} can be written as

$$\sigma^{CS} = \hat{\sigma}_{CS}^{(\gamma)}\gamma + \hat{\sigma}_{CS}^{(\beta^2)}\beta^2 + \hat{\sigma}_{CS}^{(\phi^2)}\phi^2 + \hat{\sigma}_{CS}^{(\psi^2)}\psi^2 \qquad (16)$$

where β^2 denotes the square of the solvent's $\sum\beta_2^H$ descriptor, ϕ^2 is the square of the fraction of the nonhydrogenic solvent atoms that are aromatic carbon, and ψ^2 is the square of the fraction of the nonhydrogenic solvent atoms that are electronegative halogen atoms. In other words, ϕ is defined as the number of aromatic carbon atoms in the solvent divided by the total number of non-hydrogenic atoms in the solvent, and ψ is defined as the number of F, Cl, and Br atoms in the solvent divided by the total number of non-hydrogenic atoms in the solvent.

3. Parametrization

3.1. Choice of Training Set. To avoid using ambiguous data for parametrization, we required that any solute included in our training set has a single dominant configuration in the gas phase. As a general rule, we considered a solute to have a single dominant gas-phase configuration if the calculated energy for a given configuration is at least 2 kcal/mol lower in energy than any alternative configuration. In implementing this rule, we ignore conformational minima that for all practical purposes need not be distinguished from the lowest-energy minimum. For example, torsional minima of methyl groups on benzene rings are usually not significant.

Initially, we started with the SM5.4 aqueous model neutral training set.[1] The experimental free energies of aqueous solvation for these molecules were taken from the references that follow and were converted to a consistent standard state of 1 mole of solute per liter of ideal gas for the gas phase and 1 mole of solute per liter of ideal aqueous solution in water. The sources for the aqueous experimental data are, in order of preference, Suleiman and Eckert,[43] Cabani et al.,[44] Abraham et al.,[45] Hine and Mookerjee,[46] Wagman,[47] Wolfenden,[48] Wauchope and Haque,[49] and Han and Bartles.[50]

To these data, we added the additional nitrogen and halogen-containing solutes identified in the SM5.0R aqueous model neutral training set,[27] for which the experimental aqueous free energy of solvation values were taken from the MedChem database,[51] and from Ferguson et al.[52] To this set of solutes, we added any additional solutes that were included in the SM5.4

organic solvation model training set,[3,4] the SM5.4 benzene or toluene solvation training set,[4] or SM5.4 chloroform solvation training set.[5] For these solutes the aqueous free energy of solvation was taken from the MedChem database.[51] This produced a training set that contains 260 solutes and covers a wide and balanced variety of organic functional groups containing H, C, N, O, F, P, S, Cl, Br, and/or I atoms.

To parametrize the SM5.2R model for organic solvents, we considered the 90 organic solvents used when parametrizing the SM5.4 organic solvation models.[3−5] When available, partition coefficient data from the MedChem database[51] were usually used to determine experimental free energy of transfer data for a given solute from the aqueous phase to each of the 90 organic solvents, and these data were combined with the aqueous free energy of solvation data just described to obtain free energies of solvation in organic solvents. The only exceptions to the above protocol were for solutes in 1-octanol that were available in Dallas's thesis,[53] and for solutes in n-hexadecane where data were taken from selected sources[40,54,55] as described previously.[10]

When extracting data from the MedChem database, first we discarded any measurements that were made under conditions that do not correspond to conditions for which this model was developed. These exclusionary criteria are listed elsewhere.[4] If one or more solvent/air partition coefficients remain, they are used to determine the solvation free energy for a given solute. When there is more than one nonexcluded solvent/air partition coefficient, data are processed by a statistical procedure explained previously.[4] Altogether, there are 260 solute molecules and a total of 2084 solute/solvent data points for the 91 solvents (including water) in the SM5.2R neutral training set.

The SM5.2R ionic training set consists of the 34 ions included in the SM5.4 ionic training set.[1] Since the publication of the SM5.4 paper, a new paper by Florian and Warshel[56] has been published, which contains experimental values for the ions that differ from the values published earlier by Pearson.[57] For this work, the experimental ionic aqueous solvation free energies are taken from Pearson[57] only if they are not available in the more recent paper of Florian and Warshel.[56]

Note that the experimental value we use for the aqueous solvation free energy of anisole has changed from our previous[1−5,8,9,27,29] papers. The aqueous free energy of solvation predicted by Hine and Mookerjee,[46] −1.03 kcal/mol, was included in the Cabani et al.[44] compilation and thus by our priority scheme should be used for the aqueous solvation free energy for anisole, but we have chosen to use the value given by Abraham et al.,[45] −2.45 kcal/mol. The MedChem database contains a footnote for the Hine and Mookerjee[46] prediction that indicates that it is unreliable. Another reason for making this change is that using Abraham's number in conjunction with the partition coefficient measurement from the MedChem database[51] for $P_{1-octanol/water}$ yields a solvation free energy for anisole in 1-octanol of −5.31 kcal/mol. This value is much closer to the solvation free energy predicted with the direct $P_{1-octanol/water}$ data in Dallas's thesis,[53] −5.47 kcal/mol, than is the solvation free energy for anisole in 1-octanol arrived at using Hine and Mookerjee's[46] free energy of aqueous solvation with the $P_{1-octanol/water}$ measurement in the MedChem database,[51] −3.88 kcal/mol.

3.2. Gas-Phase Geometries. The SM5.2R model is specifically designed to be used with any accurate gas-phase geometry. The parameters of the model are defined to absorb any solute geometry relaxation that occurs when a particular solute is placed in a given solvent. The gas-phase geometries for the

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3261**

TABLE 1: Atomic Radii (Å) Used in Calculating the Electrostatic (ρ_k) and Nonelectrostatic (r_k) Contributions to the Free Energy of Solvation in the SM5.2R Model

k	$\rho_k{}^a$	$r_k{}^b$
H	0.91[c]	1.20
C	1.78	1.70
N	1.92	1.55
O	1.60	1.52
S	2.05[d]	1.80
F	1.50	1.47
Cl	2.13	1.75
Br	2.31	1.85
I	2.66	1.98

[a] Except where indicated, these values are from Chambers et al.[1] [b] From Bondi.[59] [c] From Hawkins et al.[2] [d] Optimized in this work.

molecules in our SM5.2R neutral and ionic training sets were optimized at the Hartree−Fock (HF) level using the MIDI![58] basis set. MIDI! is a heteroatom-polarized split valence basis,[58] whose polarization functions were selected to provide accurate and economical geometry optimizations.

To test the stability of the SM5.2R/MNDO/d and SM5.2R/ MNDO parametrizations to changes in the gas-phase geometry used in the calculation, results are tabulated for the SM5.2R/ MNDO/d model using MNDO/d gas-phase geometries and for the SM5.2R/MNDO model using MNDO gas-phase geometries, in addition to using the HF/MIDI! geometries for each type of calculation.

3.3. Parameters within the Solvation Models. The parameters used in the SM5.2R model can be divided into five different groups: (i) the electrostatic radius for each atomic number used in the calculation of the effective Born radii as part of the generalized Born approximation to the polarization effects within the solute/solvent system; (ii) the parameters used in the generalized Born formula, which are the $d_{kk'}$ parameters in eq 3; (iii) the van der Waals radius for each atomic number, used in calculating the exposed van der Waals surface area for each atom within a molecule for use in computing the nonelectrostatic contribution to the overall solvation free energy; (iv) the parameters in the surface tension functionals that specify the dependence of the atomic surface tensions upon the local molecular geometry within the calculation of the nonelectrostatic effects to the overall solvation free energy; (v) the surface tension coefficients.

The solvation energies are nonlinear functions of parameters of types i−iv, but linear functions of parameters of type v. With the exception of the electrostatic radii for sulfur (which will be discussed later) and hydrogen, parameters of types i and ii were taken from the original SM5.4 paper.[1] In the models presented in this paper, the electrostatic radius for H is always fixed to 0.91 Å, a value that was originally optimized for the SM5.4PD models.[2] For the van der Waals radii, the atomic radii suggested by Bondi[59] were utilized. Almost all of the type iv parameters ($R_{kk'}$ and W_k) were taken from previous work,[1,27] with the exception of $\bar{R}_{NN}$, which is utilized in the surface tension functional related to the $\tilde{\sigma}_{HN}^{(2)}$ surface tension coefficient that was developed in the current work. The surface tension coefficients were optimized to fit the difference between the experimental solvation free energies for the data points in our training set and the electrostatic component calculated with the SM5.2R model and a given Hamiltonian. Further details of this procedure are included in section 3.4 (see Tables 1−3).

Equations 4−16 allow for 104 possible surface tension coefficients in each parametrization of the SM5.2R model, 25

TABLE 2: $\bar{R}$ COT Parameters (Å) Used in SM5.2R Model

form[a]	k'	$\bar{R}$
$R_H k'$	C, O, N	1.55[b]
	S	2.14[c]
$R_C k'$	C, N	1.84[c]
	O, F	1.84[b]
	S	2.20[b]
	C1	2.10[b]
	I	2.60[d]
$R_O k'$	N	1.50[b]
	O	2.75[c]
$R_N k'$	N	1.85[d]
$R_S k'$	S	2.75[c]
$R_C^{(2)} k'$	C	1.27[c]
$R_O^{(2)} k'$	C	1.33[c]
$R_N^{(2)} k'$	C	1.225[d]
$W_{k'C}$	C, O	0.07[c,b]
	N	0.065[d]
W	0.3 Å[c]	
W_{TT}	0.4 Å[c]	
R_{TT}	−0.4 Å[c]	

[a] Parameters $R_{kk'}$ and $R_{k'}k$ have identical values. [b] From Hawkins et al.[27] [c] From Chambers et al.[1] [d] New in this work.

TABLE 3: Surface Tension Coefficients (cal mol⁻¹ Å⁻²) Optimized for SM5.2R/MNDO/d

k	$\hat{\sigma}_k^{(n)}$	$\hat{\sigma}_k^{(\alpha)}$	$\hat{\sigma}_k^{(\beta)}$	$\hat{\sigma}_k^{(water)}$
H	41.39			98.86
C	47.12	29.65	−23.44	66.41
N	−19.53	−165.99	36.92	−193.07
O	−66.62	−61.40		−237.31
F	12.29			55.09
S	−72.71	−49.76	41.04	−91.00
Cl	−31.73			−0.02
Br	−44.21			−19.31
I	−50.44			−30.18
H, C	−97.57			−127.61
H, N	−102.16		−158.03	−240.72
H, N (2)	−152.04			−134.82
H, O	−66.24	−330.88	−472.23	−412.19
H, O (2)	153.33			320.76
H, S	23.71			−9.13
C, C	−64.18			−63.08
C, C (2)	16.89			29.16
O, C	42.50		−32.88	158.83
O, O	14.83	146.52	8.66	152.86
C, N	−60.77	218.33		122.14
N, C	−13.01	−62.52	3.61	−71.45
N, C (2)				−383.81
N, C (3)				
O, N	102.50	86.06	81.05	382.00
S, S	−5.43			27.08

$\sigma_{CS}^{(\gamma)}$	$\sigma_{CS}^{(\beta2)}$	$\sigma_{CS}^{(\phi2)}$	$\sigma_{CS}^{(\psi2)}$
0.2424	10.99	−2.75	−10.07

for water as a solvent, and 79 for the organic solvents, but it is not necessary to use all potential surface tension coefficients when fitting the SM5.2R model for organic solvents. Initially, any potential surface tension coefficient that was found unnecessary in the development of the original SM5.4 model[4] was set to zero in the SM5.2R model. The remaining surface tension coefficient and functional combinations were evaluated on the basis of three general rules. (1) If a surface tension coefficient and related functional affect less than two different solutes, the coefficient is set to zero. (2) If using a surface tension coefficient and related functional do not improve the mean unsigned error for the affected solutes by at least 0.1 kcal/mol, the coefficient is set to zero. (3) Any surface tension coefficient that is not set to zero by either of these rules is retained.

Of the nonzero surface tension coefficients used in the SM5.4 organic model,[4] we set the $\hat{o}_{HS}^{(\beta)}$ and $\hat{o}_{OC}^{(\alpha)}$ to zero because of rule 1. The $\hat{o}_{SS}^{(\beta)}$ was set to zero to satisfy rule 2. The $\hat{o}_{HO}^{(\beta)}$, which was set to zero in the SM5.4 organic models, was used in the fitting because preliminary tests indicated that its inclusion is consistent with rules 1 and 2. As discussed in section 2, the surface tension functionals associated with the $\tilde{\sigma}_{CN}$, $\tilde{\sigma}_{NC}^{(2)}$, $\tilde{\sigma}_{NC}^{(3)}$, $\tilde{\sigma}_{HN}^{(2)}$, and $\tilde{\sigma}_{HO}^{(2)}$ surface tension coefficients have not been used in any previous SM*x* model for organic solvents. Thus, by eq 15 there are 15 potential new surface tension coefficients to use in the general organic parametrization of the SM5.2R model. As mentioned earlier, the $\tilde{\sigma}_{NC}^{(3)}$ surface tension coefficients were set to zero for aqueous and organic solvents owing to rule 2. The $\tilde{\sigma}_{NC}^{(2)}$ surface tension coefficients were set to zero for organic solvents to satisfy rule 1. Then, to satisfy rule 3, we added the index-of-refraction-dependent surface tension coefficient for the remaining three surface tension functionals. The α- and β-dependent surface tension coefficients for these three functionals were then tested. Only $\tilde{\sigma}_{NC}^{(\alpha)}$ satisfied both rules 1 and 2, and thus, it was added to the set of parameters to be optimized in the SM5.2R model. Altogether, there are 24 surface tension coefficients used in the water portion of the SM5.2R model and 45 surface tension coefficients optimized for the organic portion of the model.

Once the surface tension coefficients were selected, they were optimized to fit the difference between the experimental solvation free energies for the data points in our training set and the electrostatic component calculated with the SM5.2R model. Further details of this procedure are included in the following section.

3.4. Parametrization of the Model. *3.4.1. Determination of an Unfitness Function for a Given Parameter Set.* As discussed in section 3.3, the SM5.2R model contains both linear and nonlinear parameters. For a given set of nonlinear parameters, we can determine the best values for the linear surface tension coefficients, $\tilde{\sigma}_k$, by minimizing the sum of the squares of the error between the experimental aqueous solvation free energy and the predicted solvation free energy over a given subset of the training set solute molecules. Thus, to determine the best overall parameters for the SM5.2R model, we must first devise a method for evaluating the unfitness of a given set of nonlinear parameters.

The selection of an unfitness function for a set of nonlinear parameters requires considering the systematic differences between ionic and nonionic solutes. As we have mentioned in previous papers,[1,9] the ionic compounds in our training set have experimental free energies of aqueous solvation that are favorable, in the range $55-110$ kcal/mol, and generally have experimental uncertainties of about 5 kcal. The favorable interaction between the water solvent and the ionic solute can be attributed primarily to the electrostatic interactions between the two. The neutral solutes in our training set have experimental aqueous solvation free energies that range from -15 to $+5$ kcal and have uncertainties on the order of 0.2 kcal. In neutral compounds, the electrostatics are not necessarily the predominant source of solvation free energy in every case. (Keep in mind that there is no rigorous method for separating the electrostatic and nonelectrostatic components of the solvation free energy for any molecule.)

Since the results for neutral solutes are more sensitive to the fit of the surface tension coefficients (linear nonelectrostatic parameters) and the potential errors for the ionic solutes are much larger than for the neutral solutes, we decided not to consider the ionic solutes when determining the best values for

the surface tension coefficients for a given set of nonlinear parameters. However, to compare the relative success of one set of nonlinear parameters to another, we define an error function that does include both neutral and ionic solutes. The unfitness function, U, that we chose was first developed in earlier work:[2]

$$U = \frac{\displaystyle\sum_{n=1}^{N}|G_S^{\circ}(\exp_n) - G_S^{\circ}(\mathrm{calc}_n)| + \frac{1}{6}\left(\sum_{i=1}^{I}|G_S^{\circ}(\exp_i) - G_S^{\circ}(\mathrm{calc}_i)|\right)}{N + I} \tag{17}$$

where N is the total number of neutral molecules in a given subset, I is the total number of ionic compounds, $G_S^{\circ}(\exp)$ is the experimental standard-state free energy of solvation, and $G_S^{\circ}(\mathrm{calc})$ is the standard-state free energy of solvation calculated for a given set of nonlinear parameters and reoptimized linear parameters.

3.4.2. Water as a Solvent. Initially, the HF/MIDI! geometries were obtained for all 260 neutral solutes and 34 ions. Since the parametrization of the SM5.2R model for predicting free energies of solvation in water is independent of the parametrization for the SM5.2R model for predicting free energies of solvation in organic solvents, we first limited our attention to the compounds for which we had experimental aqueous solvation free energy data.

We began the parametrization for aqueous solution by considering only the 157 neutral solutes and 25 ionic compounds in our aqueous training set, which contain at most H, C, N, and/or O atoms. We selected a set of the nonlinear parameters as indicated in section 3.3, and then we fit the linear surface tension coefficients to minimize the sum of the squares of the error in the predicted free energies of solvation for the neutral compounds. At that point, we systematically perturbed the nonlinear electrostatic radii used for H, C, N, and O while refitting the linear parameters to the neutral solutes for each perturbation. The unfitness of each perturbed set was calculated using eq 17 for all 157 neutral compounds and 25 ionic solutes. None of the perturbed parameter sets resulted in unfitness values that were significantly better than the nonperturbed nonlinear parameter set, so we chose to use the electrostatic radii optimized in earlier work.[1,2]

Table 1 includes a list of the atomic radii used in calculating the electrostatic and nonelectrostatic contributions to the free energy of solvation for the SM5.2R model. The COT parameters used in the surface tension functionals are listed in Table 2.

The H, C, N, and O parameters were then frozen for the aqueous portion of the SM5.2R/MNDO/d model. Using these parameters, we considered the 91 neutral solute compounds and 9 ionic solutes that contain F, S, Cl, Br, or I. (These molecules may also contain H, C, N, and/or O.) The remaining surface tension coefficients were optimized for our initial choice of nonlinear parameters. We then perturbed the electrostatic radii used for F, S, Cl, Br, and I and refit the nonfrozen linear parameters after each perturbation. When perturbing the electrostatic radius for sulfur, we discovered that setting the sulfur radius larger than the 1.92 Å used previously[1] resulted in a significant improvement in the unfitness value (eq 18) calculated for the compounds containing sulfur. By the increase of the electrostatic radius for sulfur to 2.05 Å, the unfitness of the sulfur-containing subset of the training set improved by 0.2 kcal for the SM5.2R/MNDO/d parametrization. This aspect of

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3263**

TABLE 4: Performance of the SM5.2R/MNDO/d Model by Solvent Functional Group Class

	number of			SM5.2R/MNDO/d//HF/MIDI!		SM5.2R/MNDO/d//MNDO/d	
solvent class	solvents[a]	solute classes[b]	data[c]	signed[d]	unsigned[d]	signed[d]	unsigned[d]
aqueous	1	31	248	−0.05	0.47	−0.09	0.66
alkanes	11	30	475	−0.05	0.29	−0.10	0.34
cycloalkanes	2	24	106	0.00	0.36	−0.08	0.37
arenes	12	16	256	0.08	0.30	−0.07	0.33
aliphatic alcohols	12	31	299	−0.07	0.46	−0.22	0.59
aromatic alcohols	2	7	12	0.16	0.49	0.08	0.59
ketones	4	10	35	0.02	0.35	−0.27	0.64
esters	2	8	36	0.42	0.50	0.36	0.51
aliphatic ethers	4	19	99	0.02	0.46	−0.09	0.48
aromatic ethers	3	5	15	−0.40	0.43	−0.34	0.43
amines	2	6	12	−0.04	0.45	−0.14	0.58
pyridines	3	5	15	−0.09	0.30	−0.04	0.39
nitriles	2	5	10	−0.27	0.45	−0.24	0.48
nitro compounds	4	8	27	−0.02	0.55	−0.29	0.87
tertiary amides	2	5	10	0.13	0.23	0.18	0.38
haloaliphatics	12	27	269	−0.05	0.45	−0.21	0.59
haloaromatics	6	11	106	−0.06	0.27	−0.12	0.34
miscellaneous acidic solvents	3	5	15	0.16	0.38	0.21	0.45
miscellaneous nonacidic solvents	4	12	39	−0.13	0.42	−0.26	0.56
total	91	31	2084	−0.02	0.38	−0.12	0.48

[a] Number of solvents in this solvent class. [b] Number of solute classes for which data exist in this solvent class. [c] Total number of solute/solvent data involving this solvent class. [d] Mean errors over data in this solvent class (kcal/mol).

the models is further discussed in the Results and Discussion section. The electrostatic radius for sulfur of 2.05 Å was accepted for the SM5.2R model, while the rest of the electrostatic radii were left as developed in conjunction with earlier models. Once these decisions were finalized, the linear surface tension coefficients were optimized separately for each parametrization of the SM5.2R model (see Tables 1−3).

The parametrizations of the SM5.2R model for the MNDO, AM1, and PM3 Hamiltonians were completed using the final nonlinear parameters developed for the SM5.2R/MNDO/d parametrization and surface tension coefficients, which were optimized for each parametrization using the two-step method of finding the surface tension coefficients first for solutes containing H, C, N, and/or O, freezing those parameters, and then finding the surface tension coefficients for the solutes containing F, S, Cl, Br, and/or I. Finally, with all these parameters frozen, we considered solutes containing P.

3.4.3. Organic Solvents. The SM5.2R model uses the same nonlinear parameters for predicting free energies of solvation in organic solvents that were developed for treating aqueous systems (see Tables 1 and 2). To obtain the linear parameters (i.e., surface tension coefficients) for organic solvents for one of the Hamiltonians, e.g., the MNDO/d model, we first restricted our attention to the 1618 experimental solvation free energy data points for solutes containing at most H, C, N, and/or O atoms. This subset spans 164 solutes and all 90 organic solvents. We fit the 36 surface tension coefficients affecting these compounds to minimize the sum of the squares of the errors in the predicted free energies of solvation for all 1618 pieces of data. These surface tension coefficients were then frozen before progressing with the parametrization. In the next stage, the best surface tension coefficients for F, S, Cl, Br, or I solute atoms in organic solvents were determined, again by least squares from the 216 pieces of data in our database, which contains these solute atoms but do not contain P. This subset contains 96 solutes in 34 organic solvents. For this subset there were 9 surface tension coefficients. In the third stage, we parametrized phosphorus compounds. This subset of the training set for organic solvents involves 37 data for 13 solutes in 10 organic solvents and 3 surface tension coefficients.

In some cases a four-stage procedure was used instead of the above three-stage procedure. Stage 1 was accomplished using all training set data for compounds containing only H, C, and/or O and a preliminary N training set. Stages 2 and 3 proceeded as before. Then in stage 4 all surface tension coefficients involving N were given their final values by simultaneously refitting the training set data for all nitrogen compounds, including those containing both N and halogens, with all other parameters fixed. We found little difference between the two procedures.

3.4.4. Final Coefficients. The surface tension coefficients for MNDO/d are in Table 3, and those for MNDO, AM1, and PM3 are included in the Supporting Information.

3.4.5. Parametrization for P. The parametrization of SM5.2R for solutes containing P was less successful than for the other solutes, and all further discussion of solutes containing P is relegated to Supporting Information.

4. Results and Discussion

Overall, the SM5.2R/MNDO/d model using HF/MIDI! geometries achieves a mean unsigned error of 0.38 kcal for the 2084 data of free energy of solvation of neutral molecules. By use of optimized MNDO/d gas-phase geometries, the SM5.2R/MNDO/d parametrization achieves a mean unsigned error of 0.48 kcal for the 2084 data points in the neutral portion (excluding phosphorus) of the training set. Considering the 43 ionic solutes that do not contain phosphorus, that have available experimental aqueous solvation energies, and that were used in the parametrization, the SM5.2R/MNDO/d//HF/MIDI! model yields a mean unsigned error of 3.8 kcal/mol and a mean signed error of 0.4 kcal/mol.

Table 4 shows the performance of the MNDO/d parametrization of the SM5.2R models by solvent class, while Table 5 contains similar results for the MNDO, AM1 and PM3 parametrizations of the SM5.2R model. Among solvent classes in the SM5.2R/MNDO/d parametrization, the error is relatively constant with the largest unsigned error of 0.55 kcal coming from the 27 data points for solutes dissolved in nitro compounds while the smallest error of 0.23 kcal arises for solvation free energies in tertiary amide solvents. The largest signed error occurs in ester solvents, where the SM5.2R/MNDO/d param-

TABLE 5: Performance of the SM5.2R Models for the MNDO, AM1, and PM3 Hamiltonian Parametrizations by Solvent Functional Group Class

	SM5.2R/MNDO// HF/MIDI!		SM5.2R/AM1// HF/MIDI!		SM5.2R/PM3// HF/MIDI!		SM5.2R/MNDO// MNDO	
solvent class	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]
aqueous	−0.07	0.47	−0.03	0.44	−0.01	0.51	−0.11	0.65
alkanes	−0.05	0.29	−0.06	0.32	−0.05	0.30	−0.10	0.34
cycloalkanes	0.00	0.36	0.07	0.37	0.04	0.35	−0.07	0.38
arenes	0.08	0.30	0.16	0.34	0.13	0.33	−0.06	0.33
aliphatic alcohols	−0.09	0.47	−0.03	0.53	−0.04	0.49	−0.23	0.61
aromatic alcohol	0.16	0.49	0.16	0.52	0.16	0.51	0.09	0.60
ketones	0.02	0.35	−0.07	0.42	−0.02	0.40	−0.27	0.64
esters	0.42	0.50	0.50	0.59	0.46	0.56	0.36	0.52
aliphatic ethers	0.02	0.46	0.08	0.50	0.07	0.47	−0.09	0.49
aromatic ethers	−0.40	0.43	−0.43	0.49	−0.43	0.46	−0.34	0.43
amines	−0.04	0.45	0.10	0.43	0.03	0.40	−0.14	0.58
pyridines	−0.09	0.30	−0.13	0.43	−0.11	0.38	−0.04	0.39
nitriles	−0.27	0.45	−0.53	0.60	−0.40	0.56	−0.24	0.48
nitro compounds	−0.01	0.55	−0.23	0.47	−0.14	0.55	−0.28	0.87
tertiary amides	0.13	0.23	0.03	0.36	0.09	0.34	0.18	0.38
haloaliphatics	−0.05	0.45	−0.10	0.45	−0.07	0.45	−0.21	0.59
haloaromatics	−0.06	0.27	−0.22	0.34	−0.15	0.29	−0.12	0.34
miscellaneous acidic solvents	0.16	0.38	0.17	0.52	0.18	0.48	0.21	0.45
miscellaneous nonacidic solvents	−0.13	0.42	−0.21	0.46	−0.18	0.44	−0.26	0.56
total	−0.03	0.39	−0.02	0.41	−0.01	0.40	−0.12	0.48

[a] Mean errors over data in this solvent class (kcal/mol).

TABLE 6: Performance of the SM5.2R/MNDO/d Model by Solute Functional Group Class

	number of			SM5.2R/MNDO/d// HF/MIDI!		SM5.2R/MNDO/d// MNDO/d	
solute class	solutes[a]	solvent classes[b]	data[c]	signed[d]	unsigned[d]	signed[d]	unsigned[d]
unbranched alkanes	9	19	84	0.05	0.38	0.13	0.40
branched alkanes	5	3	12	0.07	0.45	0.19	0.42
cycloalkanes	5	6	18	−0.14	0.34	−0.06	0.33
alkenes	9	4	27	−0.05	0.21	−0.04	0.21
alkynes	5	3	14	0.04	0.13	0.09	0.16
arenes	9	19	134	−0.27	0.39	−0.26	0.39
alcohols	17	19	385	−0.01	0.33	0.08	0.33
ethers	12	19	93	0.09	0.39	0.35	0.48
aldehydes	7	8	38	−0.27	0.46	−0.52	0.61
ketones	12	18	203	−0.09	0.33	−0.30	0.39
carboxylic acids	5	14	124	−0.04	0.35	−0.85	0.88
esters	14	8	249	0.02	0.29	−0.45	0.51
bifunctional compounds containing H, C, O	5	8	28	0.41	0.86	0.32	0.69
inorganic compounds containing H and O	2	9	22	0.00	0.58	0.04	0.59
aliphatic amines	15	10	168	0.01	0.27	0.18	0.36
aromatic amines	11	12	81	0.14	0.43	0.13	0.42
nitriles	4	6	22	0.09	0.35	0.07	0.36
nitrohydrocarbons	6	8	38	0.00	0.23	−0.88	0.90
amides & ureas	4	6	11	1.38	1.49	1.06	1.27
bifunctional compounds containing N	6	3	11	−0.89	0.96	−0.75	1.15
inorganic compounds containing N	2	8	15	−0.18	0.57	−0.24	0.58
thiols	4	5	14	0.34	0.48	0.38	0.51
sulfides	6	6	23	−0.13	0.98	−0.13	1.08
disulfides	2	3	5	0.00	0.16	0.03	0.20
fluorinated hydrocarbons	9	5	19	−0.41	0.67	−0.31	0.63
chloroalkanes	13	5	35	−0.01	0.27	0.03	0.29
chloroalkenes	5	4	16	0.65	0.65	0.60	0.60
chloroarenes	8	6	37	−0.54	0.56	−0.50	0.52
brominated hydrocarbons	14	6	50	−0.19	0.33	−0.04	0.38
iodinated hydrocarbons	9	6	28	−0.06	0.41	0.11	0.54
multifunctional halogenated solutes	26	9	80	0.08	0.64	−0.13	0.79
total	260	19	2084	−0.02	0.38	−0.12	0.48

[a] Number of solutes in this solute class. [b] Number of solvent classes for which there are data for this solute class. [c] Total number of solute/solvent data involving solutes in this solute class. [d] Mean errors over this solute class data (kcal/mol).

etrization predicts solvation free energies that are undersolvated by an average of 0.42 kcal/mol for the data points in the training set. In general, the errors obtained with the SM5.2R model parametrizations for the MNDO, AM1, and PM3 Hamiltonians are of a similar or slightly larger magnitude compared with those achieved with the MNDO/d parametrization.

The performance of the SM5.2R model by solute class is shown in Tables 6 and 7. In the MNDO/d parametrization, relatively large mean unsigned errors ($\sim$1−1.5 kcal/mol) are obtained for the amide, urea, bifunctional-nitrogen-containing, and sulfide solutes. The largest signed error in the SM5.2R/MNDO/d parametrization is nearly 1.4 kcal/mol and occurs for

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3265**

TABLE 7: Performance of the SM5.2R Models for the MNDO, AM1, and PM3 Hamiltonian Parametrizations by Solute Functional Group Class

solute class	SM5.2R/MNDO// HF/MIDI!		SM5.2R/AM1// HF/MIDI!		SM5.2R/PM3// HF/MIDI!		SM5.2R/MNDO// MNDO	
	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]
unbranched alkanes	0.05	0.38	0.16	0.45	0.16	0.42	0.13	0.40
branched alkanes	0.07	0.45	0.01	0.39	0.08	0.42	0.19	0.42
cycloalkanes	−0.14	0.34	−0.25	0.31	−0.12	0.30	−0.06	0.33
alkenes	−0.05	0.21	0.21	0.31	0.11	0.25	−0.04	0.21
alkynes	0.04	0.13	−0.01	0.18	0.00	0.17	0.09	0.16
arenes	−0.27	0.39	−0.42	0.53	−0.39	0.47	−0.26	0.39
alcohols	−0.01	0.33	−0.02	0.35	−0.01	0.34	0.08	0.33
ethers	0.09	0.39	0.09	0.40	0.11	0.39	0.35	0.48
aldehydes	−0.27	0.46	−0.15	0.41	−0.23	0.43	−0.52	0.60
ketones	−0.09	0.33	−0.18	0.34	−0.15	0.35	−0.30	0.39
carboxylic acids	−0.04	0.35	−0.06	0.39	−0.06	0.38	−0.85	0.88
esters	0.02	0.29	0.04	0.33	0.04	0.31	−0.45	0.50
H, C, O bifunctional compounds	0.41	0.86	0.76	1.07	0.62	0.99	0.31	0.69
H and O inorganic compounds	0.00	0.58	−0.02	0.57	0.00	0.58	0.04	0.59
aliphatic amines	0.01	0.27	0.07	0.32	0.02	0.30	0.18	0.36
aromatic amines	0.14	0.43	−0.02	0.35	0.14	0.39	0.13	0.43
nitriles	0.09	0.35	0.32	0.39	−0.09	0.26	0.06	0.36
nitrohydrocarbons	0.00	0.23	0.02	0.57	−0.06	0.40	−0.88	0.89
amides & ureas	1.38	1.49	1.38	1.48	1.26	1.32	1.03	1.24
N bifunctional compounds	−0.89	0.96	−0.83	0.88	−0.84	0.91	−0.75	1.15
N inorganic compounds	−0.18	0.57	−0.28	0.62	−0.08	0.53	−0.24	0.58
thiols	0.30	0.45	0.31	0.43	0.29	0.40	0.33	0.49
sulfides	−0.10	0.84	−0.12	0.90	−0.06	0.83	−0.07	0.92
disulfides	0.00	0.17	0.00	0.16	0.00	0.17	0.37	0.37
fluorinated hydrocarbons	−0.44	0.68	−0.11	0.42	−0.10	0.46	−0.34	0.65
chloroalkanes	−0.16	0.36	−0.27	0.37	0.43	0.51	−0.17	0.33
chloroalkenes	0.52	0.54	0.73	0.73	0.88	0.88	0.40	0.42
chloroarenes	−0.62	0.64	−0.18	0.33	−0.32	0.37	−0.64	0.65
brominated hydrocarbons	−0.26	0.35	−0.10	0.42	−0.21	0.33	0.05	0.35
iodinated hydrocarbons	−0.04	0.48	−0.03	0.49	−0.09	0.37	0.03	0.56
multifunctional halogenated solutes	0.15	0.69	0.10	0.58	0.03	0.79	−0.03	0.85
total	−0.03	0.39	−0.02	0.41	−0.01	0.40	−0.12	0.48

[a] Mean errors over this solute class data (kcal/mol).

solutes in the amide and urea class. All solute classes that contain over 100 data points from our training set have mean unsigned errors that are smaller than 0.4 kcal/mol. These trends are generally mirrored for the MNDO, AM1, and PM3 parametrizations, although relatively large increases (∼0.1 kcal/mol) in the mean unsigned error for arene solvents are apparent in the AM1 and PM3 parametrizations.

One source of variation in the predictions of the various Hamiltonians (MNDO, AM1, ...) is the quality of the point charges. Orozco and Luque[60] have studied the MNDO and AM1 point charges as obtained by deorthogonalized Mulliken analysis, but we use zero-overlap Mulliken analysis, which is different. Although the MNDO and AM1 charges often differ significantly, we find we can obtain good-quality parametrizations using either.

The results for the ionic solutes in the training set are presented in Table 8 for the MNDO/d parametrization of the SM5.2R model using both HF/MIDI! and MNDO/d geometries. The results for the ionic solutes for the MNDO, AM1, and PM3 parametrizations are included in the Supporting Information.

Tables 6 and 7 also demonstrate the performance across the solute classes for the SM5.2R/MNDO/d and SM5.2R/MNDO parametrizations at the MNDO/d and MNDO geometries, respectively. In general, the models have slightly larger errors when using the geometries obtained using semiempirical Hamiltonians rather than the HF/MIDI! geometries for which the SM5.2R model was parametrized. The largest increase in errors occurs for carboxylic acids and nitrohydrocarbon solutes with increases in the mean unsigned error over the respective solute class of more than 0.5 kcal/mol. The amides and ureas as well

as the chloroalkene solute classes show an opposite trend, improving slightly when the less accurate method for obtaining geometries is employed.

The calculated solvation free energies for selected molecules in selected solvents are shown in Table 9. The results with the SM5.2R/MNDO/d parametrization are shown using the HF/MIDI! geometries as well as using MNDO/d geometries. For reference, the SM5.4/AM1 model,[1,3−5] which represents our most rigorous electrostatic treatment to date, including the CM1A model of class IV charges,[13] is included along with the experimental values. In general the SM5.4/AM1 and SM5.2R/MNDO/d//HF/MIDI! models perform similarly, with mean unsigned errors of 0.39 and 0.34 kcal, respectively, over the subset of molecules in Table 9. The mean unsigned error for the SM5.2R/MNDO/d parametrization applied using MNDO/d geometries rises to 0.41 kcal for this subset. This difference is due to the changes in geometry (and hence the changes in the solvated charge distribution) caused by moving from the more accurate geometries provided by HF/MIDI! to the MNDO/d geometries.

As mentioned in section 3, a new surface tension coefficient, $\sigma_{HN}^{(2)}$, and related surface tension functional were added to treat N−N single bonds. This functional works by adding a solvation free energy correction to each hydrogen atom that is attached to a nitrogen atom that is attached to a second nitrogen atom. Hydrazine is the only solute in our training set that is affected by this correction and has experimental solvation free energy data available, but this correction will affect any solute that contains the hydrazone functionality and has a hydrogen attached

TABLE 8: Calculated and Experimental Free Energies of Solvation (kcal/mol) for Ionic Solutes Used in the Parametrization of SM5.2R/MNDO/d

	SM5.2R/MNDO/d//HF/MIDI!			SM5.2R/MNDO/d//MNDO/d	expt
	ΔG_{EP}	G_{CDS}	ΔG_S°	ΔG_S°	ΔG_S°
HC_2^-	−80.5	1.7	−78.8	−78.8	−73
$CH_3OH_2^+$	−81.4	−1.6	−83.0	−83.3	−87
$(CH_3)_2OH^+$	−69.6	−1.1	−70.7	−70.0	−70
$CH_3CH_2OH_2^+$	−76.4	−1.7	−78.1	−78.9	−81
$CH_3C(OH)CH_3^+$	−66.5	−2.6	−69.1	−68.3	−64
H_3O^+	−100.1	0.7	−99.4	−99.1	−105
CH_3O^-	−80.6	−2.0	−82.6	−82.3	−98
$CH_3CO_2^-$	−71.7	−1.2	−72.8	−75.6	−77
$CH_3COCH_2^-$	−69.0	−1.6	−70.6	−71.3	−81
$C_6H_5O^-$	−61.7	−3.3	−65.0	−62.9	−75
$C_6H_5CH_2^-$	−56.2	−0.2	−56.4	−56.2	−59
OH^-	−101.7	−7.4	−109.0	−109.3	−110
HO_2^-	−87.9	−7.2	−95.1	−97.7	−101
O_2^-	−86.8	−5.9	−92.7	−93.5	−87
$CH_3NH_3^+$	−74.6	−2.3	−77.0	−76.8	−73
$HC(OH)NH_2^+$	−72.7	−7.6	−80.3	−80.3	−78
CH_3CNH^+	−67.1	−1.2	−68.4	−68.0	−69
$CH_3C(OH)NH_2^+$	−66.2	−8.0	−74.1	−73.6	−70
$(CH_3)_2NH_2^+$	−67.5	0.0	−67.5	−66.9	−66
$(CH_3)_3NH^+$	−62.2	2.6	−59.6	−58.7	−59
imidazole·H^+	−61.4	−0.3	−61.7	−61.3	−64
pyridine·H^+	−58.3	−1.1	−59.4	−58.8	−58
$C_6H_5NH_3^+$	−64.1	−3.4	−67.5	−67.3	−68
NH_4^+	−84.5	−4.1	−88.6	−88.6	−81
CN^-	−77.5	−0.1	−77.6	−77.4	−75
CH_2CN^-	−68.2	−1.9	−70.2	−70.0	−75
NH_2^-	−85.6	−5.0	−90.5	−90.4	−95
NO_2^-	−76.1	0.5	−75.6	−75.7	−73
NO_3^-	−68.8	5.2	−63.6	−63.6	−66
N_3^-	−66.1	−10.1	−76.2	−76.1	−74
$CH_3SH_2^+$	−74.0	0.8	−73.2	−73.5	−74
$(CH_3)_2SH^+$	−68.4	1.3	−67.1	−66.9	−61
HS^-	−79.9	−2.4	−82.2	−82.3	−76
CH_3S^-	−75.7	−1.8	−77.5	−76.6	−76
$CH_3CH_2S^-$	−73.7	−1.8	−75.5	−74.6	−74
$n\text{-}C_3H_7S^-$	−73.9	−1.7	−75.6	−73.8	−76
$C_6H_5S^-$	−63.0	−3.4	−66.3	−64.5	−65
F^-	−109.3	1.5	−107.8	−107.8	−107
$CHF_2CO_2^-$	−64.6	0.4	−64.2	−64.8	−70
Cl^-	−76.9	0.0	−76.9	−76.9	−78
$CHCl_2CO_2^-$	−57.5	−1.4	−58.9	−63.0	−66
Br^-	−70.9	−0.8	−71.8	−71.8	−72
I^-	−61.6	−1.5	−63.1	−63.1	−63
mean unsigned error			3.6	3.3	
mean signed error			0.7	0.7	

to one of the involved nitrogen atoms. For this reason, we searched the MedChem database[51] for experimental data that could test the effects of the new $\sigma_{HN}^{(2)}$ surface tension and related functional on other molecules outside our training set. We were able to locate a few octanol/water partition coefficients for molecules without complicating conformational issues, and these are listed in Table 10. We then applied the SM5.2R/ MNDO/d method to the three solutes listed at the MNDO/d geometries both with the $\sigma_{HN}^{(2)}$ surface tension coefficient as parametrized in the full model and with the coefficient set to zero. In all three cases, the hydrazine correction was in the correct direction, leading us to believe that the surface tension did identify a true physical effect and that it will also generally improve calculated results for solutes containing hydrazone character.

When models such as SM5.2R are parametrized to predict solvation free energies, it is important to be concerned with free energy *trends* across a group of solutes in addition to the overall mean unsigned error in the solvation free energy obtained for a group of solutes. One such important trend is the effect of adding a CH$_2$ group to a solute. Table 11 lists the

approximate change in solvation free energy when adding a CH$_2$ group for alkanes, alcohols, aldehydes, and ketones using water, 1-octanol, and hexadecane as solvents. Across the SMx models tabulated, the cost of adding a CH$_2$ group is relatively stable within a model. The magnitude of the error in the trends for the SM5.2R/MNDO/d parametrization is always less than one-fourth the magnitude of the experimental trend for adding a CH$_2$ group. The error in the trend for adding a CH$_2$ group in the SM5.2R/MNDO/d model could have been improved by considering the trends while fitting the linear surface tension coefficient parameters, but the mean unsigned error in the solvation free energies predicted by the resulting model would have increased significantly. Given these difficulties and the always present uncertainties in the experimental results, we feel that the trends predicted by the SM5.2R/MNDO/d and SM5.2R/ MNDO models are acceptable.

Table 12 shows the breakdown of the predicted free energy of solvation for the SM5.2R/MNDO/d model into the electrostatic (ΔG_{EP}) and nonelectrostatic (G_{CDS}) components. As mentioned earlier, there is no rigorous physical method to define an accurate separation of these components. In the SM5.2R

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3267**

TABLE 9: Solvation Free Energies (kcal/mol) for Selected Solutes in Selected Solvents Using SM5.2R/MNDO/d at both the HF/MIDI! Geometries and the MNDO/d Geometries

solute	method	solvent						
		hexadecane	cyclohexane	benzene	1-octanol	ethylether	chloroform	water
n-octane								
	SM5.2R/MNDO/d//HF/MIDI!	−4.50	−4.63	−5.40	−3.98	−5.01	−5.49	2.10
	SM5.2R/MNDO/d//MNDO/d	−4.41	−4.55	−5.32	−3.89	−4.94	−5.43	2.31
	SM5.4/AM1	−4.61	−4.76	−5.55	−4.40	−4.99	−5.07	2.15
	expt	−5.02	−5.63	−5.35	−4.18	−5.62	−5.25	2.89
benzene								
	SM5.2R/MNDO/d//HF/MIDI!	−4.40	−4.46	−5.10	−4.40	−5.06	−5.02	−1.18
	SM5.2R/MNDO/d//MNDO/d	−4.37	−4.43	−5.08	−4.40	−5.07	−5.02	−1.17
	SM5.4/AM1	−3.86	−3.95	−4.53	−3.87	−4.67	−4.69	−0.97
	expt	−3.80	−4.19	−4.55	−3.72	−4.21	−4.64	−0.87
toluene								
	SM5.2R/MNDO/d//HF/MIDI!	−4.75	−4.82	−5.53	−4.71	−5.48	−5.49	−0.89
	SM5.2R/MNDO/d//MNDO/d	−4.72	−4.80	−5.52	−4.71	−5.48	−5.49	−0.85
	SM5.4/AM1	−4.52	−4.61	−5.28	−4.66	−5.38	−5.38	−0.89
	expt	−4.54	−4.90	−5.32	−4.55	−5.23	−5.48	−0.89
1-butanol								
	SM5.2R/MNDO/d//HF/MIDI!	−3.64	−3.71	−4.67	−6.26	−5.39	−5.09	−4.82
	SM5.2R/MNDO/d//MNDO/d	−3.50	−3.57	−4.52	−6.12	−5.26	−4.93	−4.48
	SM5.4/AM1	−3.39	−3.47	−4.31	−5.56	−5.11	−4.95	−4.39
	expt	−3.55	−3.52	−4.45	−5.71	−5.69	−5.28	−4.72
phenol								
	SM5.2R/MNDO/d//HF/MIDI!	−5.39	−5.44	−6.57	−8.13	−7.38	−6.79	−6.46
	SM5.2R/MNDO/d//MNDO/d	−5.32	−5.38	−6.53	−8.17	−7.41	−6.76	−6.36
	SM5.4/AM1	−5.47	−5.52	−6.45	−7.91	−7.71	−7.21	−6.55
	expt	−5.14	−5.57	−7.12	−8.69	−8.75	−7.14	−6.62
1,4-dioxane								
	SM5.2R/MNDO/d//HF/MIDI!	−4.11	−4.15	−4.73	−5.02	−4.99	−5.50	−5.31
	SM5.2R/MNDO/d//MNDO/d	−3.84	−3.89	−4.47	−4.79	−4.77	−5.25	−4.79
	SM5.4/AM1	−3.65	−3.73	−4.47	−5.49	−5.13	−6.31	−6.03
	expt	−3.82	−4.17	−5.21	−4.89	−4.67	−6.21	−5.05
butanone								
	SM5.2R/MNDO/d//HF/MIDI!	−3.48	−3.54	−4.15	−4.72	−4.54	−4.87	−3.26
	SM5.2R/MNDO/d//MNDO/d	−3.61	−3.67	−4.30	−5.00	−4.76	−5.11	−3.64
	SM5.4/AM1	−3.31	−3.38	−4.07	−4.96	−4.83	−5.00	−3.62
	expt	−3.12	−3.48	−4.46	−3.78	−4.09	−5.43	−3.64
propanoic acid								
	SM5.2R/MNDO/d//HF/MIDI!	−3.69	−3.73	−4.74	−6.82	−6.08	−5.41	−6.13
	SM5.2R/MNDO/d//MNDO/d	−4.14	−4.17	−5.25	−7.75	−6.83	−6.17	−7.28
	SM5.4/AM1	−2.94	−3.00	−3.98	−6.22	−6.02	−5.22	−5.85
	expt	−3.12	−3.78	−4.75	−6.86	−6.75	−5.37	−6.47
butylamine								
	SM5.2R/MNDO/d//HF/MIDI!	−3.65	−3.73	−4.46	−5.25	−4.55	−4.87	−4.05
	SM5.2R/MNDO/d//MNDO/d	−3.58	−3.67	−4.40	−5.15	−4.50	−4.80	−3.82
	SM5.4/AM1	−3.95	−4.04	−4.71	−5.39	−4.80	−5.52	−4.26
	expt	−3.57	−3.72	−4.33	−5.35	−4.44	−5.35	−4.29
pyridine								
	SM5.2R/MNDO/d//HF/MIDI!	−4.56	−4.61	−5.17	−5.27	−5.12	−5.64	−4.27
	SM5.2R/MNDO/d//MNDO/d	−4.55	−4.60	−5.17	−5.31	−5.17	−5.68	−4.29
	SM5.4/AM1	−4.54	−4.59	−5.01	−6.24	−5.40	−6.56	−5.94
	expt	−4.10	−4.30	−5.28	−5.34	−4.81	−6.45	−4.70
aniline								
	SM5.2R/MNDO/d//HF/MIDI!	−5.40	−5.46	−6.39	−6.86	−6.71	−6.50	−5.09
	SM5.2R/MNDO/d//MNDO/d	−5.36	−5.43	−6.37	−6.82	−6.70	−6.48	−5.01
	SM5.4/AM1	−5.71	−5.78	−6.61	−7.16	−7.19	−7.15	−5.14
	expt	−5.44	−5.52	−6.88	−6.71	−6.51	−7.34	−5.49
nitrobenzene								
	SM5.2R/MNDO/d//HF/MIDI!	−6.04	−6.07	−6.68	−6.32	−7.21	−7.95	−4.17
	SM5.2R/MNDO/d//MNDO/d	−6.60	−6.62	−7.32	−7.42	−8.13	−8.92	−5.41
	SM5.4/AM1	−5.88	−5.92	−6.43	−5.51	−7.10	−7.75	−3.07
	expt	−6.22	−6.62	−7.60	−6.63	−6.85	−7.78	−4.12

model, the electrostatics are calculated using class II charges centered at each of the atoms in the molecule. In general, these charges are smaller in magnitude than class IV charges,[13] which were developed to accurately reproduce gas-phase dipole moments. Thus, as is demonstrated in Table 12, the ΔG_{ENP} terms for models based on class IV charges, such as SM5.4/ AM1,[1] are generally larger in magnitude than the ΔG_{EP} terms for the models presented here. (Note that the error due to neglecting the changes in the nuclear internal energy is much

smaller than the differences in the electrostatic terms shown in Table 12.) Since ΔG_{EP} or ΔG_{ENP} varies as the square of the point charges, the smaller class II point charges must be compensated for by a more negative or less positive G_{CDS} term (when compared to a model based on class IV charges) to predict accurate solvation free energies. If this argument is taken to the limit, where there is a zero charge on all atoms in a neutral solute, the ΔG_{ENP} or ΔG_{EP} term is zero and the G_{CDS} term must compensate for the lack of electrostatics in order to predict

TABLE 10: Effect of $\sigma_{HN}^{(2)}$ and Related Surface Tension Functional on Solutes Containing Hydrazone Functionality

	SM5.2R/MNDO/d//MNDO/d with $\sigma_{HN}^{(2)}$ correction			SM5.2R/MNDO/d//MNDO/d without $\sigma_{HN}^{(2)}$ correction			expt
solute compound	$\Delta G_S^{\circ a}$ (octanol)	$\Delta G_S^{\circ a}$ (water)	log $P_{oct/wat}$	$\Delta G_S^{\circ a}$ (octanol)	$\Delta G_S^{\circ a}$ (water)	log $P_{oct/wat}$	log $P_{oct/wat}{}^b$
semicarbazide	−14.4	−18.6	−3.1	−11.6	−16.9	−3.9	−2.75
4-aminobenzaldehydesemicarbazone	−19.5	−15.5	−0.7	−15.7	−18.1	−1.8	0.80
benzophenonehydrazone	−14.1	−10.3	2.8	−12.2	−9.1	2.2	2.79

^{a} In units of kcal/mol. ^{b} Partition coefficient for solute for octanol/water from the MedChem database.[51]

TABLE 11: Calculated and Experimental Free Energy Cost (kcal/mol) of Adding CH₂ to a Solute

ref solute, But	ref solute, Oct	SM5.2R/MNDO/d//HF/MIDI! $\Delta\Delta G_S^{\circ a}$	SM5.4/AM1 $\Delta\Delta G_S^{\circ a}$	SM5.0R $\Delta\Delta G_S^{\circ a}$	expt $\Delta\Delta G_S^{\circ a}$
		Solvent: Water			
n-butane	n-octane	0.15	0.18	0.13	0.20
1-butanol	1-octanol	0.16	0.18	0.14	0.16
butanal	octanal	0.17	0.17	0.14	0.22
butanone	2-octanone	0.18	0.23	0.14	0.19
		Solvent: Octanol			
n-butane	n-octane	−0.55	−0.72	b	−0.58
1-butanol	1-octanol	−0.55	−0.72	b	−0.61
butanone	2-octanone	−0.53	−0.68	b	−0.65
		Solvent: Hexadecane			
n-butane	n-octane	−0.61	−0.74	b	−0.71
1-butanol	1-octanol	−0.61	−0.74	b	−0.69
butanal	octanal	−0.60	−0.74	b	−0.72
butanone	2-octanone	−0.59	−0.72	b	−0.67

^{a} $\Delta\Delta G_S^{\circ} = (\Delta G_S^{\circ}$(solute Oct) $- (\Delta G_S^{\circ}$(solute But))/4. Dividing by 4 yields the change per CH₂ group. ^{b} The SM5.0R model has not been parametrized for organic solvents yet.

accurate solvation free energies. This is essentially the idea behind the SM5.0R model,[27] which contains only implicit electrostatics. Thus, the SM5.2R model displays characteristics of both the SM5.4 model and the SM5.0 model. In aqueous solution, the ΔG_{EP} term for alcohols using an SM5.2R model is typically one-third of the magnitude obtained with an SM5.4 model. The nonelectrostatic term in the SM5.2R models is substantially larger than in an SM5.4 model to compensate for the difference. For the halogenated hydrocarbons, the ΔG_{EP} and ΔG_{ENP} terms are similar for both the SM5.2R/MNDO/d and SM5.4/AM1 parametrizations, and hence, the G_{CDS} terms are also comparable.

Users of models based on class II charges such as SM5.2R need to be extra cautious when using the model to predict solvation free energies when the point charges calculated from the underlying Hamiltonian are suspect. Consider the solutes in our training set that contain sulfur. In every sulfur-containing nonaromatic solute, the magnitudes of the aqueous solvated class II charges are diminished from those predicted by the class IV charges, and the ΔG_{EP} term for these molecules is less negative for SM5.2R/MNDO/d than the ΔG_{ENP} term in SM5.4/AM1 (see Table 12). However, the class II charges in solutes containing both a sulfur atom and an aromatic ring are not systematically smaller than the class IV charges. In thiophene, the SM5.2R/MNDO/d aqueous solvated class II charge on the sulfur atom is 0.53, nearly 5 times the charge predicted by CM1A class IV charges. Thiophenol and thioanisole also have class II point charges relaxed by the aqueous reaction field that are sometimes larger in magnitude than CM1A relaxed class IV charges, but the difference is much less pronounced than for thiophene where the resulting ΔG_{EP} term is larger for SM5.2R/MNDO/d than it is for the ΔG_{ENP} term in SM5.4/AM1. The atomic surface tension functionals are not capable of compensating for these large fluctuations in the underlying point charge distribution provided by class II charges and the resulting changes in the

ΔG_{EP} term. Hence, the mean unsigned error in the predicted solvation free energies for sulfur-containing compounds is significantly larger for all parametrizations of the SM5.2R model than it was for the SM5.4 parametrizations. The SM5.2R model oversolvates thiophene and undersolvates the nonaromatic thiols and sulfide solutes. Our fit of the sulfur-containing compounds (other than thiophene) could have been improved by over 30% by removing thiophene from the training set, but we believe it is more important to find parameters that have been averaged over all the functionalities a user is likely to be interested in rather than finding parameters that are limited to a small subset of the potential functionalities. Thus, thiophene remained in the parametrization suite.

To test the predictive power of the SM5.2R model, we created four experimental data sets that each contained only 75% of the data from our overall training set. This was accomplished by first separating the aqueous data from the organic data. Each of these two subsets was then put in order by solvent class. (See list of solvent classes in Table 4.) Within each solvent class the data were ordered by number of carbons in the solvent, then by the number of oxygen atoms in the solvent, then by the number of nitrogen atoms in the solvent, ..., until all the 91 solvents were ordered. Then the data within each solvent were ordered by the solute class as listed in Table 6. Within a given solute class, the data were ordered by the number of carbons in the solute molecule, then by the number of oxygen atoms, then by the number of nitrogen atoms, ..., until the data set was entirely ordered. This order for both the organic and aqueous subsets of the data points is reproduced in the Supporting Information for this paper. Once the data were ordered, set 1 was created by dropping out data 4, 8, 12, ... from both the organic and aqueous data subsets. Set 2 excluded data 3, 7, 11, ..., and so forth. Thus, each molecule in the overall training set was removed from set 1, set 2, set 3, or set 4. We then refit all the linear surface tensions in the SM5.2R/MNDO/d param-

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3269**

TABLE 12: Calculated and Experimental Free Energies of Solvation (kcal/mol) for Selected Molecules

solute	SM5.2R/MNDO/d//HF/MIDI!			SM5.4/AM1			SM5.0R ΔG_S°	expt ΔG_S°
	ΔG_{EP}	G_{CDS}	ΔG_S°	ΔG_{ENP}	G_{CDS}	ΔG_S°		
				Solvent: Water				
n-butane	0.0	1.5	1.5	−0.2	1.7	1.4	1.6	2.1
n-octane	0.0	2.1	2.1	−0.1	2.2	2.2	2.2	2.9
1-butanol	−1.1	−3.8	−4.8	−3.7	−0.7	−4.4	−4.6	−4.7
1-octanol	−1.1	−3.1	−4.2	−3.5	−0.1	−3.7	−4.0	−4.1
butanal	−2.2	−1.0	−3.2	−4.3	1.0	−3.3	−3.1	−3.2
octanal	−2.1	−0.4	−2.5	−4.2	1.6	−2.6	−2.6	−2.3
butanone	−2.2	−1.1	−3.3	−4.6	1.0	−3.6	−3.2	−3.6
2-octanone	−2.1	−0.4	−2.6	−4.3	1.6	−2.7	−2.6	−2.9
butanoic acid	−3.0	−2.8	−5.9	−7.1	1.7	−5.4	−5.8	−6.4
hexanoic acid	−3.0	−2.5	−5.5	−7.0	2.0	−5.0	−5.6	−6.2
methyl butanoate	−2.9	0.2	−2.7	−4.8	1.9	−2.9	−3.0	−2.8
methyl octanoate	−2.9	0.9	−2.0	−4.7	2.5	−2.2	−2.9	−2.0
pentylamine	−0.2	−3.7	−3.9	−1.7	−2.3	−4.1	−4.1	−4.1
ethanethiol	−0.2	−0.5	−0.7	−1.3	0.2	−1.0	−0.8	−1.3
thiophene	−2.6	−1.6	−4.2	−2.5	−0.4	−2.9	−3.2	−1.4
diethyl sulfide	−0.1	0.0	−0.2	−1.5	0.4	−1.1	−1.1	−1.4
chloromethane	−1.7	1.1	−0.6	−1.5	1.0	−0.5	−0.2	−0.6
2-chlorobutane	−1.4	1.4	0.0	−1.4	1.3	−0.1	0.7	0.1
bromomethane	−1.1	0.4	−0.7	−1.5	0.5	−1.0	−0.6	−0.8
1-bromobutane	−0.9	0.6	−0.3	−1.2	0.6	−0.6	−0.2	−0.4
				Solvent: 1-Octanol				
n-butane	0.0	−1.8	−1.8	−0.2	−1.3	−1.5		−1.9
n-octane	0.0	−4.0	−4.0	−0.1	−4.3	−4.4		−4.2
1-butanol	−1.0	−5.3	−6.3	−3.4	−2.2	−5.6		−5.7
1-octanol	−1.0	−7.5	−8.4	−3.2	−5.2	−8.4		−8.1
butanal	−1.9	−2.7	−4.6	−3.8	−0.8	−4.6		−4.6
butanone	−1.9	−2.8	−4.7	−4.0	−0.9	−5.0		−3.8
2-octanone	−1.9	−5.0	−6.8	−3.7	−3.9	−7.7		−6.4
butanoic acid	−2.7	−4.6	−7.3	−6.5	−0.3	−6.7		−7.6
hexanoic acid	−2.7	−5.7	−8.3	−6.4	−1.8	−8.1		−8.8
methyl butanoate	−2.6	−2.0	−4.6	−4.3	−1.4	−5.7		−4.6
thiophene	−2.2	−3.4	−5.7	−2.3	−2.6	−4.9		−3.9
diethyl sulfide	−0.1	−3.1	−3.2	−1.3	−3.3	−4.6		−4.1
bromomethane	−1.0	−1.4	−2.4	−1.3	−1.0	−2.3		−2.4
1-bromobutane	−0.8	−3.4	−4.2	−1.0	−3.4	−4.5		−4.2
				Solvent: Hexadecane				
n-butane	0.0	−2.1	−2.1	−0.1	−1.5	−1.7		−2.2
n-octane	0.0	−4.5	−4.5	0.0	−4.6	−4.6		−5.0
1-butanol	−0.5	−3.1	−3.6	−1.8	−1.6	−3.4		−3.6
1-octanol	−0.5	−5.5	−6.1	−1.7	−4.6	−6.3		−6.3
butanal	−0.9	−2.4	−3.4	−1.9	−1.2	−3.1		−3.1
octanal	−0.9	−4.8	−5.8	−1.8	−4.3	−6.1		−6.0
butanone	−0.9	−2.6	−3.5	−2.0	−1.3	−3.3		−3.1
2-octanone	−0.9	−4.9	−5.8	−1.8	−4.4	−6.2		−5.8
butanoic acid	−1.4	−2.8	−4.2	−3.4	−0.2	−3.6		−3.9
hexanoic acid	−1.4	−4.1	−5.4	−3.3	−1.7	−5.0		−5.4
methyl butanoate	−1.4	−2.5	−3.8	−2.2	−1.9	−4.1		−4.0
pentylamine	−0.1	−4.2	−4.3	−0.8	−3.8	−4.7		−4.3
ethanethiol	−0.1	−2.3	−2.4	−0.6	−1.9	−2.5		−3.0
thiophene	−1.1	−3.7	−4.8	−1.2	−3.7	−4.9		−4.0
diethyl sulfide	−0.1	−3.5	−3.5	−0.7	−3.6	−4.2		−4.2
1-bromobutane	−0.4	−3.8	−4.2	−0.5	−3.6	−4.1		−4.2

etrization suite by fitting to the experimental results for the aqueous and organic portions of each of sets 1−4. In these fittings, any surface tension coefficient was set to zero if it did not affect any of the data in the training subset. Each set of resulting parameters was then used to predict the solvation free energies for all the molecules in the overall training set. The results of this test are presented in Table 13. Note that all linear parameters for all atom types were fit at one time for this test rather than using the multistage procedure presented in the parametrization section; thus, the error quoted for fitting the full set of data in the training set is slightly lower than the actual error in the SM5.2R/MNDO/d parametrization. The mean unsigned error over the entire training set using the linear parameters determined with each data subset is very similar. In other words, using a data set only 75% as large as the one we

fit, we can predict the remaining 25% of the data quite well. This indicates that our model does indeed have predictive power for solutes similar to those in our training set but not actually included in our training set.

Table 14 demonstrates the reliability of the SM5.2R model when applied to geometries optimized with various semiempirical Hamiltonians. For the SM5.2R/MNDO/d parametrization, using semiempirical geometries increases the mean unsigned error in the predicted solvation free energies for the set of neutrals (excluding P) in the training set by 0.05−0.09 kcal/mol. The largest deviations from the results obtained with the SM5.2R/MNDO/d parametrization using HF/MIDI! geometries occur for carboxylic acids and nitrohydrocarbons where the mean unsigned deviations are slightly larger than 0.8 kcal/mol using MNDO/d geometries. The typical solute class deviates

TABLE 13: Testing the Predictive Ability of the SM5.2R/MNDO/d Parametrization of the SM5.2R Model

set of molecules	number of data fit		total number of data		MUE[a] over all data		
	water[b]	organic[c]	water[b]	organic[c]	water[b]	organic[c]	total
set 1[d]	186	1376	248	1834	0.46	0.37	0.377
set 2[e]	186	1376	248	1834	0.53	0.36	0.384
set 3[f]	186	1375	248	1834	0.47	0.37	0.380
set 4[g]	186	1375	248	1834	0.48	0.37	0.381
full set	248	1834	248	1834	0.45	0.36	0.375

[a] MUE is the mean unsigned error over the set of data. [b] Number of data for which water is the solvent. [c] Number of data for which the solvent is organic. [d] Excludes the 4th, 8th, 12th, ... solute molecules from the parametrization. [e] Excludes the 3rd, 7th, 11th, ... solute molecules from the parametrization. [f] Excludes the 2nd, 6th, 10th, ... solute molecules from the parametrization. [g] Excludes the 1st, 5th, 9th, ... solute molecules from the parametrization.

TABLE 14: Deviation from the SM5.2R/MNDO/d//HF/MIDI! Results for the SM5.2R/MNDO/d Parametrization at MNDO/d, MNDO, AM1, and PM3 Geometries by Solute Functional Group Class

	SM5.2R/MNDO/d//							
	MNDO/d		MNDO		AM1		PM3	
solute class	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]	signed[a]	unsigned[a]
unbranched alkanes	0.08	0.08	0.08	0.08	0.04	0.04	0.03	0.04
branched alkanes	0.12	0.12	0.06	0.07	0.09	0.10	0.05	0.07
cycloalkanes	0.08	0.08	0.06	0.06	0.05	0.05	0.03	0.03
alkenes	0.01	0.02	0.05	0.06	−0.01	0.03	0.04	0.07
alkynes	0.05	0.05	0.00	0.06	0.01	0.03	−0.03	0.04
arenes	0.00	0.02	−0.02	0.03	−0.01	0.01	−0.03	0.03
alcohols	0.09	0.13	0.09	0.12	0.05	0.08	0.05	0.08
ethers	0.26	0.26	0.20	0.20	0.03	0.05	0.04	0.05
aldehydes	−0.26	0.26	−0.20	0.21	−0.21	0.21	−0.13	0.14
ketones	−0.20	0.20	−0.19	0.19	−0.14	0.14	−0.11	0.12
carboxylic acids	−0.81	0.81	−0.76	0.76	−0.58	0.58	−0.54	0.54
esters	−0.48	0.48	−0.42	0.43	−0.35	0.35	−0.30	0.30
H, C, O bifunctional compounds	−0.10	0.27	−0.05	0.25	−0.10	0.17	0.00	0.15
H and O inorganic compounds	0.04	0.05	0.04	0.04	0.01	0.02	0.01	0.02
aliphatic amines	0.12	0.12	0.06	0.06	0.05	0.07	0.01	0.03
aromatic amines	−0.01	0.03	0.00	0.05	−0.06	0.06	−0.04	0.08
nitriles	−0.01	0.03	−0.05	0.10	−0.02	0.03	−0.06	0.10
nitrohydrocarbons	−0.88	0.88	−0.72	0.72	−0.64	0.64	−0.57	0.57
amides & ureas	−0.33	0.33	−0.34	0.47	−0.26	0.27	−0.20	0.35
N bifunctional compounds	0.12	0.51	−0.02	0.21	−0.21	0.42	−0.13	0.24
N inorganic compounds	−0.07	0.07	−0.06	0.08	−0.05	0.05	−0.04	0.05
thiols	0.04	0.04	0.00	0.06	−0.01	0.05	−0.03	0.04
sulfides	0.00	0.14	−0.05	0.29	−0.13	0.20	−0.02	0.16
disulfides	0.04	0.08	0.18	0.20	0.08	0.15	0.01	0.10
fluorinated hydrocarbons	0.09	0.14	0.14	0.33	−0.09	0.12	0.06	0.28
chloroalkanes	0.05	0.15	0.20	0.30	0.18	0.20	0.26	0.33
chloroalkenes	−0.05	0.09	0.07	0.23	0.13	0.13	0.21	0.24
chloroarenes	0.04	0.05	0.06	0.11	0.17	0.17	0.20	0.20
brominated hydrocarbons	0.15	0.15	0.12	0.18	0.14	0.14	0.05	0.11
iodinated hydrocarbons	0.18	0.18	0.10	0.29	0.24	0.25	0.12	0.31
multifunctional halogenated solutes	−0.21	0.30	−0.08	0.41	−0.25	0.39	−0.09	0.42
total mean deviation	−0.11	0.23	−0.09	0.23	−0.09	0.17	−0.07	0.16
total mean error in predicted solvation free energy	−0.13	0.47	−0.12	0.46	−0.12	0.44	−0.10	0.43

[a] Mean deviations over this solute class data in kcal/mol.

by an average of less than 0.2 kcal/mol when using MNDO/d, MNDO, AM1, or PM3 semiempirical geometries, indicating that the method is relatively stable to the origin of the gas-phase geometries used in conjunction with the SM5.2R model.

5. Concluding Remarks

We have presented a new solvation model, SM5.2R, that has proved to be remarkably successful for calculating solvation free energies in both aqueous and organic solvents. In fact, it is our most successful model to date probably because we are now able to build on considerable experience in parametrizing general-purpose solvation models and also because of our new choice of effective solvent radius. When the geometrical relaxation effects are built into the parametrization, the present model affords one the opportunity to calculate solvation energies on the basis of accurate or high-level gas-phase structures, and

thus, it should be very useful for state-of the art calculations on complex systems.

Aside from the improvements in the mean error, why should one prefer SM5.2R to the previous SM5.4 models? We can give six more reasons. (1) The data set used for parametrization is larger, and hence, the parametrizations may be more robust. (2) Solvation in benzene, toluene, or chloroform is treated using the same parameters as for other solvents, whereas previously they required special treatment; this results from the addition of new solvent descriptors that allow us to parametrize all solvents in a consistent way. (3) The new model is parametrized for four Hamiltonians, including one with d orbitals, whereas the SM5.4 model is parametrized for only two, both without d orbitals. (4) The SM5.2R models are easier to code. (5) The calculations are less expensive. (6) The "rigid" character of the SM5.2R model makes it suitable for calculations on systems

Solvation Free Energies

J. Phys. Chem. B, Vol. 102, No. 17, 1998 **3271**

where geometry optimization at the semiempirical level gives unacceptable results. Examples of the latter include five-membered rings (which tend to be too flat at the semiempirical level), hydrogen bonds (which usually have qualitatively wrong structures in AM1), hydrazine (for which the true global minimum is not even a local minimum in PM3), and sugars (where semiempirical calculations have severe deficiencies for conformational analysis).

Do the SM5.4 models have compensating advantages such that they will still often be preferred to SM5.2R? Yes, they do. We note the following. (1) The SM5.4 model is based on more accurate charges than the SM5.2R model; thus, it provides a more physical explanation of the *origin* of solvation effects and it may be more robust for solutes differing qualitatively from any molecule in the parametrization set. (2) As a corollary, calculated dipole moments and charge distributions for molecules in solution will be more accurate with the SM5.4 model. (3) The use of a finite solvent radius in the first-solvent-shell terms may be more accurate for treating hard-to-access binding sites such as sterically hindered reaction centers or binding pockets of enzymes. (4) In some cases it is essential to explicitly model the relaxation of solute geometry in solution. Examples of the latter include 2,4-pentanedione,[61] transition states for S_N1 reactions,[62] the transition state for the 1,2-hydrogen migration in methylphenylnitrenium cation in acetonitrile,[63] or any case with a soft mode that might be especially sensitive to solvation.

The existence of multiple models and parametrizations for calculations of the structures and energetics of condensed-phase systems brings the status of this field closer to that for the gas phase where the computational chemist has a healthy choice of basis sets, levels of approximation, and parametrizations. The present paper increases the number of universal parametrizations (i.e., parametrizations valid for water and all organic solvents) from two to six and makes it more likely that a suitable modeling tool will be available for any given application.

Supporting Information Available: Five tables that include the surface tension coefficients optimized for the MNDO, AM1, and PM3 parametrizations of the SM5.2R model, the predicted solvation free energies for all 2084 neutral data points and all four parametrizations of the SM5.2R model, and the calculated free energies of aqueous solvation for the ionic solutes in the parametrization suite for the MNDO, AM1, and PM3 parametrizations of the SM5.2R model, and seven tables that contain the extension of the SM5.2R model to phosphorus (106 pages). Ordering information is given on any current masthead page.

References and Notes

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

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

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

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

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

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

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

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

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

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

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

(12) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Viellard, A., Eds.; Reidel: Dordrecht, 1980; Vol. 2, p 25.

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

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

(15) Dewar, M. J. S.; Zoebisch, E. G. *J. Mol. Struct.: THEOCHEM* **1988**, *180*, 1.

(16) Dewar, M. J. S.; Yate-Ching, Y. *Inorg. Chem.* **1990**, *29*, 3881.

(17) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.

(18) Thiel, W.; Voityuk, A. A. *Int. J. Quantum Chem.* **1992**, *44*, 807.

(19) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1996**, *93*, 315.

(20) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616.

(21) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

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

(23) Becke, A. *J. Chem. Phys.* **1986**, *84*, 4524.

(24) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(25) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(26) Zerner, M. C. *Rev. Comput. Chem.* **1991**, *2*, 313.

(27) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B*, in press.

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

(29) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.

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

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

(32) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 343.

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

(34) Kozaki, T.; Morihashi, M.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.

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

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

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

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

(39) *CRC Handbook of Chemistry and Physics*, 75th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1995.

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

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

(42) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085.

(43) Suleiman, D.; Eckert, C. A. *J. Chem. Eng. Data* **1994**, *39*, 692.

(44) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.

(45) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291.

(46) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 287.

(47) Wagman, D. D. *J. Phys. Chem. Ref. Data* **1982**, *11* (Suppl. 2).

(48) Wolfenden, R. *Biochemistry* **1978**, *17*, 201.

(49) Wauchope, R. D.; Haque, R. *Can. J. Chem.* **1972**, *50*, 133.

(50) Han, P.; Bartles, D. M. *J. Phys. Chem.* **1990**, *94*, 7294.

(51) Leo, A. J. Masterfile. *MedChem Software*; BioByte Corp.: Claremont, CA, 1994.

(52) Ferguson, D. M.; Pearlman, D. A.; Swope, W. C.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 362.

(53) Dallas, A. J. Ph.D. Thesis, University of Minnesota, Minneapolis, 1995.

(54) Zhang, Y.; Dallas, A. J.; Carr, P. W. *J. Chromatogr.* **1993**, *638*, 43.

(55) Abraham, M. H. *J. Chromatogr.* **1993**, *644*, 98.

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

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

(58) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281. Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.*, in press.

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

(60) Orozco, M.; Luque, F. J. *J. Comput. Chem.* **1990**, *11*, 909.

(61) Cramer, C. J.; Truhlar, D. G. In *Solvent Effects and Chemical Reactivity*; Tapia, O., Bertrán, J., Eds.; Kluwer: Dordrecht, 1996; p 1.

(62) Mathis, J. R.; Hynes, J. T. *J. Phys. Chem.* **1994**, *98*, 5445.

(63) Cramer, C. J.; Truhlar, D. G.; Falvey, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 12338.